CUT&RUN detects distinct DNA footprints of RNA polymerase II near the transcription start sites

Michi Miura* and Honglin Chen

State Key Laboratory for Emerging Infectious Diseases and

Department of Microbiology, The University of Hong Kong,

Hong Kong SAR, China

*E-mail: michi.miura.res@gmail.com

Abstract

CUT&RUN is a powerful tool to study protein-DNA interactions *in vivo*. DNA fragments cleaved by the targeted micrococcal nuclease identify the footprints of DNA-binding proteins on the chromatin. We performed CUT&RUN on human lung carcinoma cell line A549 maintained in a multi-well cell culture plate to profile RNA polymerase II. Long (>270 bp) DNA fragments released by CUT&RUN corresponded to the bimodal peak around the transcription start sites, as previously seen with chromatin immunoprecipitation. However, we found that short (<120 bp) fragments identify a well-defined peak localised at the transcription start sites. This distinct DNA footprint of short fragments, which constituted only about 5% of the total reads, suggests the transient positioning of RNA polymerase II before promoter-proximal pausing, which has not been detected in the physiological settings by standard chromatin immunoprecipitation. We showed that the positioning of the large-size-class DNA footprints around the short-fragment peak was associated with the directionality of transcription, demonstrating the biological significance of distinct CUT&RUN footprints of RNA polymerase II.

Keywords: CUT&RUN; chromatin profiling; RNA polymerase II; transcription initiation; promoter-proximal pausing; divergent transcription

37 **Abbreviations**

38 CUT&RUN: Cleavage Under Targets and Release Using Nuclease

39 pAG-MNase: Protein A/G-fused micrococcal nuclease

40 ChIP: chromatin immunoprecipitation

41 Pol II: RNA polymerase II

42 TSS: transcription start site

43 S5p Pol II: Serine 5-phosphorylated RNA polymerase II

44 mNET-seq: mammalian native elongating transcript-sequencing

45 # Introduction

46 CUT&RUN (Cleavage Under Targets and Release Using Nuclease) (Skene and Henikoff
47 2017) is a powerful tool to map protein-DNA interactions *in vivo*. CUT&RUN directs
48 Protein A/G-fused micrococcal nuclease (pAG-MNase) (Meers et al. 2019a) to the
49 antibody-bound protein-DNA complex, with a subsequent release of DNA fragments
50 cleaved in the presence of calcium ions. There are two major advantages of CUT&RUN
51 over conventional chromatin immunoprecipitation (ChIP). First, CUT&RUN does not
52 require sonication for chromatin fragmentation. Physical shearing of chromatin by
53 sonication does not occur randomly, which can skew signal detection; and excessive
54 sonication may attenuate the epitope of the target protein (Marx 2019). Second,
55 CUT&RUN identifies the footprints of DNA-binding proteins on the chromatin, which
56 revealed distinct binding configurations of transcription factors in yeast (Skene and
57 Henikoff 2017) and mammalian cells (Meers et al. 2019b).

58 In this study, we performed CUT&RUN on the human lung carcinoma cell line A549
59 maintained in a multi-well cell culture plate (Fig. 1a) to profile RNA polymerase II (Pol
60 II) around the transcription start sites (TSS). We found that CUT&RUN could validate
61 the bimodal signal around the TSS as previously detected using ChIP (Erickson et al.
62 2018). However, we also identified a distinct peak at the TSS from short DNA fragments
63 released only by pAG-MNase, suggesting a transient positioning of RNA polymerase II
64 before promoter-proximal pausing, which has eluded detection by conventional ChIP in
65 the steady state of transcription (Erickson et al. 2018).

66 # Results and Discussion

67 **Overall performance of "CUT&RUN on plate" for profiling Pol II**

68 First, we tested the feasibility of performing CUT&RUN on adherent cells attached to a
69 multi-well polystyrene cell culture plate with Pol II antibodies (Fig. 1a). A549 cells were
70 efficiently permeabilised on the plate with 0.1% Triton X-100 and remained attached
71 throughout the protocol with gentle manipulation. The background cutting was
72 estimated with two irrelevant antibodies (anti-influenza A PA and anti-influenza A PB1).
73 The resulting signals were inspected in the genomic region chr6 p22.2 (Fig. 1b). This
74 location is ideal for efficient visual inspection of Pol II signals because of the compact
75 size and close spacing of multiple histone genes. Our Pol II signals consistently mapped
76 to the transcription units of histone genes with low background as previously reported
77 (Meers et al. 2019a; Skene and Henikoff 2017) (Fig. 1b). The signals within 1.5 kb
78 around the TSS were averaged over "singletons" (genes that are separated from other
79 genes by >2 kb) (Fig. 1c), showing consistent enrichment around the TSS. Therefore,
80 through the successful capture of Pol II signals *in situ* with low background, the
81 performance of our "CUT&RUN on plate" (or CROP) protocol was shown to be
82 comparable with the conventional CUT&RUN (Meers et al. 2019a; Skene and Henikoff
83 2017).

84

85 **CUT&RUN fragments of different length reveal distinct Pol II footprints**

86 The size distribution of aligned paired-end reads showed multiple peaks (Fig. 2a and
87 Online Resource 1). We categorised the aligned reads into four subclasses by length
88 (Fig. 2a), each of which is expected to reflect a unique footprint of the protein-DNA

interaction (Meers et al. 2019b; Skene and Henikoff 2017). Indeed, we found that the fragments from different size classes show distinct localisations of Pol II around the TSS (Fig. 2b and Online Resource 2). The signal of the longer reads (>270 bp) conformed to the characteristic bimodal peak as previously reported by ChIP (i.e. the peaks located at either ~150 bp upstream or downstream of the TSS) (Erickson et al. 2018). By contrast, the shorter fragments (<120 bp) constituted a single narrow peak at the TSS (Fig. 2b). The signals of the shorter (<120 bp) and longer (>270 bp) fragments were mutually exclusive, showing two distinct footprints of Pol II adjacent to each other.

Recent studies using ChIP-seq, as well as nascent transcript-sequencing, have shown that Pol II is paused at ~50 bp downstream of the TSS (Core et al. 2008; Erickson et al. 2018; Kwak et al. 2013; Nojima et al. 2015; Shao and Zeitlinger 2017). Since the duration of Pol II in the promoter-proximal pausing state is about ten times longer than that of Pol II in the pre-initiation state (Darzacq et al. 2007; Steurer et al. 2018), a majority of reads obtained by ChIP is from the paused site, approximately 50 bp downstream of the TSS. Because ChIP fragments are typically hundreds of base pairs long, the rare signal of the poised Pol II at the TSS is often obscured. While Erickson *et al.* had previously identified the signal of the poised Pol II at the TSS by ChIP, this signal was only exposed following the depletion of Pol II in hypotonic condition with the use of triptolide to block the transcription initiation (Erickson et al. 2018). In contrast, we accurately identified the unique positioning of Pol II at the TSS in the physiological condition through the recovery of its unique footprint.

113 **Skewed positioning of the large-size-class Pol II footprints is associated with the**
114 **directionality of transcription**

115 As expected, the bimodal signal of the longer fragments (>270 bp) is skewed towards the
116 region downstream of the TSS (Fig. 2b) (Erickson et al. 2018); when genes were inspected
117 individually, however, we observed a variation in the skewness of large-size-class Pol II
118 footprints around the TSS (Fig. 2c). This observation prompted us to stratify genes by
119 the pattern of the large-size-class Pol II footprints.

120 We observed slight deviation of the short-fragment peak from the annotated
121 transcription start site in some genes (e.g. ALKBH5 in Fig. 2c), possibly due to the
122 alternative transcription start site. Therefore, we quantified the read coverage of the
123 large-size-class fragments in 500 bp upstream or downstream from the short-fragment
124 peak for each gene (Fig. 3a). We selected genes with the short-fragment peak above the
125 background and identified the peak position by local polynomial fitting (Fig. 3b). Genes
126 which had the peak within 500 bp from the annotated transcription start site were
127 included in this analysis (Fig. 3b). The signals were quantified for each of the two size
128 classes (120-270 bp and 270-440 bp), and the skewness of the footprint is expressed by
129 the ratio of the downstream signal to the upstream signal (Fig. 3a). The two size classes
130 showed similar skewness for each gene (Fig. 3c and Online Resource 3). About 14% of
131 the genes were positively skewed (skewness ≥ 1 for both two size classes) and 8%
132 negatively skewed (skewness ≤ -1 for both two size classes) (Fig. 3c and Online Resource
133 3).

134 We identified recurrent genes in which the large-size-class signals were either positively
135 or negatively skewed (e.g. RBM6, SLC35B1, CCDC6 and HNRMPA0) (Fig. 3c and Online
136 Resource 3). We anticipated that the skewed distribution of large-size-class reads

137 correlates with the directionality of transcription from the promoter. To test this
138 hypothesis, we turned to previously published mNET-seq data (Nojima et al. 2015)
139 (Online Resource 4). The mNET-seq reads were aligned around the TSS of these genes
140 (Fig. 3d). We observed that the genes were almost exclusively transcribed towards
141 downstream (i.e. the same direction as the gene) from the promoter when the large-size-
142 class Pol II footprints were positively skewed. On the other hand, bidirectional
143 transcription was observed from the promoter of genes with negatively skewed Pol II
144 footprints (Fig. 3d). These observations were confirmed by clustering genes by the
145 skewness of the large-size-class signals and visualising the nascent transcription in
146 these genes (Fig. 3e and Online Resource 5). In genes with positively skewed signals,
147 sense transcription from the promoter (i.e. the same direction as the gene) was
148 dominant. On the other hand, in genes with negatively skewed signals, abundant
149 antisense transcription from the promoter was observed (Fig. 3e and Online Resource
150 5). Therefore, we conclude that the negatively skewed positioning of the large-size-class
151 Pol II footprints is associated with the bidirectional transcription from the gene promoter
152 (Nojima et al. 2015; Preker et al. 2008).

153

154 **Impact of formaldehyde fixation on "CUT&RUN on plate"**

155 Formaldehyde fixation is often used to freeze transient protein interactions in
156 biochemical assays. We tested if formaldehyde fixation can be implemented in
157 "CUT&RUN on plate". Cells were fixed with 1.5% formaldehyde before performing
158 CUT&RUN. The resulting Pol II signals were mapped on the HOXA gene cluster region
159 alongside the signals from unfixed cells (Fig. 4a). The formaldehyde fixation did not
160 significantly affect the background cutting (Fig. 4b). However, the Pol II signal at the

161 TSS (i.e. the maximum height of the Pol II peak) was ubiquitously reduced to some 10%
162 to 30% of that in the "Unfixed" assays (Fig. 4c). We wish to see whether the observed
163 signal reduction was due to the formaldehyde fixation or within the range of
164 experimental variation. To this end, we compared CUT&RUN signals from eight
165 independent assays (four "Unfixed" and four "Fixed" assays) (Fig. 4d). Difference in the
166 signal intensity between unfixed and fixed samples were always greater than that
167 between any two assays under the same condition (experimental variation) (Fig. 4e, cf.
168 red and black symbols). Mean signal ratio between "Fixed" and "Unfixed" assays
169 ($\log_{10}0.7$, n = 8) (Fig. 4e) indicates that the signal was reduced to about 20% on average
170 in formaldehyde-fixed cells. It is possible that the formaldehyde fixation reduces the
171 efficiency of chromatin cutting by pAG-MNase and/or affects the release of cleaved DNA
172 fragments.

173

174 **Summary**

175 In this study, we have devised "CUT&RUN on plate" (CROP) for profiling protein-DNA
176 interactions in adherent cells maintained in a multi-well cell culture plate. The current
177 CUT&RUN protocol (Hainer and Fazzio 2019; Meers et al. 2019a; Skene et al. 2018)
178 relies on Concanavalin A-coated magnetic beads to immobilise the cells before the
179 targeted pAG-MNase cleavage. This immobilisation step introduces a range of
180 shortcomings associated with the use of Concanavalin A-beads. For example,
181 trypsinisation while harvesting the cells may alter the behaviour of DNA-binding
182 proteins and lower the binding affinity of cells to Concanavalin A. Concanavalin A also
183 stimulates certain types of cells. When the plasma membrane is removed to investigate
184 the nuclei, Concanavalin A does not bind efficiently to the nuclear membrane. Finally,

185 Concanavalin A-beads tend to aggregate when cells are attached. Our CROP protocol
186 described here overcomes these key problems associated with the use of Concanavalin-
187 A beads and may find widespread application. A fundamental limitation of the CROP
188 method is that it is only applicable to adherent cells that are stably attached to the
189 polystyrene surface of the culture plate throughout the protocol. In addition, it is crucial
190 to seed the optimal number of cells, as well as to inspect the confluency of cells before
191 an assay, to ensure that the cell count is consistent between assays for spike-in
192 normalisation.

193 Using CUT&RUN, we profiled RNA polymerase II near the transcription start sites. Our
194 observations suggest two distinct binding configurations of RNA polymerase II near the
195 TSS. These configurations are: (1) RNA polymerase II at the TSS in the pre-initiation
196 state where nucleosomes are depleted, and (2) paused RNA polymerase II upstream or
197 downstream of the TSS where Pol II pausing is regulated by a higher-molecular complex
198 (e.g. NELF) after transcription initiation (Fig. 5). The short-fragment (<120 bp) peak at
199 the TSS were also observed for Serine 5-phosphyrylated Pol II (Fig. 2b), raising the
200 possibility that some Pol II molecules detected by these short fragments are already
201 released from the poised site, yet are unable to extend the transcription, which is known
202 as abortive initiation (Goldman et al. 2009) The paused Pol II peaks were visible
203 exclusively in the longer (>120 bp) fragments adjacent to the TSS. We reason that this
204 is because Pol II is associated with a higher-molecular complex at the paused site (Vos
205 et al. 2018), which protects the chromatin from the pAG-MNase cleavage. As a result,
206 pAG-MNase is unable to produce shorter (<120 bp) fragments associated with the
207 paused Pol II. We showed that the positioning of these large-size-class footprints was
208 associated with the directionality of transcription from the gene promoter,
209 demonstrating the biological significance of these CUT&RUN footprints.

210 # Materials and Methods

211 **Cells**

212 Human lung carcinoma cell line A549 was maintained in Dulbecco's Modified Eagle
213 Medium (Gibco, 11965-092) supplemented with 10% FBS (Gibco, 16140-071), 100 U/ml
214 penicillin and 100 μg/ml streptomycin (Gibco, 15140-122).

215

216 **CUT&RUN in a cell culture plate**

217 CUT&RUN (Skene and Henikoff 2017) was performed on a standard 24-well cell culture
218 plate (TPP, 92012); cells were left attached to the bottom of the well throughout the
219 protocol.

220 A549 was seeded in wells and grown to ~70 to 90% confluency (~0.5 million cells) at the
221 time of an assay. Cells were washed with PBS and permeabilised with Perm buffer (0.1%
222 Triton X-100 (Sigma, X100), 20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.5 mM
223 Spermidine (Sigma, S2626) and proteinase inhibitor (Roche, 04 693 132 001)) for 15
224 min at room temperature. After removing Perm buffer, cells were washed again with
225 Perm buffer and incubated with an antibody (~0.7 μg) in 150 μl Perm buffer per well for
226 one hour at room temperature. The antibodies used in this study were: mouse anti-
227 RNA polymerase II (Millipore, 05-623; clone CTD4H8; 0.75 μg per well); rabbit anti-RNA
228 polymerase II (phospho S5) (Abcam, ab5131; polyclonal; 0.675 μg per well); rabbit anti-
229 influenza A PA (Invitrogen, PA-32223; polyclonal; 1.1 μg per well); and rabbit anti-
230 influenza A PB1 (Invitrogen, PA-34914; polyclonal; 0.86 μg per well). After incubation,
231 unbound antibody was removed, and cells were washed twice with Perm buffer. Cells
232 were then incubated with the recombinant pAG-MNase (Cell Signaling, 40366; 1:33

233 volume) in 150 µl Perm buffer per well for an hour at room temperature. Unbound pAG-
234 MNase was removed, and cells were washed twice with Perm buffer. During the second
235 wash of this stage, the plate was placed on ice-cold water. For pAG-MNase activation,
236 150 µl ice-cold Perm buffer containing 5 mM $CaCl_2$ was dispensed per well, and the plate
237 was incubated on ice-cold water for 30 min. The pAG-MNase digestion was halted by
238 the addition of 50 µl 4× STOP solution (680 mM NaCl, 40 mM EDTA (Sigma, E5134),
239 8mM EGTA (AG Scientific, E-2491), 100 µg/ml Rnase A (Invitrogen, 12091021) and
240 0.1% Triton X-100). The plate was incubated at 37°C for 30 min to facilitate the release
241 of digested DNA fragments from the nucleus. The supernatant was collected, and DNA
242 was extracted using Dneasy Blood & Tissue Kit (Qiagen, 69504).

243 In addition to the native "CUT&RUN on plate" described above, two protocols for cell
244 fixation with 1.5% formaldehyde were tested: (1) cells were fixed with 1.5% formaldehyde
245 for 10 min prior to the cell permeabilisation with Perm buffer; and (2) cells were fixed
246 following the permeabilisation (5 min in Perm buffer). Both protocols produced similar
247 results. The supernatant collected at the end of the CUT&RUN protocol was
248 supplemented with 1% SDS (final concentration) and incubated at 65°C overnight for
249 de-crosslinking. The DNA was purified with QIAquick PCR Purification Kit (Qiagen,
250 28104).

251

252 **Pair-end sequencing and read alignment**

253 DNA libraries were constructed using NEBNext Ultra II DNA Library Prep Kit for Illumina
254 (NEB, E7645) with index primers (NEB, E7335, E7710 and E7600). Fragmented
255 genomic DNA from *S. cerevisiae* (Cell Signaling, 40366) was used (10 pg per sample) as
256 spike-in.

257 DNA libraries were sequenced (150 bp × 2) with NovaSeq 6000 (Illumina), and the initial
258 50 bases of each read were aligned to a concatenation of human reference genome
259 (GRCh38) and *S. cerevisiae* (sacCer3) with a BWA algorithm (version 0.7.17) `bwa mem`
260 (Li and Durbin 2009). Duplications of reads were marked with a Picard tool (version
261 2.18.9) `MarkDuplicates` (http://broadinstitute.github.io/picard). Only a minority of
262 the aligned reads constituted the yeast genome, thereby confirming the successful
263 recovery of DNA fragments cleaved by pAG-MNase. A subset of aligned reads was
264 selected for the downstream analyses using Samtools (version 1.9) with the following
265 commands: SAM flag 3 (read paired; read mapped in proper pair) (`samtools view -f`
266 `3`); mapping quality greater than 20 (`samtools view -q 20`); without tags "XA:Z:" and
267 "SA:Z:" (`grep -v -e 'XA:Z:' -e 'SA:Z:'`); and aligned in either one of the
268 chromosomes chr1 to chr22, chrX or chrY (`samtools view chr1 chr2 ... chrY`).

269

270 **Data analyses and visualisation**

271 Aligned reads were fractionated by size (the 9th field of SAM) and stored in the BAM
272 format. Read coverage was computed and stored in the bigWig format with deepTools's
273 (version 3.1.3) `bamCoverage` (Ramirez et al. 2016), normalised per 10,000 spike-in
274 reads. The spike-in-normalised bigWig files were converted into the bedGraph format
275 using `bigWigToBedGraph` (UCSC) and the read coverage was visualised with the R
276 package Sushi (version 1.24.0) (Phanstiel et al. 2014) in R (version 3.6.2). A matrix
277 which stores the read coverage over a given region of genes was computed using
278 deepTools's (version 3.3.2) `computeMatrix` with an annotation
279 GCA_000001405.15_GRCh38_full_analysis_set.refseq_annotation (NCBI). A subset of
280 genes that were separated from other genes by >2 kb was considered for this analysis;

281 such genes were extracted using bedtools's (version 2.26.0) (Quinlan and Hall 2010)
282 `merge` and `intersect`. The matrix was visualised as a heatmap and a metaplot using
283 the R package EnrichedHeatmap (version 1.16.0) (Gu et al. 2018).

284

285 **External mNET-seq data**

286 Publicly available mNET-seq datasets (GSM1474225, GSM1474226, GSM1474228 and
287 GSM1480181) were obtained from Gene Expression Omnibus (GSE60358) (Nojima et
288 al. 2015). The original bigWig files were converted to the bedGraph format using
289 bigWigToBedGraph (UCSC), and then the genome coordinates were converted to hg38
290 using liftOver (UCSC).

291

292 **Signal quantification and statistical tests**

293 Pol II signal intensity were quantified by the maximum height of the spike-in-normalised
294 read coverage within 1.5 kb around the annotated transcription start site (Fig. 3b and
295 Figs. 4b-d). Local polynomial fitting to the short-fragment peak was performed using
296 the R function `loess` (`span = 0.05`) (Fig. 3b). The signal in the upstream or downstream
297 of the short-fragment peak (Fig. 3a) was quantified by the summation of signals in the
298 500 bp window stored in the matrix file from deepTools's `computeMatrix` (Figs. 3a and
299 c). Linear regression was performed using the R function `lm` to fit a straight line to the
300 data points and estimate the parameters (Fig. 4c). Mean difference of two populations
301 was calculated with the R function `wilcox.test` (Figs. 4d and e).

302

Data reproducibility

The assays performed and analysed in this work are listed in Table 1. The source data for figure production are listed in Table 2.

Table 1 Assays performed in this study and data availability

Assay ID	Antibody [a]	Fixation
5-200331	Total Pol II	Unfixed
1-200518	Total Pol II	Unfixed
2-200518	Total Pol II	Fixed [b]
3-200518	Total Pol II	Fixed [c]
7-200331	S5p Pol II	Unfixed
7-200518	S5p Pol II	Unfixed
8-200518	S5p Pol II	Fixed [b]
9-200518	S5p Pol II	Fixed [c]
1-200820	Total Pol II	Unfixed
3-200820	S5p Pol II	Unfixed
4-200820	Flu PA	Unfixed
5-200820	Flu PB1	Unfixed
1-1-200710	Total Pol II	Fixed [c]
1-3-200710	S5p Pol II	Fixed [c]
1-4-200710	Flu PA	Fixed [c]
1-5-200710	Flu PB1	Fixed [c]

[a] Total Pol II, Millipore 05-623; S5p Pol II, Abcam ab5131
[b] Fixed after cell permeabilisation
[c] Fixed before cell permeabilisation

Table 2 Source data for figure production

Figure	Source (Assay ID)
Figs. 1b and c	Total Pol II: 1-200820 S5p Pol II: 3-200820 Flu PA (negative control): 4-200820 Flu PB1 (negative control): 5-200820
Fig. 2a	Total Pol II: 5-200331 S5p Pol II: 7-200518
Online Resource 1 (Supplementary to Fig. 2a)	Total Pol II: 1-200518 S5p Pol II: 7-200331
Fig. 2b	Total Pol II: 5-200331 S5p Pol II: 7-200518
Online Resource 2 (Supplementary to Fig. 2b)	Total Pol II: 1-200518 S5p Pol II: 7-200331
Fig. 2c	Total Pol II: 5-200331 (Assay 1) and 1-200518 (Assay 2) S5p Pol II: 7-200331 (Assay 1) and 7-200518 (Assay 2)
Figs. 3b and c	5-200331
Online Resource 3 (Supplementary to Fig. 3c)	Total Pol II: 1-200518 S5p Pol II (left): 7-200331 (left) and 7-200518 (right)
Fig. 3d	Total Pol II: 5-200331 (top) and 1-200518 (bottom) S5p Pol II: 7-200331 (top) and 7-200518 (bottom)
Fig. 3e	5-200331
Online Resource 5 (Supplementary to Fig. 3e)	Total Pol II (1-200518) S5p Pol II: 7-200331 (top) and 7-200518 (bottom)
Figs. 4a-c	Total Pol II: 1-200820 (Unfixed) and 1-1-200710 (Fixed) S5p Pol II: 3-200820 (Unfixed) and 1-3-200710 (Fixed) Flu PA: 4-200820 (Unfixed) and 1-4-200710 (Fixed) Flu PB1: 5-200820 (Unfixed) and 1-5-200710 (Fixed)
Figs. 4d and e	Total Pol II (Unfixed): 5-200331 and 1-200518 Total Pol II (Fixed): 3-200518 and 2-200518 S5p Pol II (Unfixed) :7-200331 and 7-200518 S5p Pol II (Fixed): 9-200518 and 8-200518

Acknowledgements

We thank Liew Jun Mun, University of Cambridge, for reviewing the manuscript to enhance the readability. We also thank Prof. Charles R. M. Bangham, Imperial College London, for reviewing the manuscript and critical comments. Illumina NovaSeq 6000 sequencing was performed at Centre for PanorOmic Sciences (CPOS) Genomics Core, LKS Faculty of Medicine, The University of Hong Kong. This study was supported in part by the Research Grants Council of the Hong Kong SAR (17119618, 17107019 and C7027-16G).

Declarations

Funding

This study was supported in part by the Research Grants Council of the Hong Kong SAR (17119618, 17107019 and C7027-16G).

Competing interests

The authors declare no competing interests.

Ethics approval

Not applicable

Consent for publication

Not applicable

Authors' contributions

MM: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Visualization, Writing - original draft preparation, Writing - review and editing

HC: Resources, Project administration, Funding acquisition, Writing - review and editing

334 **References**

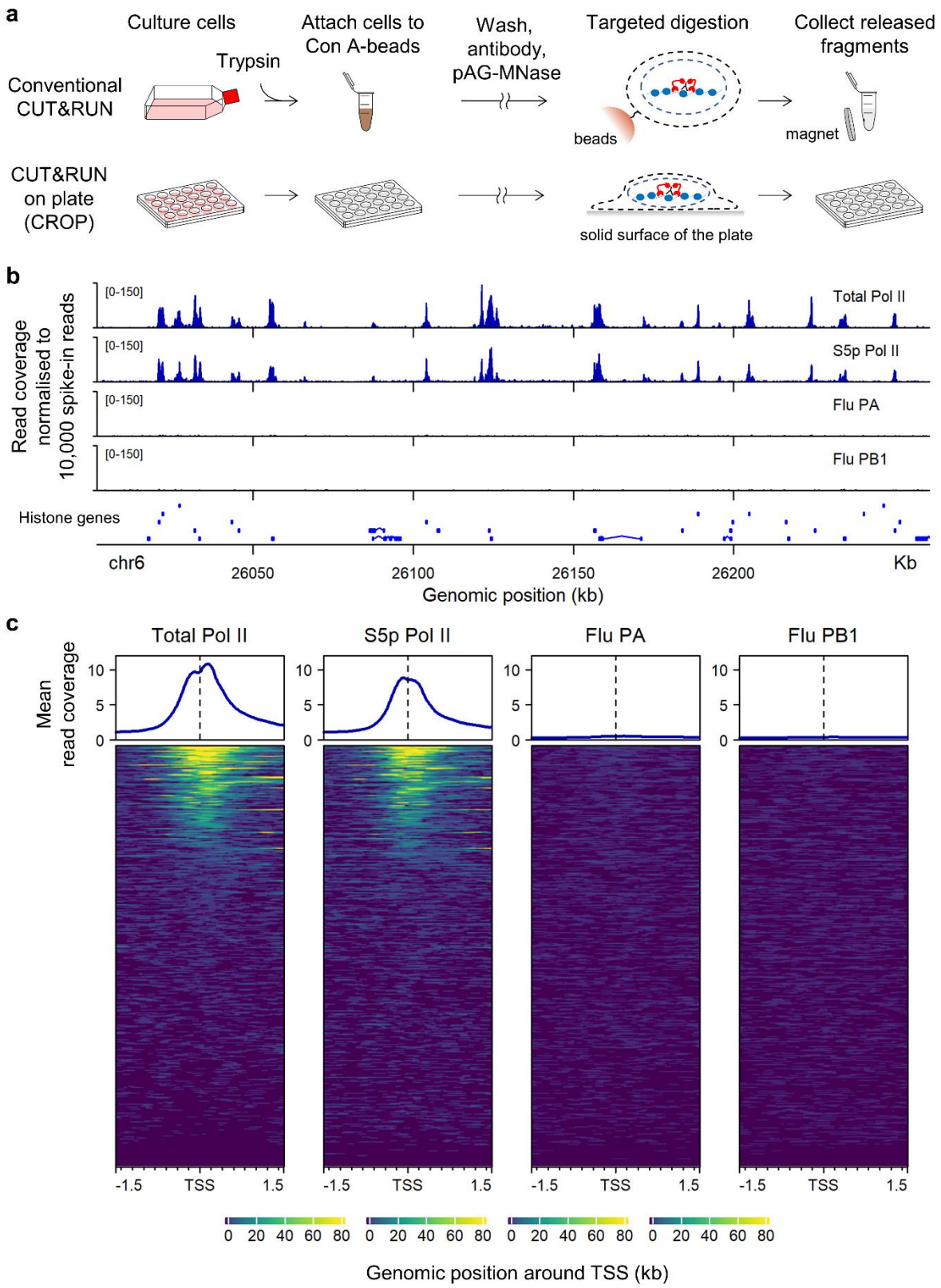
335 Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread
336 pausing and divergent initiation at human promoters Science 322:1845-1848
337 doi:10.1126/science.1162228
338 Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH (2007)
339 In vivo dynamics of RNA polymerase II transcription Nat Struct Mol Biol 14:796-
340 806 doi:10.1038/nsmb1280
341 Erickson B, Sheridan RM, Cortazar M, Bentley DL (2018) Dynamic turnover of paused
342 Pol II complexes at human promoters Genes Dev 32:1215-1225
343 doi:10.1101/gad.316810.118
344 Goldman SR, Ebright RH, Nickels BE (2009) Direct detection of abortive RNA transcripts
345 in vivo Science 324:927-928 doi:10.1126/science.1169237
346 Gu Z, Eils R, Schlesner M, Ishaque N (2018) EnrichedHeatmap: an R/Bioconductor
347 package for comprehensive visualization of genomic signal associations BMC
348 Genomics 19:234 doi:10.1186/s12864-018-4625-x
349 Hainer SJ, Fazzio TG (2019) High-Resolution Chromatin Profiling Using CUT&RUN Curr
350 Protoc Mol Biol 126:e85 doi:10.1002/cpmb.85
351 Kwak H, Fuda NJ, Core LJ, Lis JT (2013) Precise maps of RNA polymerase reveal how
352 promoters direct initiation and pausing Science 339:950-953
353 doi:10.1126/science.1229386
354 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler
355 transform Bioinformatics 25:1754-1760 doi:10.1093/bioinformatics/btp324
356 Marx V (2019) What to do about those immunoprecipitation blues Nat Methods 16:289-
357 292 doi:10.1038/s41592-019-0365-3

358 Meers MP, Bryson TD, Henikoff JG, Henikoff S (2019a) Improved CUT&RUN chromatin
359 profiling tools Elife 8 doi:10.7554/eLife.46314
360 Meers MP, Janssens DH, Henikoff S (2019b) Pioneer Factor-Nucleosome Binding Events
361 during Differentiation Are Motif Encoded Mol Cell 75:562-575 e565
362 doi:10.1016/j.molcel.2019.05.025
363 Nojima T et al. (2015) Mammalian NET-Seq Reveals Genome-wide Nascent Transcription
364 Coupled to RNA Processing Cell 161:526-540 doi:10.1016/j.cell.2015.03.027
365 Phanstiel DH, Boyle AP, Araya CL, Snyder MP (2014) Sushi.R: flexible, quantitative and
366 integrative genomic visualizations for publication-quality multi-panel figures
367 Bioinformatics 30:2808-2810 doi:10.1093/bioinformatics/btu379
368 Preker P et al. (2008) RNA exosome depletion reveals transcription upstream of active
369 human promoters Science 322:1851-1854 doi:10.1126/science.1164096
370 Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic
371 features Bioinformatics 26:841-842 doi:10.1093/bioinformatics/btq033
372 Ramirez F et al. (2016) deepTools2: a next generation web server for deep-sequencing
373 data analysis Nucleic Acids Res 44:W160-165 doi:10.1093/nar/gkw257
374 Shao W, Zeitlinger J (2017) Paused RNA polymerase II inhibits new transcriptional
375 initiation Nat Genet 49:1045-1051 doi:10.1038/ng.3867
376 Skene PJ, Henikoff JG, Henikoff S (2018) Targeted in situ genome-wide profiling with
377 high efficiency for low cell numbers Nat Protoc 13:1006-1019
378 doi:10.1038/nprot.2018.015
379 Skene PJ, Henikoff S (2017) An efficient targeted nuclease strategy for high-resolution
380 mapping of DNA binding sites Elife 6 doi:10.7554/eLife.21856

381 Steurer B et al. (2018) Live-cell analysis of endogenous GFP-RPB1 uncovers rapid
382 turnover of initiating and promoter-paused RNA Polymerase II Proc Natl Acad Sci
383 U S A 115:E4368-E4376 doi:10.1073/pnas.1717920115
384 Vos SM, Farnung L, Urlaub H, Cramer P (2018) Structure of paused transcription
385 complex Pol II-DSIF-NELF Nature 560:601-606 doi:10.1038/s41586-018-0442-
386 2

387

Fig. 1

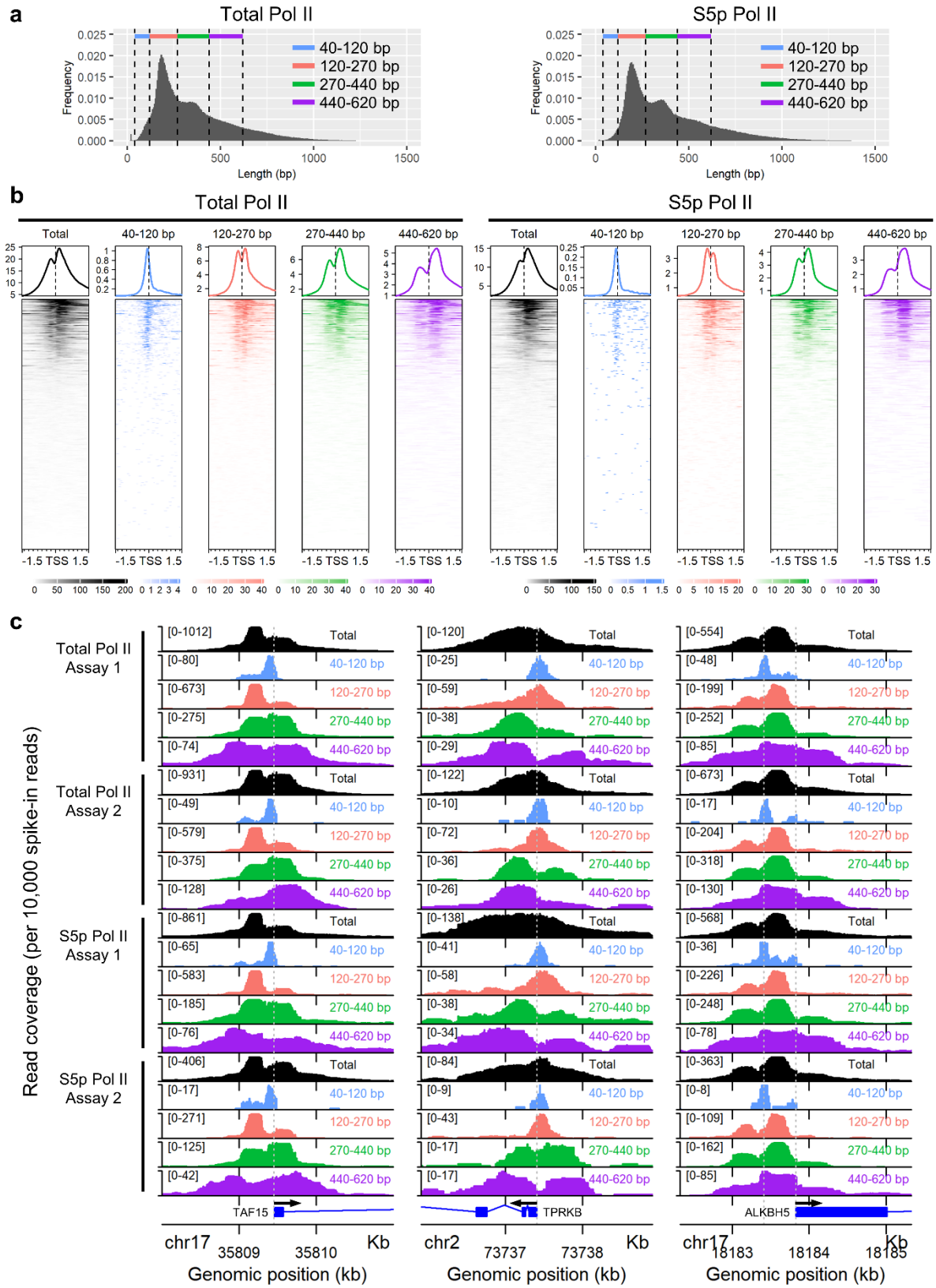

a
Culture cells
Attach cells to Con A-beads
Wash, antibody, pAG-MNase
Targeted digestion
Collect released fragments
Trypsin
Conventional CUT&RUN
beads
magnet
CUT&RUN on plate (CROP)
solid surface of the plate
b
Read coverage normalised to 10,000 spike-in reads
[0-150]
Total Pol II
S5p Pol II
Flu PA
Flu PB1
Histone genes
chr6
26050
26100
26150
26200
Kb
Genomic position (kb)
c
Mean read coverage
Total Pol II
S5p Pol II
Flu PA
Flu PB1
-1.5
TSS
1.5
0 20 40 60 80
Genomic position around TSS (kb)

388

389 **Fig. 1 Overall performance of "CUT&RUN on plate"**

390 (a) Protocol overview. In conventional CUT&RUN (top row), cells are harvested and attached on
391 Concanavalin A-magnetic beads. Subsequent cell permeabilisation, antibody binding, and targeted pAG-
392 MNase digestion are performed on these cells immobilised on the beads. The magnetic beads are captured
393 each time the buffers are changed. The alternative approach, "CUT&RUN on plate" (or CROP) (bottom
394 row), circumvents the use of Concanavalin A-magnetic beads. Adherent cells are anchored on the solid
395 surface of the multi-well plate throughout the protocol, enabling rapid buffer change for multiple assays.
396 (b) Pol II signals in the histone gene cluster obtained by CROP. Signals are expressed by the coverage of
397 total aligned reads normalised to 10,000 spike-in read counts. Total Pol II and Serine 5-phosphorylated
398 (S5p) Pol II signals were obtained using the antibodies against RNA polymerase II (Millipore 05-623 and
399 Abcam ab5131, respectively). Flu PA and Flu PB1 are negative controls in which two irrelevant antibodies
400 were used (anti-influenza A PA and anti-influenza A PB1, respectively). (c) Metaplot analysis of the signals
401 from total aligned reads. Spike-in-normalised read coverage was averaged over annotated RefSeq genes
402 that are at least 2 kb away from other genes (n = 13,452). The region within 1.5 kb around the annotated
403 transcription start site (TSS) is presented.

Fig. 2

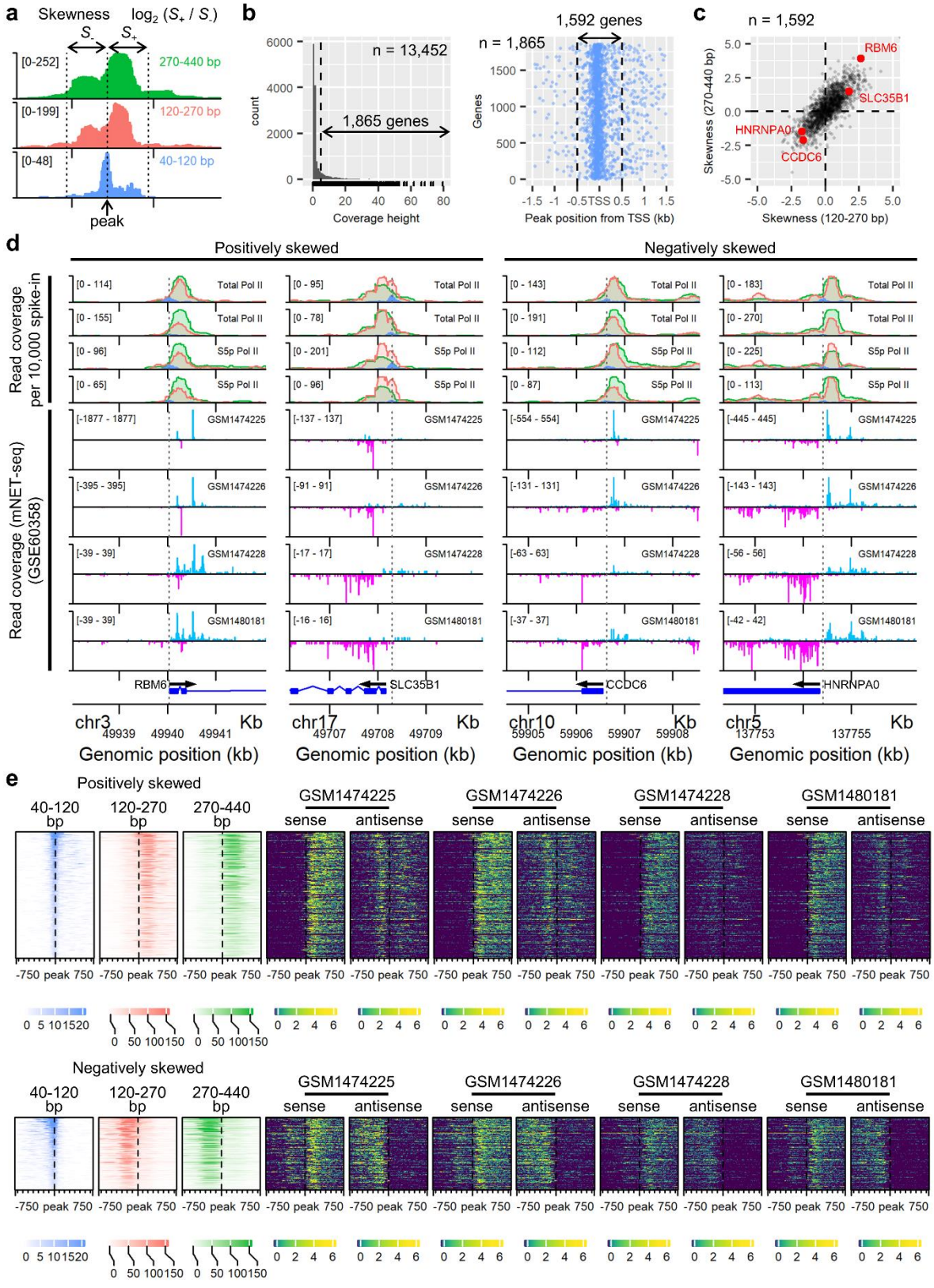

a
Total Pol II
S5p Pol II
Frequency
Length (bp)
40-120 bp
120-270 bp
270-440 bp
440-620 bp
b
Total Pol II
S5p Pol II
Total
40-120 bp
120-270 bp
270-440 bp
440-620 bp
-1.5 TSS 1.5
c
Total Pol II Assay 1
Total Pol II Assay 2
S5p Pol II Assay 1
S5p Pol II Assay 2
Read coverage (per 10,000 spike-in reads)
TAF15
TPRKB
ALKBH5
chr17
chr2
Kb
Genomic position (kb)

404

405 **Fig. 2 Size-fractionation identifies distinct Pol II footprints.**

406 (a) Size distribution of aligned reads (the insert size between the sequence adaptors). The reads were
407 categorised into four subclasses by size, each of which is coded by a unique colour. (b) Total Pol II and S5p
408 Pol II signals from each size class within 1.5 kb around the transcription start site (TSS). The signals in 13,452
409 annotated RefSeq singleton genes are presented in the heatmap. The metaplot above the heatmap
410 indicates the mean spike-in-normalised read coverage from these 13,452 genes. (c) Total Pol II and S5p Pol
411 II signals from each size classes in individual genes. A 3 kb window of the genomic region around the
412 transcription start site is presented. The vertical axis in each row shows the spike-in-normalised read
413 coverage. The scale is indicated in brackets. The same colour codes are used as per panel (a) to indicate
414 the size classes. The results of four independent assays (two for Total Pol II and the other two S5p Pol II)
415 are shown for each gene. The grey dashed line indicates the annotated transcription start site. An
416 additional line is shown in ALKBH5 to indicate the deviation of the short-fragment peak from the TSS.

Fig. 3

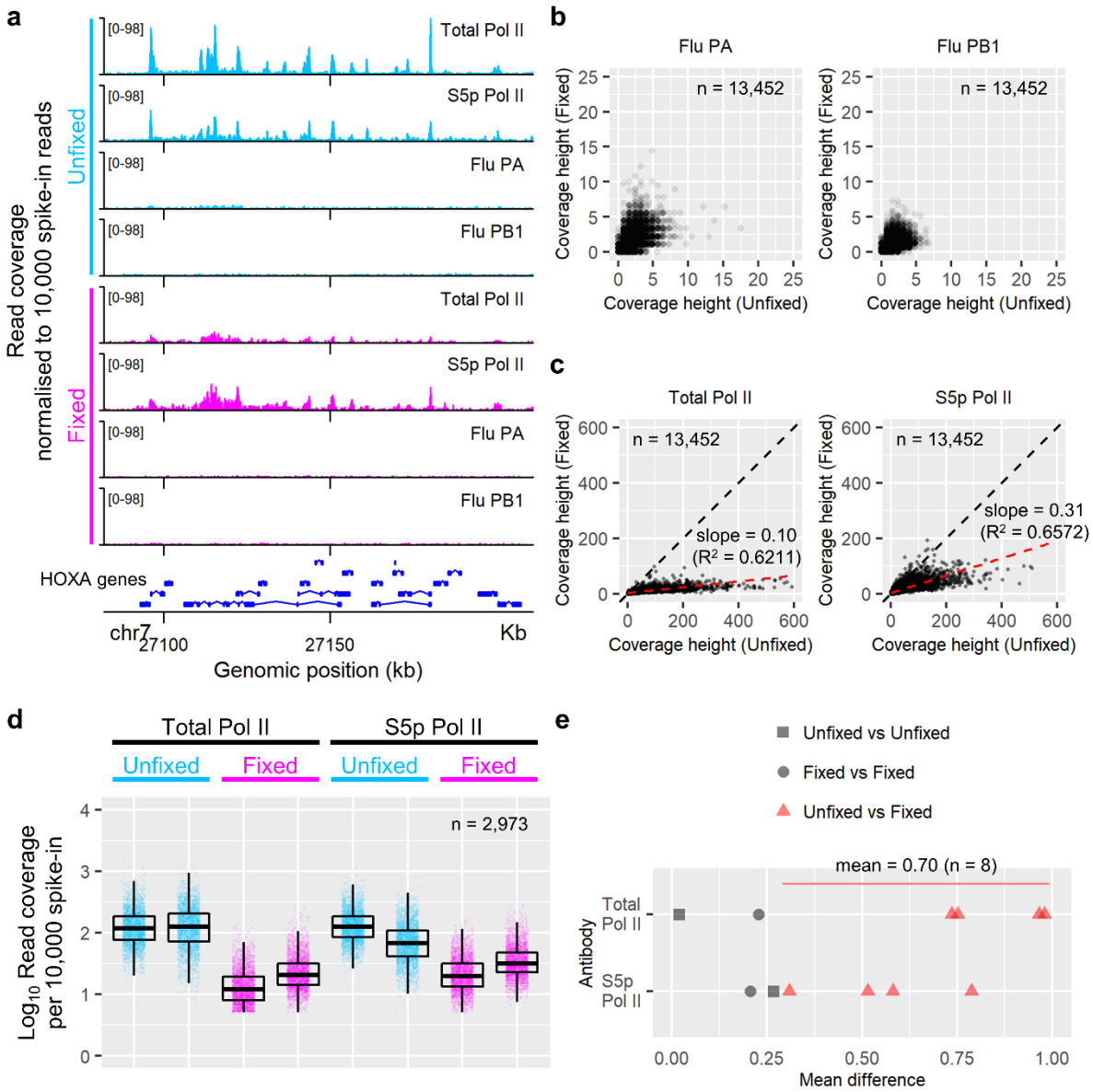

a
Skewness
log2 (S+ / S-)
S-
S+
[0-252]
270-440 bp
[0-199]
120-270 bp
[0-48]
40-120 bp
peak
b
n = 13,452
1,865 genes
count
Coverage height
n = 1,865
1,592 genes
Genes
Peak position from TSS (kb)
c
n = 1,592
RBM6
SLC35B1
HNRNPA0
CCDC6
Skewness (270-440 bp)
Skewness (120-270 bp)
d
Positively skewed
Negatively skewed
Read coverage per 10,000 spike-in
Read coverage (mNET-seq) (GSE60358)
Total Pol II
S5p Pol II
GSM1474225
GSM1474226
GSM1474228
GSM1480181
RBM6
SLC35B1
CCDC6
HNRNPA0
chr3
chr17
chr10
chr5
Kb
Genomic position (kb)
e
Positively skewed
Negatively skewed
40-120 bp
120-270 bp
270-440 bp
GSM1474225
GSM1474226
GSM1474228
GSM1480181
sense
antisense
-750 peak 750

417

418 **Fig. 3 Skewed distribution of the large-sized Pol II footprints around the TSS is**
419 **associated with the directionality of transcription.**

420 (a) Schematic diagram of quantifying the skewness of the Pol II signals at the TSS. The position of the short-
421 fragment peak was identified and then the painted area (i.e. the read coverage) in 500 bp upstream from
422 the peak (denoted as S-) and 500 bp downstream from the peak (S+) was calculated. Skewness is expressed
423 by the logarithmic ratio of S+ to S- in this study. (b) (left) Distribution of the hight of the short-fragment
424 coverage from the 13,452 singleton genes. Genes that had the short-fragment signal above the background
425 (dashed line) were selected to allow robust identification of the short-fragment peak (1,865 genes). (right)
426 The position of the short-fragment peak relative to the annotated transcription start site for each gene.
427 The vertical axis indicates the index of 1,865 genes. Genes that had the peak position within 500 bp around
428 the annotated transcription start site were further selected (1,592 genes). (c) Distribution of the skewness
429 from 1,592 genes. The skewness for the two size classes is plotted for each gene (120-270 bp on the x-axis
430 and 270-440 bp on the y-axis). Four genes that were recurrently identified as skewed, either positively or
431 negatively, are indicated in red. (d) DNA footprints of RNA polymerase II identified by CUT&RUN in this
432 study, and the nascent transcript-sequencing data from the public database. The alignment is shown near
433 the TSS of the four genes from panel (c). The first four rows in each column show CUT&RUN signals from
434 three size classes (40-120 bp, 120-270 bp and 270-440 bp). The same colour code is used as per Fig. 2a to
435 indicate the size classes. The signals were normalised to 10,000 spike-in reads. The scale of the vertical
436 axis in each panel is indicated in bracket. The second four rows show nascent RNA transcription from Gene
437 Expression Omnibus (GSE60358), which contains four mNET-seq assays. The signal of the plus-strand-
438 nascent transcripts is indicated in light blue, and the minus-strand-transcripts in magenta, regardless of the
439 gene orientation shown at the bottom of each column. The grey dashed line indicates the position of the
440 short-fragment peak. (e) Heatmaps showing the footprints of RNA polymerase II and the directionality of
441 transcription. The top row contains 227 genes that had the large CUT&RUN fragments directed towards
442 downstream of the short-fragment peak (the first three column, positively skewed), and the bottom row
443 contains 131 genes that had the large CUT&RUN fragments directed towards upstream of the short-
444 fragment peak (the first three column, negatively skewed). Each gene is oriented from left to right in the
445 heatmaps (i.e. genes encoded on the minus strand are all flipped horizontally). The heatmaps cover ±750
446 bp around the short-fragment peak of each gene. The fourth column onwards show the mNET-seq datasets
447 from GSE60358. Each of the four datasets contains two heatmaps showing the nascent transcription from
448 either the same strand (sense) or the opposite strand (antisense) of the gene.

Fig. 4

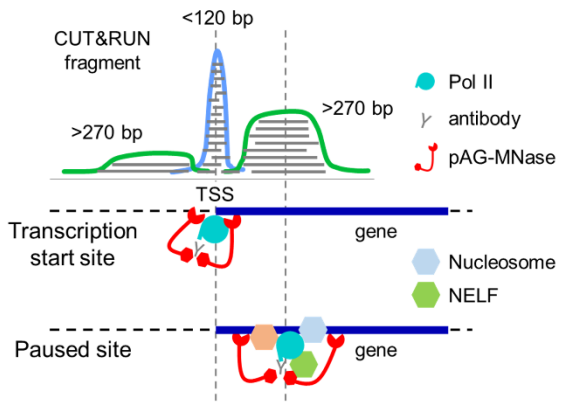

a
Read coverage normalised to 10,000 spike-in reads
Unfixed
Fixed
[0-98]
Total Pol II
S5p Pol II
Flu PA
Flu PB1
HOXA genes
chr7
27100
27150
Kb
Genomic position (kb)
b
Flu PA
Flu PB1
n = 13,452
Coverage height (Fixed)
Coverage height (Unfixed)
c
Total Pol II
S5p Pol II
slope = 0.10
(R² = 0.6211)
slope = 0.31
(R² = 0.6572)
d
Total Pol II
S5p Pol II
Unfixed
Fixed
n = 2,973
Log10 Read coverage per 10,000 spike-in
e
Unfixed vs Unfixed
Fixed vs Fixed
Unfixed vs Fixed
mean = 0.70 (n = 8)
Antibody
Total Pol II
S5p Pol II
Mean difference

449

450 **Fig. 4 Impact of formaldehyde cell fixation on "CUT&RUN on plate"**

451 (a) CUT&RUN signals of total fragments over the HOXA gene cluster (chr7 p15.2). Total Pol II, Millipore 05-
452 623; and S5p Pol II, Abcam ab5131. Flu PA and Flu PB1 are negative controls obtained with irrelevant
453 antibodies (influenza A PA and influenza A PB1, respectively). The vertical axis indicates the read coverage
454 normalised to 10,000 spike-in reads. The scale is shown in brackets. Two colour codes are used to indicate
455 whether cells were fixed with formaldehyde in the assays (light blue, "Unfixed"; and magenta, "Fixed"). (b)
456 Comparison of the background cutting between the "Fixed" and "Unfixed" assays (13,452 genes). The
457 maximum height of spike-in-normalised coverage within 1.5 kb around the TSS of each gene is shown. (c)
458 Comparison of the Pol II signal intensity between the "Fixed" and "Unfixed" assays (13,452 genes). The
459 maximum height of the spike-in-normalised Pol II peak (total reads) within 1.5 kb around the TSS is shown
460 for each gene. The black dashed line indicates the diagonal line. The red dashed line shows a linear
461 regression of these data points. The straight line was fitted to the data points and the parameters were
462 estimated. (d) Spike-in-normalised Pol II signal intensity (maximum height of the total read coverage)
463 around the TSS. The results of eight assays are presented (four for "Unfixed" and the other four "Fixed").
464 Genes with the signal above the background (peak height > 5) are shown (n = 2,973). The boxplot indicates
465 25%, 50% and 75% quantiles. (e) Mean difference of the Pol II intensity between two assays shown in panel
466 (d). Data points comparing "Fixed" and "Unfixed" assays are highlighted in red; the mean of these eight
467 datapoints (red triangles) is indicated.

Fig. 5

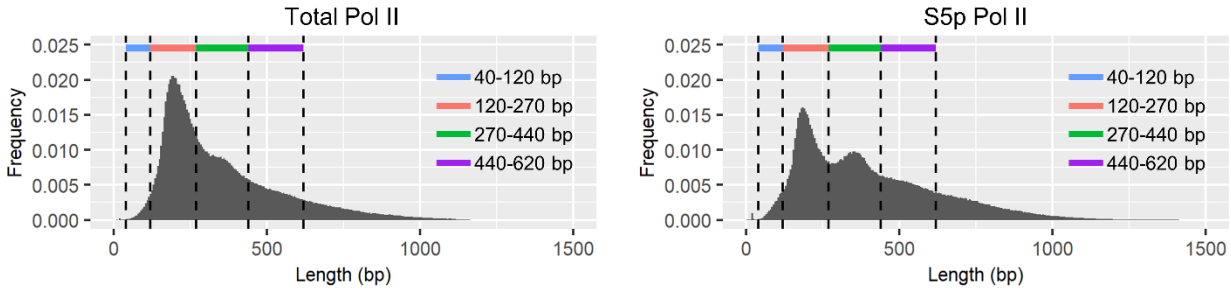


468

469 **Fig. 5 Working hypothesis on the Pol II footprints and Pol II pausing**

470 In this working hypothesis, the short-size-class Pol II footprint indicates the transient positioning of poised
471 Pol II, whereas the large-size-class Pol II footprint maps the promoter proximal pausing downstream of the
472 transcription start site.

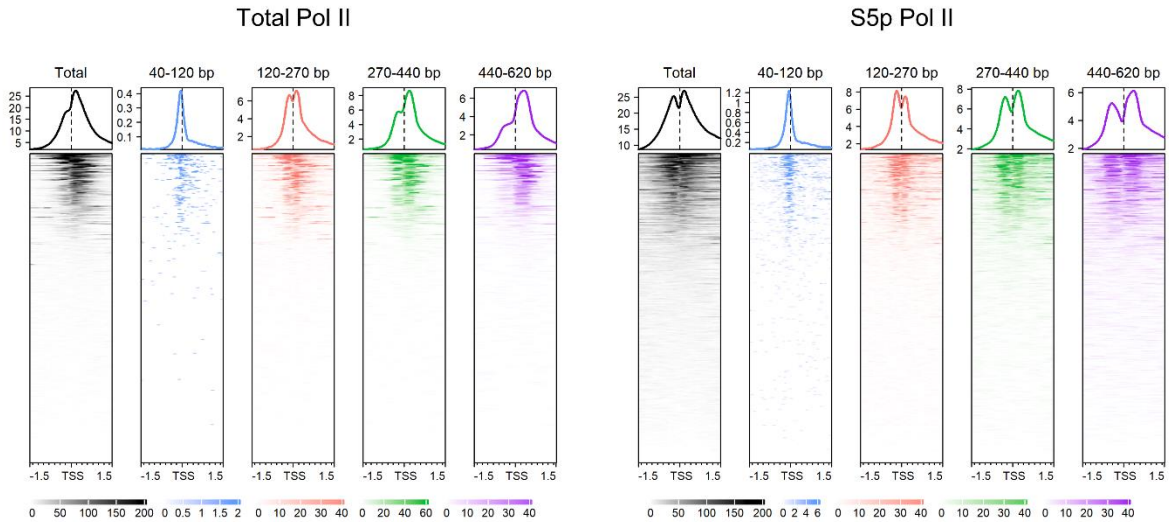
Supplementary information



Source data: 1-200518 (Total Pol II) and 7-200331 (S5p Pol II)

473 **Online Resource 1 Length distribution of aligned reads (Supplementary to Fig. 2a)**

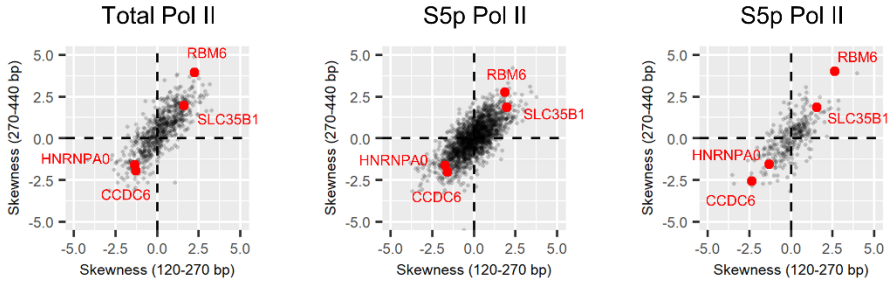
Supplementary information



Source data: 1-200518 (Total Pol II) and 7-200331 (S5p Pol II)

Online Resource 2 CUT&RUN fragments and Pol II footprints (Supplementary to Fig. 2b)

Supplementary information

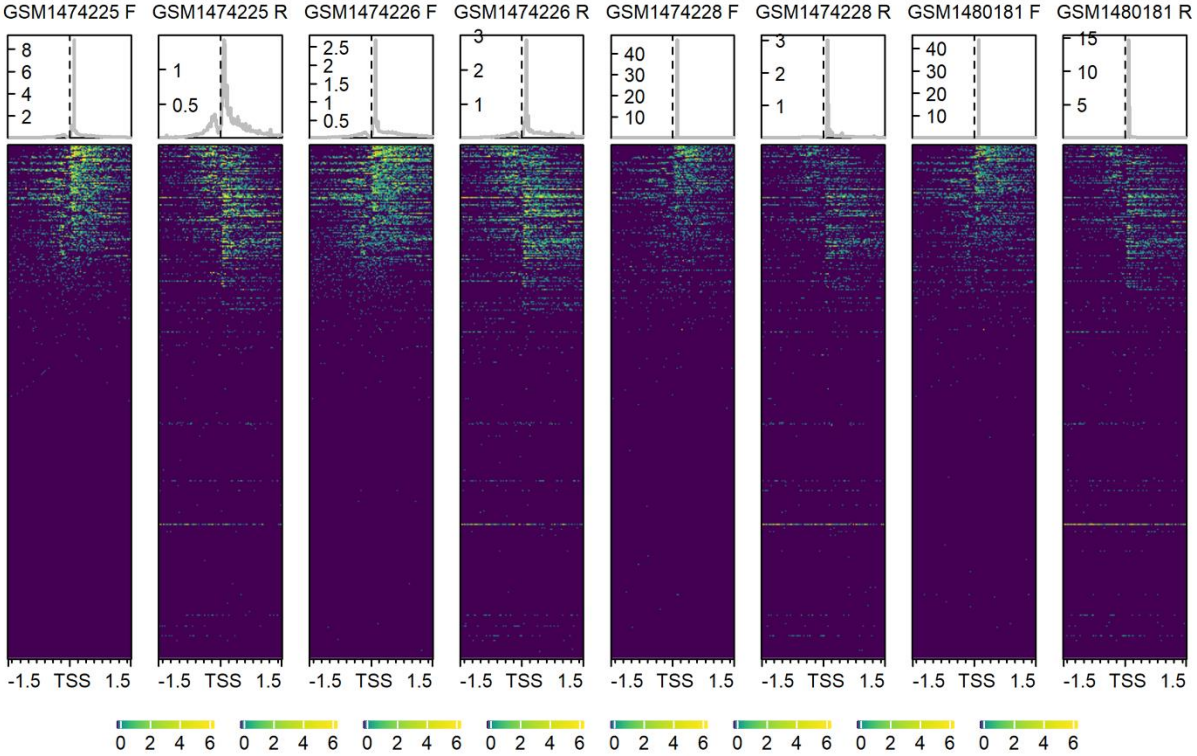


	n = 686	n = 1,899	n = 321
Positively skewed	21%	8%	7%
Negatively skewed	8%	9%	9%
Source data:	1-200518	7-200331	7-200518

Online Resource 3 Skewness of the large-size-class Pol II footprints (Supplementary to Fig. 3c)

475

Supplementary information

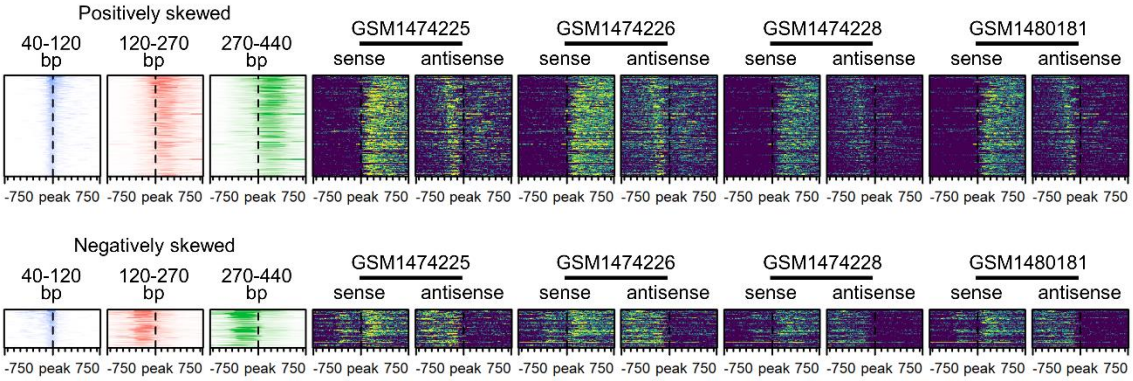


Online Resource 4 mNET-seq datasets used in this study

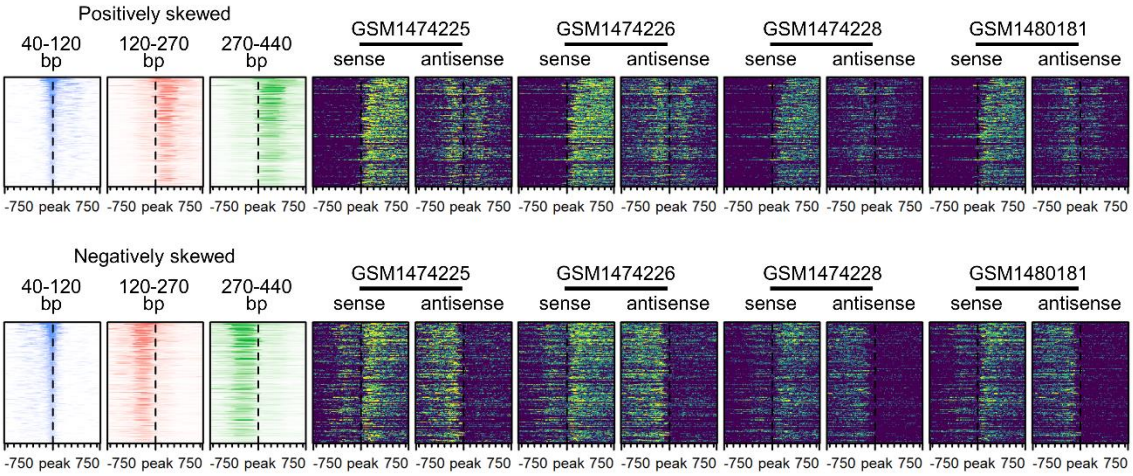
Publicly available mNET-seq data (GSE60358). The heatmap shows the nascent transcription on the plus (F) or the minus (R) strand within ±1.5 kb around the annotated transcription start site of each gene. The original data was lifted to hg38. The axis at the bottom indicates the genomic position (kb) around the transcription start site. RefSeq genes that are separated from other genes by >2 kb are presented (n = 13,452).

Supplementary information

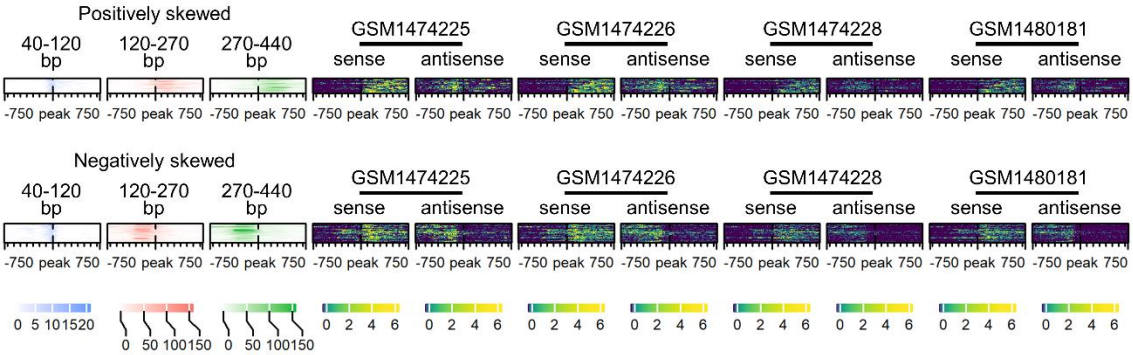
Total Pol II (Source data: 1-200518)



S5p Pol II (Source data: 7-200331)



S5p Pol II (Source data: 7-200518)



Online Resource 5 Skewed Pol II footprints and bidirectional transcription (Supplementary to Fig. 3e)