

A wave of nascent transcription on activated human genes

Youichiro Wada^{a,1}, Yoshihiro Ohta^{a,1}, Meng Xu^b, Shuichi Tsutsumi^a, Takashi Minami^a, Kenji Inoue^a, Daisuke Komura^a, Jun'ichi Kitakami^a, Nobuhiko Oshida^a, Argyris Papantonis^b, Akashi Izumi^a, Mika Kobayashi^a, Hiroko Meguro^a, Yasuharu Kanki^a, Imari Mimura^a, Kazuki Yamamoto^a, Chikage Mataka^a, Takao Hamakubo^a, Katsuhiko Shirahige^c, Hiroyuki Aburatani^a, Hiroshi Kimura^d, Tatsuhiko Kodama^{a,2}, Peter R. Cook^{b,2}, and Sigeo Ihara^a

^aLaboratory for Systems Biology and Medicine, Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8904, Japan; ^bSir Williams Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, United Kingdom; ^cGraduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259, Ngatsuda, Midori-ku, Yokohama 226-8501, Japan; and ^dGraduate School of Frontier Biosciences, Osaka University, 1-3, Yamadaoka, Suita, Osaka 565-0871, Japan

Edited by Richard A. Young, Whitehead Institute, Cambridge, MA, and accepted by the Editorial Board September 1, 2009 (received for review March 12, 2009)

Genome-wide studies reveal that transcription by RNA polymerase II (Pol II) is dynamically regulated. To obtain a comprehensive view of a single transcription cycle, we switched on transcription of five long human genes (>100 kbp) with tumor necrosis factor- α (TNF α) and monitored (using microarrays, RNA fluorescence in situ hybridization, and chromatin immunoprecipitation) the appearance of nascent RNA, changes in binding of Pol II and two insulators (the cohesin subunit RAD21 and the CCCTC-binding factor CTCF), and modifications of histone H3. Activation triggers a wave of transcription that sweeps along the genes at \approx 3.1 kbp/min; splicing occurs cotranscriptionally, a major checkpoint acts several kilobases downstream of the transcription start site to regulate polymerase transit, and Pol II tends to stall at cohesin/CTCF binding sites.

endothelial cell | polymerase II | RNA | tumor necrosis factor alpha

Transcription by RNA polymerase II (Pol II) is at the core of gene expression and hence is the basis of all cellular activities. To generate a mature messenger RNA (mRNA), Pol II traverses a transcription cycle; this involves recruitment to an activated promoter, initiation, escape into the gene, elongation, and termination (1). Processing of the nascent transcript—which can include capping, splicing, and poly (A) addition—is coupled to polymerization, and the C-terminal domain (CTD) of the polymerase acts as a scaffold for the binding of many of the factors involved (2–4).

Genome-wide analyses using chromatin immunoprecipitation followed by microarray analysis (ChIP-chip) or deep sequencing (ChIP-seq) provide powerful means of mapping comprehensively and at high resolution where proteins bind to the DNA template. Such studies have revealed widespread pausing and abortion by engaged Pol II (5–10). Pol II dynamics have also been studied using fluorescence imaging (11–15), but it remains difficult to observe both the rapid recycling of Pol II (14) and the unstable nascent transcripts.

To overcome these problems, we focus on the temporal profile of the first cycle of transcription after switching on transcription. As Pol II transcribes at more than 3 kbp per min (12), we analyzed five genes longer than 100 kbp that could be rapidly and synchronously activated by tumor necrosis factor- α (TNF α), a potent cytokine that orchestrates the inflammatory response by sequentially activating the expression of more than 6,000 genes in cultured human umbilical vein cells (HUVECs) (16, 17). To avoid problems caused by variations in sequence-specific signal in the arrays used (18) and to increase sensitivity, we developed analytical algorithms for temporal profiling that handled each probe sequence separately. Our repeated comprehensive observations reveal a wave of transcription sweeping along the genes; a major checkpoint regulates polymerase transit \approx 1–10 kbp into the genes, and polymerases tend to stall further downstream at

sites where the RAD21 subunit of cohesin and CCCTC-binding factor (CTCF) bind (19, 20).

Results

A Wave of premRNA Synthesis That Sweeps Down Activated Genes. At different times after stimulation with TNF α , total nuclear RNA was purified and hybridized to a tiling microarray bearing oligonucleotides complementary to *SAMD4A*, a long gene of 221 kbp; signals were normalized using an algorithm (see *SI Materials and Methods*). In Fig. 1, the height of the red and yellow needles reflects the intensity of signal of each probe that is given by premRNA binding to intronic and exonic probes, respectively. An alternative way of presenting these results and a control with antisense probe are illustrated in Fig. S1. A wave of intronic signal (red) appears to sweep down the gene from “Start” (at 15 min) to “End” (at 75–90 min). Exonic signal (yellow) increases significantly above the basal level only after \approx 75 min, when the polymerase has terminated. Similar waves are seen with other long genes, including *ZFPM2* (486 kbp; Fig. S2).

Abortive Transcription. Polymerases make many abortive transcripts of a few tens of nucleotides before forming stable elongation complexes (6, 8–10). However, probes covering the first thousands of nucleotides from the transcription start site (TSS) yield signal between 15 and 180 min (Fig. 1, green rectangle), and polymerases seem to escape downstream only for a limited interval (i.e., 15–30 min) to initiate the first wave. A similar pattern is seen with all long genes studied (Fig. S2). This points to a checkpoint that regulates escape, but here the checkpoint seems to act on a second polymerase once it has sensed that there is already the first on the gene (even though it might be >100 kbp downstream). This is a major checkpoint, as signal given by probes in the first three 1,000-nucleotide windows

Author contributions: H.K., T.K., P.R.C., and S.I. designed research; Y.W., M.X., S.T., T.M., K.I., A.P., A.I., M.K., H.M., Y.K., I.M., K.Y., C.M., and T.K. performed research; T.H., K.S., and H.A. contributed new reagents/analytic tools; Y.O., D.K., J.K., N.O., and H.A. analyzed data; and Y.W., Y.O., T.K., P.R.C., and S.I. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. R.A.Y. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession nos. GPL5828 and GSE9036, GSE9055, and GSE15157).

¹Y.W. and Y.O. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: kodama@lsbm.org or peter.cook@path.ox.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0902573106/DCSupplemental.

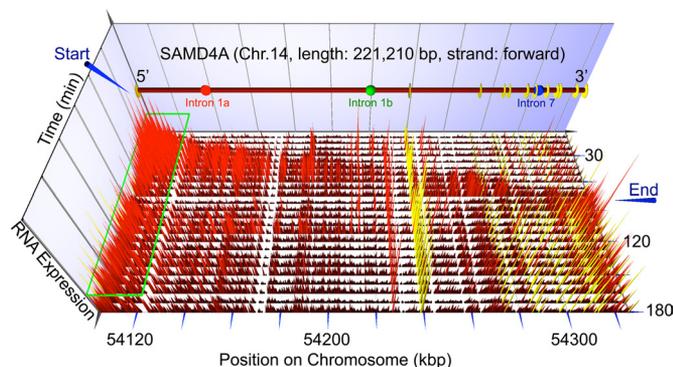


Fig. 1. Transcription waves visualized using microarrays. HUVECs were stimulated with $\text{TNF}\alpha$, samples collected every 7.5 min for 3 h, and total nuclear RNA purified and hybridized to a tiling microarray bearing 25-mers complementary to *SAMD4A*. The vertical axis gives intensity of signal detected by intronic and exonic probes (marked by red and yellow needles, respectively). Gene length and genomic location are shown at the front, probe positions within the gene from left to right; and time after stimulation from top to bottom. Blue arrowheads indicate the “Start” and “End” of the first wave of transcription that sweeps down the gene; green rectangle marks the position of probes continuously yielding signal between 7.5–180 min. Intronic targets for RNA FISH probes (1a, red; 1b, green; 7, blue) are indicated on the gene map.

in intron 1 of *ZFPM2* is more than two times higher than that in downstream windows (Fig. 2A).

Cotranscriptional Splicing and Intron Degradation. Figure 1 shows the wave reaches the middle of the second intron 60–75 min after stimulation; then, there is little signal in intron 1 (except close to the promoter, within the green rectangle). This is consistent with co-transcriptional splicing and degradation of RNA in the first intron while the polymerase is still transcribing the second. Quantitative analysis confirms this; summing signals given by all probes in 1,000-nucleotide windows between 15 and 180 min gives a “saw-tooth” pattern (Fig. 2A) with little signal at 3' ends of introns. Suppose that a constant number of Pol II molecules elongate at constant speed (without aborting) and that co-transcriptional splicing occurs at each intron-exon boundary; then the slopes of the blue regression lines in Fig. 2A should be constant, as they are (for further discussion, see Fig. S3). The initial part of the first exon is deliberately excluded from this analysis, as the changing levels at the checkpoint distort the picture (Fig. S3). These calculations support the idea that once a polymerase passes through the checkpoint, it usually then reaches the terminus (12). Similar patterns are seen with the other genes (Fig. S3).

Velocity of the Wave. We calculated the average speed of the wave as follows. In Fig. 2B, a blue dot for each probe indicates the first time when its expression reaches 50% of the maximum (and so marks the wave front); a red dot marks the average position of all blue dots at one time. The slope of the linear regression line drawn through the red dots and then reflects the velocity of the front. We estimate that the waves travel at 3.3, 3.2, 2.9, and 3.2 kb/min down *ZFPM2*, *EXT1*, *SAMD4A*, and *ALCAM*, respectively (Fig. S4B), giving an average of 3.1 kb/min. This is faster than 1.7–2.5 kb/min on human *DMD* (24), but slower than the 4.3 kb/min seen on an artificial array of ≈ 3 -kbp human genes (12). As many regulatory genes activated by $\text{TNF}\alpha$ have long and conserved introns (e.g., *ZFPM2*, *SAMD4A*, *NFKB1*), the time spent transcribing these introns must profoundly affect temporal control by the $\text{TNF}\alpha/\text{NF}\kappa\text{B}$ network (16, 17); long introns, which are conserved among many vertebrates, allow polymerases to convert space into time.

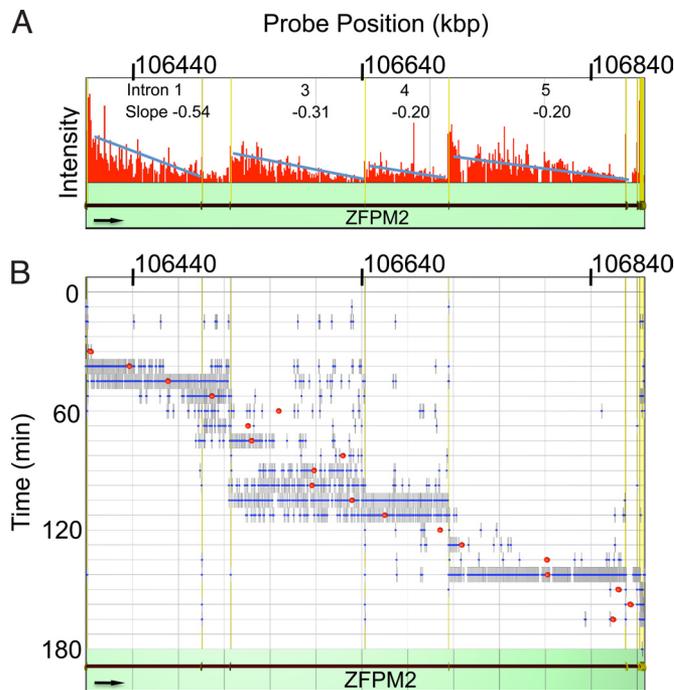


Fig. 2. Speed of the transcription wave on *ZFPM2*. (A) Decay of RNA. Sum total signals given by all probes in 1,000 base windows between 15 and 180 min was calculated and shown by red bars (height reflects intensity). Blue lines were obtained by linear regression (slope indicated in each intron), showing that signal declines to zero from beginning to end of each intron. (B) Speed of wave front. For each probe on the tiling array, a blue dot indicates where expression first reaches 50% of the maximum (using 40–70% of the maximum yields similar velocities; Fig. S3), and we use this to define the wave front. A red dot marks the average position of all blue dots at one time point. The velocity of the wave front was calculated by linear regression using the red dots (red line). Vertical axis shows time after stimulation. Genomic location is shown on top of each column. Direction of the gene is shown by the arrow.

Transcription Wave Detected by RNA Fluorescence *in Situ* Hybridization. Because tiling arrays provide information on average RNA levels in $\approx 10^6$ cells, cell-to-cell variation was monitored by fluorescence *in situ* hybridization (RNA FISH) (21, 22), using oligonucleotide probes able to detect a single transcript (23); intronic regions giving the strongest signals in tiling arrays were selected for analysis (*SI Materials and Methods*, Table S1). Probe 1a contains five 50-mers and carries ≈ 22 fluorophores; it is complementary to ≈ 600 bp in intron 1 of *SAMD4A* (Fig. 3A). Immediately after stimulation, it yields essentially no signal (Table S2). By 30 min, half of the cells possessed one or two, but never more than two, discrete nuclear foci (Fig. 3B, Table S2), which we assume mark a nascent transcript at one (or both) alleles. Later, the total number of foci in the population declined, but the intensity and size remained constant (Fig. 3D; Table S2). Probe 1b directed against a more 3' region of the same intron revealed an analogous wave that peaked later (at 60 min), and one against intron 7 later still (Fig. 3D, Table S2). Total signal in the population mimics the changes seen in arrays (Fig. 3D). In contrast, antisense probes (against regions 1a, 1b, 7) yield essentially no signal, and a probe targeting an intron in *EDNI* gave an unchanging signal (microarrays showed *EDNI* expression was unaffected by stimulation). These results show that at least half the cells in the population respond.

We confirmed that introns were removed co-transcriptionally by double labeling using probes 1a (green) and 7 (red). Although some cells possess foci of one or another (or both) colors (Fig. 3C), no foci contain both red and green signal and so appear yellow (Table S3). This suggests intron 1 must be removed before

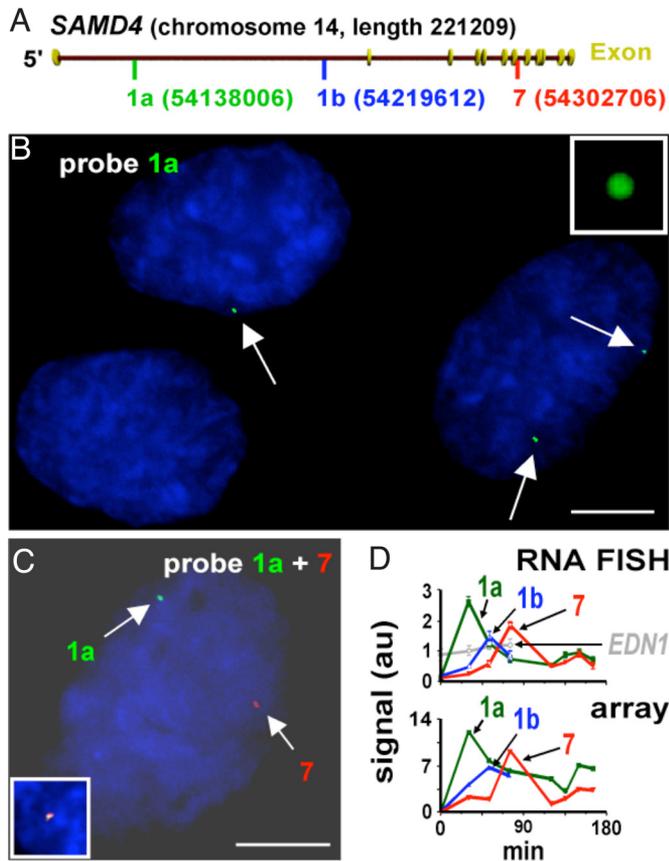


Fig. 3. A transcription wave visualized by RNA FISH. HUVECs were fixed at 0, 30, 52.5 and 75 min after stimulation with $TNF\alpha$, and nascent *SAMD4A* or *EDN1* RNA were detected by FISH using intron probes 1a, 1b, or 7, or an intron probe against *EDN1* (labeled with Alexa 488), cells counterstained with DAPI, and images were collected. *SAMD4* intron 1a (green), 1b (red), and 7 (blue) peaked as a wave of transcription passed through each region with time, whereas *EDN1* signal (gray) remains constant (D, upper graph). Similar variations were given by relevant probes in microarrays (D, lower graph). (A) *SAMD4A* locus showing probe positions. (B) A typical field 30 min after stimulation obtained using probe 1a. Cells have 0, 1, or 2 green foci/cell (arrows) marking nascent RNA at one or other allele. Intensities are normalized relative to fluorescent beads (inset) to permit comparison between different experiments. Bar, 5 μ m. (C) A nucleus 150 min after stimulation using probes 1a and 7; it contains one red and one green focus marking nascent RNA from each intron; yellow foci are never seen. (Inset) Positive control showing yellow focus given by probes 1a (green) and 1a-1 (red) 30 min after induction; these probes target intronic RNA sequences lying 1,000 nucleotides apart. Bar, 5 μ m. (D) RNA FISH and arrays give similar results. (Top) Signals (i.e., size in pixels \times intensity \times number of foci; in arbitrary units [au]) were obtained by single (open symbols; as in B) or double labeling (closed symbols, as in C). (Bottom) Similar variations are given by relevant probes in arrays.

intron 7 is made. These results confirmed that in individual cells, a wave of transcription runs along activated genes, and that splicing occurs co-transcriptionally.

Stalling/Slowing of Pol II at Cohesin-Binding Sites. The binding of Pol II was examined by ChIP-chip using antibodies against phospho-serine 5 in the heptad repeats in the CTD of the largest catalytic subunit (Rpb1); phospho-serine 5 is associated with transcriptional initiation and elongation (4, 24). Little Pol II was bound to *SAMD4A* (Fig. 4A, Fig. S5A) or *EXT1* (Fig. S5B) before stimulation (at 0 min), although some (presumably “paused”) Pol II was detected near the TSS (indicated by the single asterisk in Fig. 4A). After 30 min, Pol II was bound to the 5' half of the gene, and after 60 min it was bound more to 3' (although there

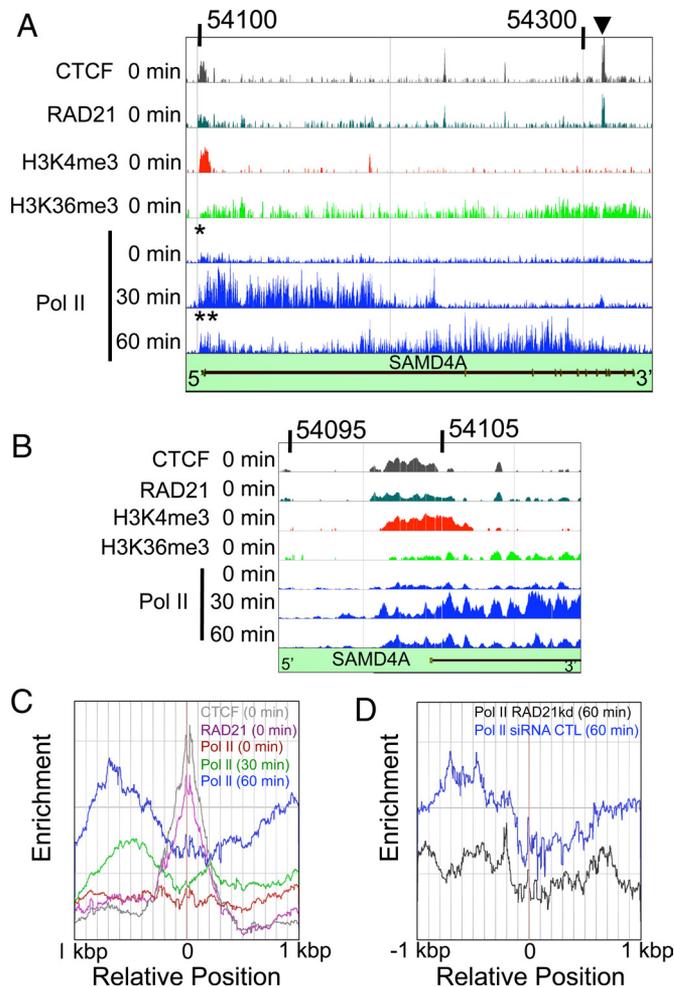


Fig. 4. Stalling of Pol II analyzed using chromatin immunoprecipitation. HUVECs were stimulated with $TNF\alpha$ and harvested after 0, 30, and 60 min; then binding of CTCF, RAD21, modified histones (H3K4me3, H3K36me3), and elongating Pol II (phospho-Ser-5 modification) to *SAMD4A* was analyzed by ChIP-chip (CTCF, RAD21, H3K4me3, H3K36me3, and Pol II). Numbers on top of A and B show the location of the genomic region of Chr14 (*SAMD4A*) analyzed. (Vertical axes) Enrichment of binding. (A) Binding to *SAMD4A*. CTCF and RAD21 are often found together, consistent with the binding of a functional insulator complex. Asterisk shows where (engaged) Pol II binds at 0 min near the TSS, suggesting that it might be paused or poised. Double asterisk shows where Pol II binds near the TSS at 60 min. Arrowhead shows representative colocalization site of RAD21 and CTCF \approx 210 kilonucleotides downstream of the TSS. (B) Binding at the TSS of *SAMD4A*. H3K4me3 and CTCF/RAD21 bind in/around the TSS. At 0 min, engaged Pol II also binds to this region. At 30 min, Pol II binding spreads into the gene, and after 60 min it becomes more concentrated around the TSS again. (C) Enrichments (number densities) of engaged Pol II near 35 sites distant from the TSS that were marked by bound RAD21 (pink) and CTCF (gray). At 0 min (red line), Pol II binds symmetrically around the RAD21/CTCF. At 30 min (green line), the amount of Pol II binding increases significantly. At 60 min (blue line), Pol II binding increases further and becomes concentrated upstream of the RAD21/CTCF; this is consistent with polymerase stalling. (D) An 80% reduction in RAD21 levels (achieved using siRNA) at 60 min (black) destroys the accumulation 5' of the RAD21/CTCF site that is seen with a control siRNA (blue).

was still some binding near the TSS; double asterisks in Fig. 4A). These results confirm those obtained using arrays and FISH.

As some Pol II was bound near the TSS, we investigated the distribution of various other markers, namely, two histone modifications (H3K4me3 and H3K36me3), and two insulator proteins (the cohesin subunit RAD21 and CTCF) (Fig. 4A, Fig. S5A and B). The distribution of bound Pol II overlaps that

of H3K4me3, a marker for active chromatin (Fig. 4A, Fig. S5A), as previously described (8, 9). RAD21 and CTCF were often, but not invariably, bound to the same sites at the same time (Fig. 4A), again as seen previously (19, 20). Closer inspection (Fig. 4B, Fig. S5 C–G) shows that RAD21 and CTCF often bind near the boundaries of regions rich in Pol II and H3K4me3. It was previously reported that histone H3K4 and H3K36 tri-methylation are associated with the elongating polymerase (25), but we could not see the correlation (Fig. 4A, Fig. S5A and B). At 60 min (i.e., after the wave had passed), the amount of paused Pol II increased near the TSS, the boundaries of which were again marked by RAD21/CTCF binding (Fig. S5A and B).

We next examined the averaged distribution of Pol II around 35 RAD21/CTCF binding sites distant from the TSS (Fig. 4C); sites were selected as described in Fig. S6. At 0 time, little Pol II (red line) was found at the sites (gray and purple lines). However, after 30 and 60 min, Pol II (green and blue lines) accumulated upstream, while being excluded from 200–400 bp around the sites. By definition, all 35 sites bound both CTCF and RAD21, and Pol II accumulated on the 5' side of 33 of the 35 sites. This suggests bound RAD21/CTCF stalls/slow the polymerase (25). To confirm this, we performed ChIP-chip after “knocking-down” RAD21 levels by 80% using siRNA; less Pol II (Fig. 4D, black line) was now bound near RAD21/CTCF sites, compared to the control (Fig. 4D, blue line). We conclude that Pol II tends to stall at bound RAD21/CTCF.

Discussion

Here we use TNF α to switch on transcription of five human genes rapidly and synchronously. As the genes are all longer than 100 kbp, and as samples are collected every 7.5 min for 3 h, there is ample time to monitor one complete transcription cycle. We use tiling microarrays and RNA FISH to follow the appearance of nascent RNA, and ChIP-chip to monitor changes in binding of Pol II and two insulator proteins (the cohesin subunit, RAD21, and CTCF), as well as two histone modifications (H3K4me3, H3K36me3). High-resolution data collection followed by statistical analysis in living human cells revealed precise mode that TNF α induces a wave of nascent RNA and Pol II to sweep along the genes from promoter to terminus (Figs. 1, 2, and 4A); these results are consistent with the polymerase initiating soon after stimulation, and transcribing \approx 3.1 kbp/min (Fig. S4). Once RNA in a 3' intron is seen, RNA in more 5' introns has disappeared; for example, RNA FISH revealed that RNA from intron 7 is never found with its counterpart from intron 1 (Fig. 3C). Moreover, once RNA at the 3' end of a long intron appears, more 5' sequences in the same intron have disappeared (Figs. 1 and 2A). Here, splicing and degradation of intronic RNA occur co-transcriptionally, with degradation of intronic RNA beginning even before the polymerase reaches the next exon, and well before it terminates. These results contrast with a genome-wide analysis that indicates that premRNA splicing in yeast is predominantly posttranscriptional (26); but the results are consistent with observations pointing to a functional coupling between transcription and splicing (27, 28).

It is easy to imagine how a brief pulse of initiation soon after stimulation would trigger the wave. However, polymerases continue to initiate throughout the period analyzed, with intronic signal falling significantly 1–10 kbp into the gene (e.g., after the green rectangle in Fig. 1). This could be because polymerases speed up, or, more likely, abort. Because polymerases continue to initiate throughout the period analyzed, the checkpoint must

operate to allow only the first ones to pass through while forcing later ones to abort, as only then can the widening trough between the peak of newly-initiated transcripts at the TSS and the advancing wave be generated. This checkpoint is much further into the gene than another major one acting only a few tens of nucleotides from the TSS (5, 6, 8–10). It seems able to sense whether another (pioneering) polymerase is already transcribing the gene, and we can only speculate on how it might do so.

Although the wave apparently progresses steadily down the gene, detailed analysis shows the polymerase tends to stall or slow at sites where RAD21, a subunit of cohesin, binds (Fig. 4A, Fig. S5A and B). RAD21 often bound to the same sites as CTCF (Fig. 4A, arrowhead); and, as CTCF physically links cohesin to chromatin (29), these complexes could act as steric barriers to the elongating polymerase. This notion was supported by statistical analysis of 35 sites; Pol II tended to accumulate just upstream of these sites (Fig. 4C), and “knocking down” RAD21 abolishes this accumulation (Fig. 4D). Although the effects at these sites are clearly more subtle than those occurring at the major checkpoint within 1 to 10 kbp of the TSS, this finding provides evidence that elongation of Pol II can be regulated by an epigenetic modification.

In conclusion, we obtained a comprehensive view of a complete transcription cycle on several long human genes. One challenge now is to see whether the rich levels of regulation that we observe in these long genes are also found in genes of more average length.

Materials and Methods

Cell Culture. Before stimulation, HUVECs were serum starved in EBM-2 (Clenetics) containing 0.5% fetal bovine serum (FBS) for 18 h, treated \pm 10 ng/ml TNF α , and harvested at various times thereafter (17).

Tiling Microarrays. We designed tiling microarrays (NimbleGen) covering *ZFPM2*, *EXT1*, *SAMD4A*, *ALCAM*, *NFKB1*, and *EDN1*. Probes (25 nucleotides) were selected so that the median of probes locate with 12 base interval, and designed to minimize cross hybridization with duplicated and repeated sequences in the human genome (*SI Materials and Methods*) (5, 30). Total RNA was purified using ISOGEN (NipponGene, Japan) and hybridized to microarrays (*SI Materials and Methods*).

RNA FISH. Transcripts were detected using RNA FISH (*SI Materials and Methods*). Probe 1a consists of five 50-mers with \approx 22 flours (Table S1) and is complementary to \approx 600 nucleotides in the middle of intron 1 of *SAMD4A* (Fig. 3A); a probe of this type is able to detect a single transcript (21). Sense signals were detected by probes the position and sequence of which are given in Fig. 1A, Fig. 3A, and Table S1. The antisense probes (against regions 1a, 1b, 7) yielded essentially no signal, and a probe targeting an intron in *EDN1* gave an unchanging signal (microarrays showed that *EDN1* expression was unaffected by stimulation).

ChIP-Chip. Detailed experimental procedures for ChIP followed by analysis using microarrays, and for the analysis of Pol II stalling at cohesin/CTCF sites are described in *SI Materials and Methods*. The stalling profiles of Pol II at RAD21 binding sites are given in Fig. 4 and Fig. S6.

ACKNOWLEDGMENTS. This work was supported by the Special Coordination Fund for Promoting Science and Technology and by grants from the New Energy and Industrial Technology Development Organization, the National Institute of Biomedical Innovation of Japan, the Strategic Information and Communications R&D Promotion Program of the Ministry of Internal Affairs and Communications, the Genome Network Project, Japan Grants-in-Aid for Scientific Research (s) 16101006, 19800009, and 19310129 from the Ministry of Education, Culture, Sports, Science and Technology, and the Japan Society for the Promotion of Science. A.P. is supported by The Wellcome Trust. M.X. thanks Oxford University (Clarendon Award), the U.K. Government (Overseas Research Student Award), and the E.P. Abraham Research Fund for support.

1. Kornberg RD (2007) The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci USA* 104:12955–12961.
2. Maniatis T, Reed R (2002) An extensive network of coupling among gene expression machines. *Nature* 416:499–506.

3. Orphanides G, Reinberg D (2002) A unified theory of gene expression. *Cell* 108:439–451.
4. Phatnani HP, Greenleaf AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* 20:2922–2936.

Supporting Information

Wada et al. 10.1073/pnas.0902573106

SI Materials and Methods

Cell Culture and Microarray Analysis. HUVECs (Clonetics) were serum starved in EBM-2 (Clonetics) containing 0.5% FBS for 18 h and then treated with or without 10 ng/ml TNF α (Peprotec). Samples were collected every 15 min for 4 h (3).

Genomic Tiling Array. Initial experiments were performed using a GeneChip Human U133 Plus 2.0 oligonucleotide array; the experiments were performed according to the Affymetrix GeneChip expression analysis technical manual (Affymetrix, Santa Clara, CA). Briefly, 3–5 μ g total RNA were used to synthesize biotin-labeled cRNA, which was then hybridized to a GeneChip Human U133 Plus 2.0 oligonucleotide array. After washing, arrays were stained with streptavidin-phycoerythrin and imaged using an Affymetrix GCS 3000 7G scanner; data were analyzed using GeneChip Analysis Suite software 5.0 (available at <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9055>). The average intensities given by all probe sets were set at 100. A total of 3,157 probe sets yielded signals that vary >200-fold (average difference) during the time course with a coefficient of variation of expression >0.2 during 0–240 min. During this analysis, we noticed 62 probe sets mapped to regions annotated as introns in the University of California Santa Cruz Genome Browser (release May 2004); most are referred to as expressed sequence tags. These results prompted us to analyze the transcription wave in detail using tiling arrays.

Probes (25mers) were designed to minimize cross hybridization with duplicated sequences in the human genome and repeated or low-complexity sequences. For *SAMD4A*, probes covered nonrepetitive genomic sequence from transcription start site (TSS) to the 3' untranslated region (3'-UTR); perfectly matched (PM) and mismatched (MM) probes for both strands were based on every 13 (median) bases. For *ZFPM2*, *EXT1*, *SAMD4A*, *ALCAM*, and *NFKB1*, probes covered 5 kilobases upstream of the TSS to 5 kbp downstream of the 3'-UTR; PM probes for both strands were based on every 25th (median) base. For the other genes, PM probes for the sense strand were based on every 25th (median) base selected. To evaluate background hybridization signals, more than 60,000 p.m. probes were designed complementary to nontranscribed sequences (i.e., sequences free of RefSeq genes).

Probe data for *ZFPM2* (12,631 probes), *EXT1* (12,581 probes), *SAMD4A* (9,857 probes), *ALCAM* (7,989 probes), and *NFKB1* (4,545 probes) are available at: <http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE9036>.

A 1-mg quantity of each sample was processed using the Affymetrix GeneChip Whole Transcript Sense Target Labeling Assay. Eight mg cRNA were input into the second-cycle cDNA reaction. Hybridization cocktails containing 3–4 mg fragmented, end-labeled cDNA were prepared and applied to the NibleGen tiling arrays. Hybridization was performed for 16 h using the MES_EukGE-WS2v5.450 fluidics wash and stain script. Arrays were scanned using the Affymetrix GCS 3000 7G and GeneChip Operating Software v. 1.3 to produce .CEL intensity files. Raw array data for each time point were analyzed using the Tiling Arrayanalysis Software (TAS) provided by Affymetrix. This program assigns probabilistic enrichment scores to collections of neighboring probes within a sliding window. We exclude probes complementary to repeats, and those with extremely high (>18 bases in the 25mer) or low (<5 bases in the 25mer) G/C content, as they gave higher backgrounds. Using the results of MM probes and thus PM-MM values, we estimate the signals for *SAMD4A*.

For other genes, signals from probes on the antisense strand were used to estimate array-wide background signals (4). The signal enrichment is mapped to the genome using exact 25-mer matching to hNCBIv36. To standardize the expression of tiling array, we normalized all expression data among all five genes using the expression data for *EDNI*. However the difference of results obtained with/without this normalization was very small. To detect time-dependent changes, we did not use any normalization procedure between time points except that the expression values at 0 min were subtracted from those at 7.5–180 min. Thus, the fully mature transcripts accumulate, and these do not influence the abundance of the comparatively lower levels of intronic RNA. Therefore, the time-dependent changes seen with the tiling arrays are not calculation artifacts.

RNA FISH. Nascent transcripts were detected by RNA FISH (5, 6). Cells were rinsed in phosphate-buffered saline (PBS), fixed (17 min; 20 °C) in 4% paraformaldehyde, 0.05% acetic acid in 0.15 M NaCl, rinsed 3 times (5 min; 20 °C) in PBS, permeabilized (5 min; 37 °C) in 0.01% pepsin (pH 2.0), rinsed in H₂O treated with diethylpyrocarbonate, postfixed (5 min; 20 °C) in 4% paraformaldehyde in PBS, and washed (10 min; 20 °C) in PBS before hybridization (16 h, 37 °C) in a moist chamber. Hybridization mix contained 50 ng labeled probe (below), 25% deionized formamide, 2 \times SSC, 200 ng/ml sheared salmon sperm DNA, 5 \times “Denhardt’s” solution (0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin [BSA]), 50 mM phosphate buffer (20 mM KH₂PO₄, 30 mM KHPO₄·2H₂O, pH 7.0), and 1 mM ethylene diamine tetraacetic acid (EDTA). Cells were then washed three times in 2 \times SSC (15 min; 37 °C), twice (5 min; 20 °C) in Tris/Saline/Tween (TST) (0.15 M NaCl, 0.1 M Tris-HCl, 0.05% Tween 20, pH 7.5), once (5 min; 20 °C) in Tris/Saline (TS) (0.15 M NaCl, 0.1 M Tris-HCl, pH 7.5), and mounted in Vectashield (Vector Laboratories; Burlingame, CA) containing 1 μ g/ml DAPI (4', 6-diamidino-2-phenylindole; Sigma) and 2.5 μ m Green and/or Orange Intensity Calibration Beads (0.02% intensity; Molecular Probes) at 6 \times 10⁴/ml. Images were collected on an Axioplan 2 microscope (Zeiss; Welwyn Garden City, Herts, UK) with a CCD camera (Photometrics CoolSNAPHO; Marlow, UK), using an exposure that gave a signal intensity of the beads of 3,500–4,095 on the gray scale. Signal intensities were measured using ImageJ (Rasband 1997–2007) and normalized relative to the intensity of the reference beads.

Five 50-mer probes were synthesized (Gene Design, Osaka, Japan) for each FISH target (Table S1). Only four probes were used for 1a-1. Every \approx 10 nucleotides in each 50-mer, a T was substituted by an amino-modifier C6-dT and was labeled with Alexa 488 using Alexa Fluor 488 reactive dye (Invitrogen) or Cy3 using Cy3 Mono Reactive Dye Pack (GE Healthcare, Amersham, UK). Briefly, 12 μ g oligonucleotide was phenol extracted, ethanol precipitated, and resuspended in 6 μ l 0.5 M NaHCO₃ (pH 9.3). One vial of dye was added into the solution, incubated (20 °C; 60 min), mixed at 15-min intervals, and incubated for another 60 min (20 °C). Another vial of dye was added using the same procedures, before a third vial was added with an additional 6 μ l 0.5 M NaHCO₃ (pH 9.3). Probes were then purified using G-50 columns (GE Healthcare), ethanol precipitated, and concentrated using a Microcon-30 column (Millipore). Labeling efficiencies were calculated from the absorption of bases and bound fluors using published extinction coefficients, and were between 19 and 23 fluors per 250 nucleotides. A single *SAMD4A* 50-mer with an average of 4.5 fluors had \approx 17% the fluorescence

intensity of a bead (determined after depositing dilutions of probe:bead mixtures on a slide; Table S2).

Chromatin Immunoprecipitation. Two million HUVECs were plated in a 15-cm culture plate, followed by serum starvation. The HUVECs were stimulated with TNF α at a concentration of 10 ng/ml at time 0, and cells were crosslinked for 10 min using 1% paraformaldehyde at the appropriate time thereafter. After neutralization using 0.2 M glycine, cells were recollected, resuspended in sodium dodecyl sulfate (SDS) lysis buffer (10 mM Tris-HCl, 150 mM NaCl, 1% SDS, 1 mM EDTA; pH 8.0), and fragmented by sonication (Branson; 3 min for microarray studies, 10 min for high-speed sequencing). Samples were stored at -80°C before use. To perform chromatin immunoprecipitation, antibodies against CTCF (Upstate 07-729), RAD21 (provided by Dr. Shirahige), H3K4me3 (Abcam, ab8580), H3K36me3 (Abcam, ab9050), and RNA polymerase II (Covance, MMS-134R) were purchased and used in combination of Protein A/G (GE Biotech) or magnetic beads (Dyna/Invitrogen). Prepared DNA was quantified using Qubit (Invitrogen), and more than 10 ng of DNA was processed as described below.

ChIP-Chip Sample Preparation and Analysis. DNA prepared by ChIP was subjected to *in vitro* transcription twice for amplification as

described previously (7). After fragmentation using DNase, samples were labeled with biotin-N11-ddATP (NEL548, Perkin-Elmer) and hybridized with a genomic tiling microarray as described above (8).

Gene Knockdown by siRNA. HUVECs cultivated in EGM2MV were removed in Opti-MEM culture medium (GIBCO, Invitrogen) and were transfected with stealth RNAi for RAD21 (Invitrogen, HSS109005) at a concentration of 10 μM using Lipofectamine RNAiMax reagent (Invitrogen). At 6 and 24 h later, culture medium was changed into EGM2MV, and 48 h later in serum-starved medium. At 18 hours later, cells were stimulated with TNF α and samples for ChIP were prepared using ChIP Reagents kit (NipponGene, Japan). Knockdown efficiencies of RAD21 were validated by real-time polymerase chain reaction (CFX96, BioRad) using primers as follows: FW5'-CTGATTCAGTGGATCCCGTT-3', RV5'-GCTTCCTCTCCTCTGGCT-3', and Western blots of whole cell fraction using the same antibody described above.

Microarray Data Viewer. Data in Fig. 1 are presented using a three-dimensional viewer written using POV-Ray.

1. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 130:77–88.
2. Bon M, McGowan SJ, Cook PR (2006) Many expressed genes in bacteria and yeast are transcribed only once per cell cycle. *FASEB J* 20:1721–1723.
3. Inoue K, Kobayashi M, Yano K, Miura M, Izumi A, et al. (2006) Histone deacetylase inhibitor reduces monocyte adhesion to endothelium through the suppression of vascular cell adhesion molecule-1 expression. *Arterioscler Thromb Vasc Biol* 26:2652–2659.
4. Johnson WE, Li W, Meyer CA, Gottardo R, Carroll JS, et al. (2006) Model-based analysis of tiling-arrays for ChIP-chip. *Proc Natl Acad Sci USA* 103:12457–12462.
5. Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, et al. (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 36:1065–1071.
6. Femino AM, Fay FS, Fogarty K, Singer RH (1998) Visualization of single RNA transcripts in situ. *Science* 280:585–590.
7. Liu CL, Schreiber SL, Bernstein BE (2003) Development and validation of a T7 based linear amplification for genomic DNA. *BMC Genomics* 4:19.
8. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, et al. (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451:796–801.

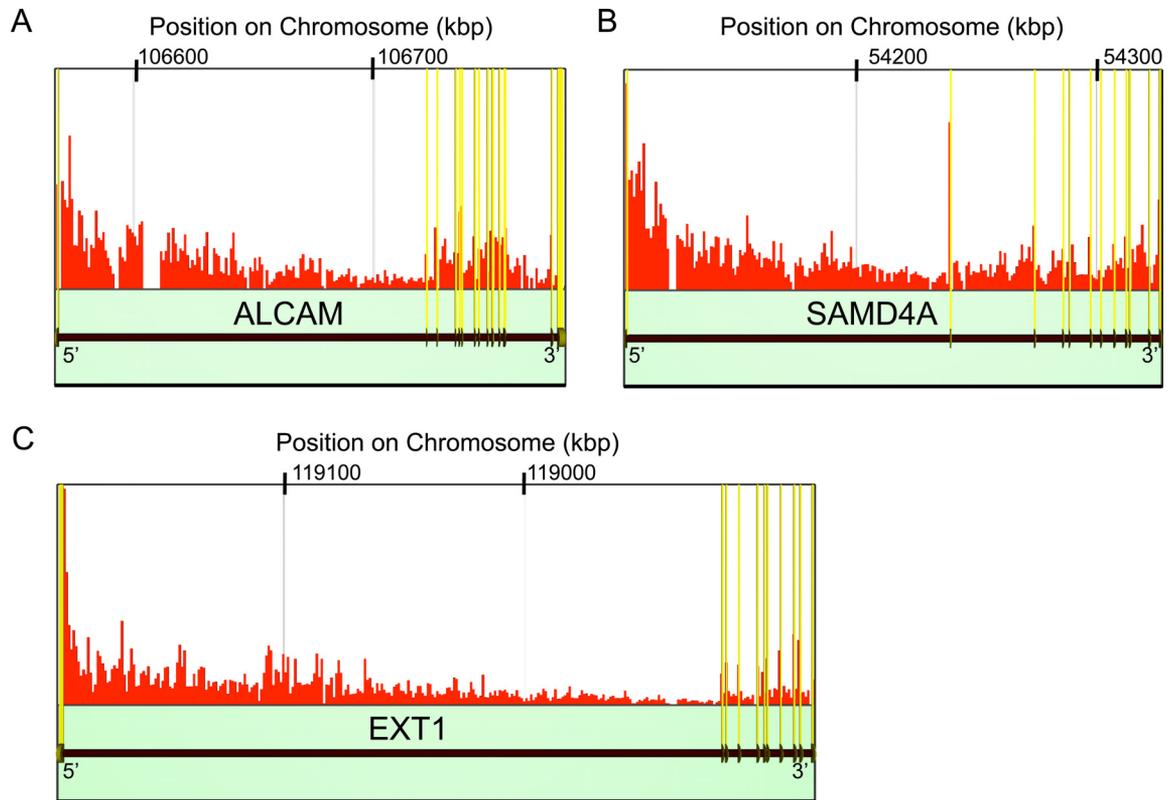


Fig. S3. Premature termination in the first long intron of *ALCAM*, *SAMD4A*, *EXT1*, and *ZFPM2*. (A–C) Using data from the tiling arrays shown in Fig. S2, sum total signals given by all probes in 1,000 base windows between 15 and 180 min were calculated and are shown by red bars (height reflects intensity). Position on chromosome is shown on top (kbp), and gene name and direction provided in lower column. (D–G) All signals detected by probes in the first long intron and within windows of 1,000 bases between 15 and 180 min are summed and plotted as a function of position as in Fig. 2A. Blue dot at position 1 kb gives the slope of the regression line through all values. Slopes at different positions are calculated by successively omitting 1-kbp windows from the 5' end. In each case, after omitting the first few thousand bases, the slopes (which reflect transcript levels) reach a plateau; this is consistent with premature termination occurring within several kbp downstream of the TSS (except when the wave is allowed to pass through). As described below, the high signal values in the first $\approx 3,000$ bases of intron 1 are, on average, more than twice the values seen at the 5' ends of the other introns in the same gene. Modeling premature termination: We analyzed intronic RNA levels and premature termination in intron 1 of a given gene using a simple model involving one polymerase elongating at constant speed, v ($= 3.3$ kb/min). When the polymerase reaches the end of k^{th} intron (at position i_k from the 5' end of the gene) and moves on to the next $(k + 1)^{\text{th}}$ exon, the k^{th} intron is spliced out and degraded immediately; then, essentially no signal will be given by the most 3' probe in the k^{th} intron. This would generate the saw-tooth pattern seen in Fig. 2A. If l_k is the length of k^{th} intron, the signal given by RNA at position x_k of the k^{th} intron ($i_k - l_k < x_k < i_k$) at time t is 0, for $t < x_k/v$. For $x_k/v < t < i_k/v$, signal at x_k (which is now detectable by probes in the body of the intron) becomes a constant, C_k , that is proportional to the number of transcripts at x_k . The time integral of the signal of intron transcripts over the length of the experiment gives $C_k (i_k/v - x_k/v)$. This means that the derivative of the time integral of the signal of the transcript at x_k gives $-C_k/v$. More generally, if C_k has x_k dependence, an extended model gives $-C_k(x)/v$. If we assume that both v and C_k are constant over a range of x_k , we can estimate C_k from the value of the slope when the time integrated signal is plotted against position. For example, we use this model to analyze $C_k(x)$ in introns of *ZFPM2*, *EXT1*, *ALCAM*, and *SAMD4A*. The sharp drop in $C_1(x)$ after 3,000 bases is consistent with the idea that a significant proportion of transcripts terminate prematurely within the first few thousand bases of intron 1 (except at the time when the wave is allowed to pass through the checkpoint).

B

Symbol	Velocity (kbp/min)	Length (kbp)	No. of junctions	Exon Length (kbp)
<i>ZFPM2</i>	3.3	485.62	7	4.48
<i>EXT1</i>	3.2	312.46	10	3.36
<i>SAMD4A</i>	2.9	221.21	11	2.34
<i>ALCAM</i>	3.2	210.03	15	4.74
<i>NFKB1</i>	2.8	115.99	23	4.09

Figure S4 (continued).

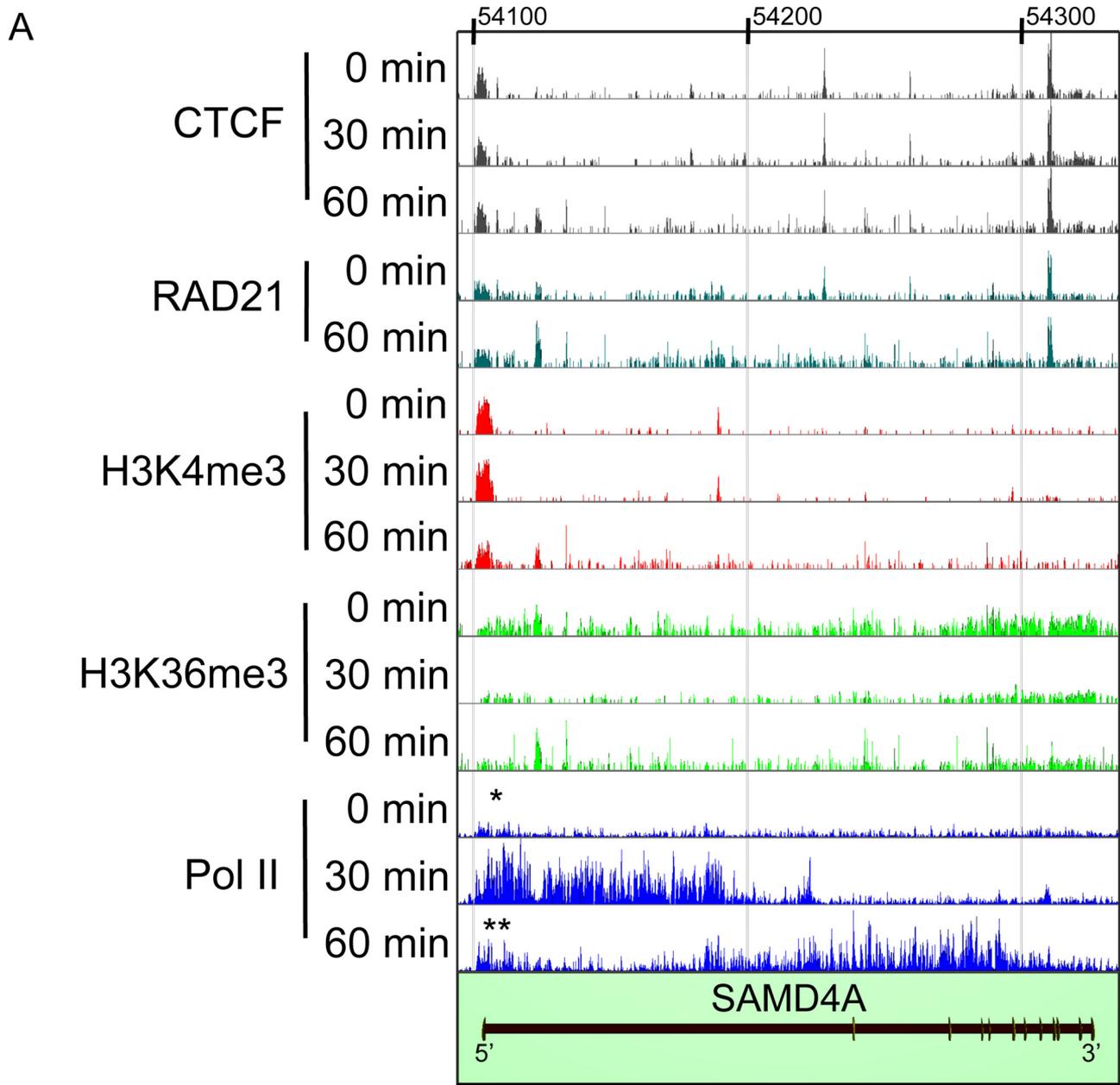


Fig. S5. Movement and stalling of Pol II analyzed using ChIP-chip. HUVECs were stimulated with $\text{TNF}\alpha$ and harvested after 0, 30, and 60 min; then binding of CTCF, RAD21, various modified histones (H3K4me3, H3K36me3), and elongating Pol II (phospho-Ser-5 modification) to *SAMD4A* (A) and *EXT1* (B) was analyzed by ChIP-chip (CTCF, RAD21, H3K4me3, H3K36me3, and Pol II). Vertical axes show the fold-enrichment of binding. The enrichment ratio was obtained as the ratio of fold enrichment at active positions relative to reference positions. Numbers on top of panels in A and B show the location of the genomic region (kbp) of Chromosome 14 (*SAMD4A*) and Chromosome 8 (*EXT1*) analyzed. Asterisk shows where (engaged) Pol II binds at 0 min, suggesting that it might be paused or poised near the TSS. Double asterisks mark Pol II binding near the TSS at 60 min. High-magnification views of the TSS of *SAMD4A* (C), *EXT1* (D), *ALCAM* (E), *NFKB1* (F), and *ZFPM2* (G) are also shown.

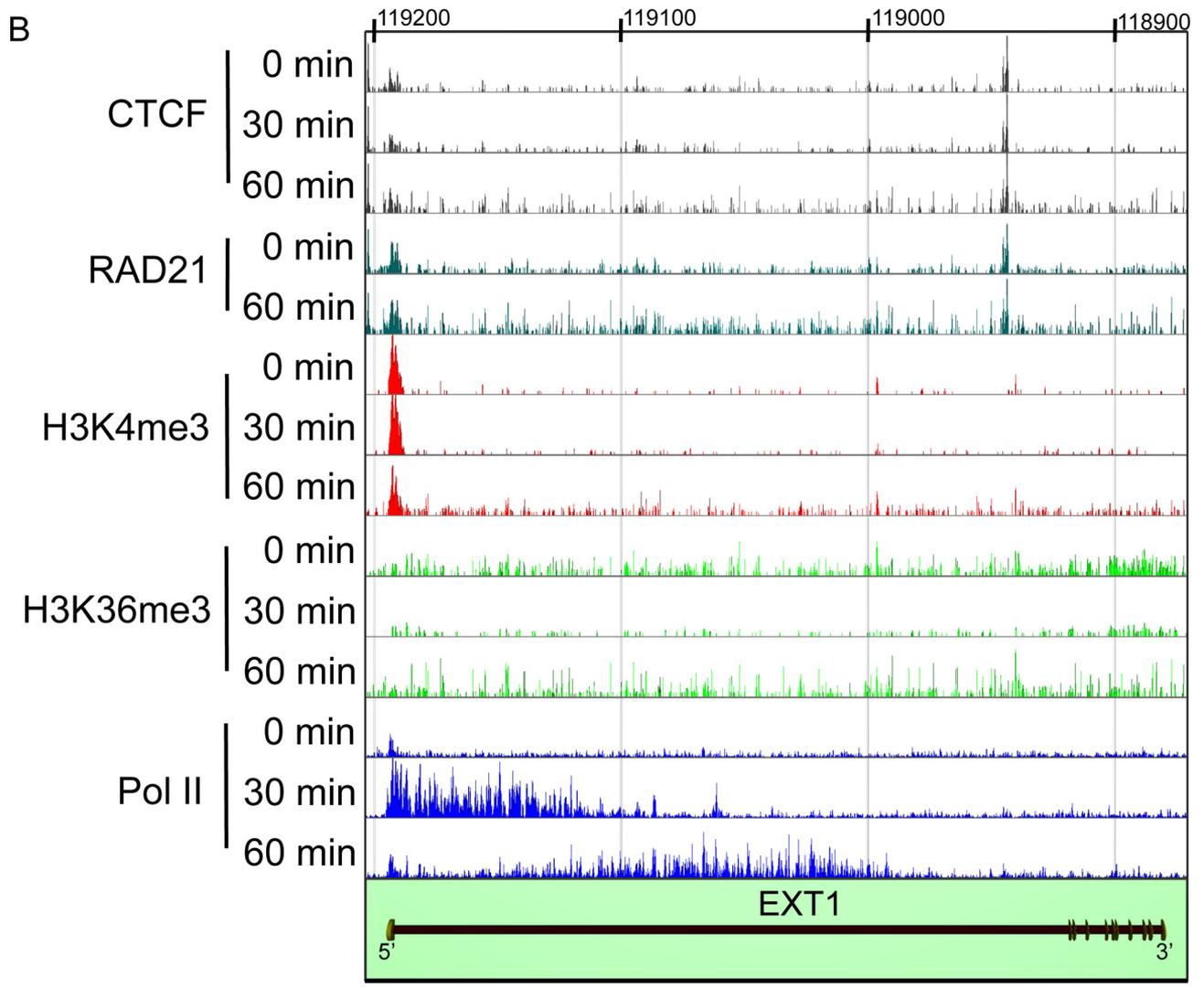


Figure S5 (continued).

C

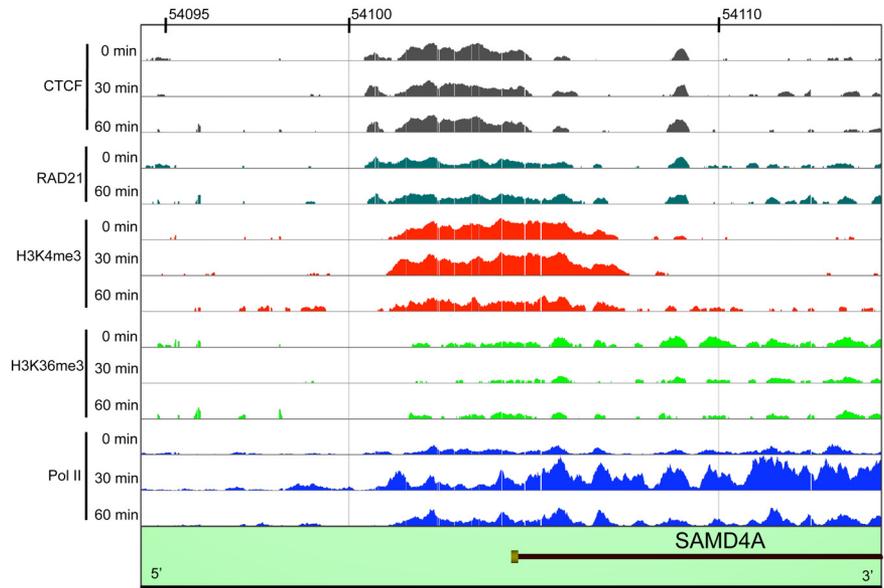


Figure S5 (continued).

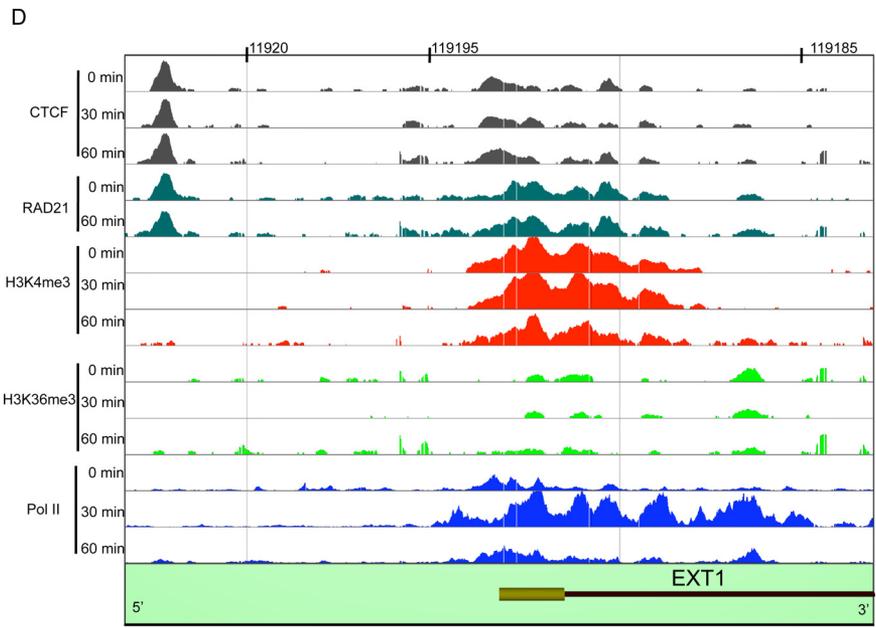


Figure S5 (continued).

E

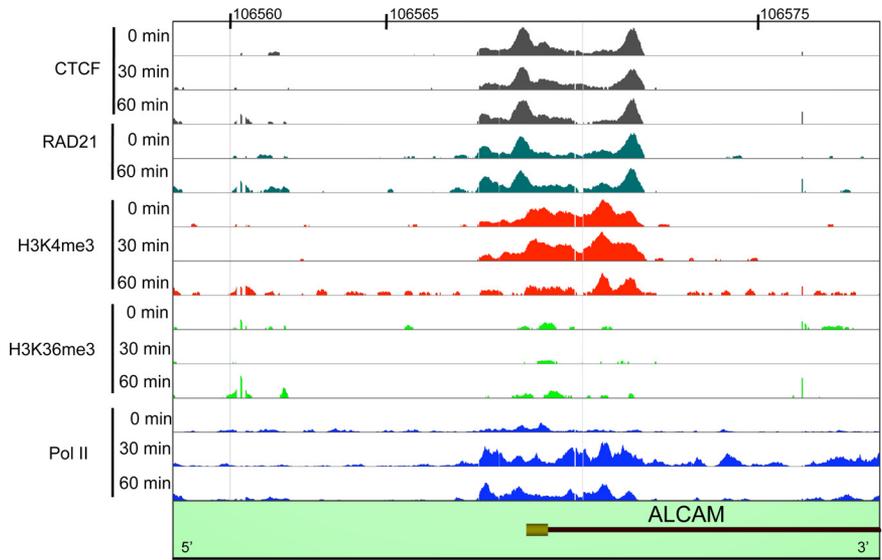


Figure S5 (continued).

F

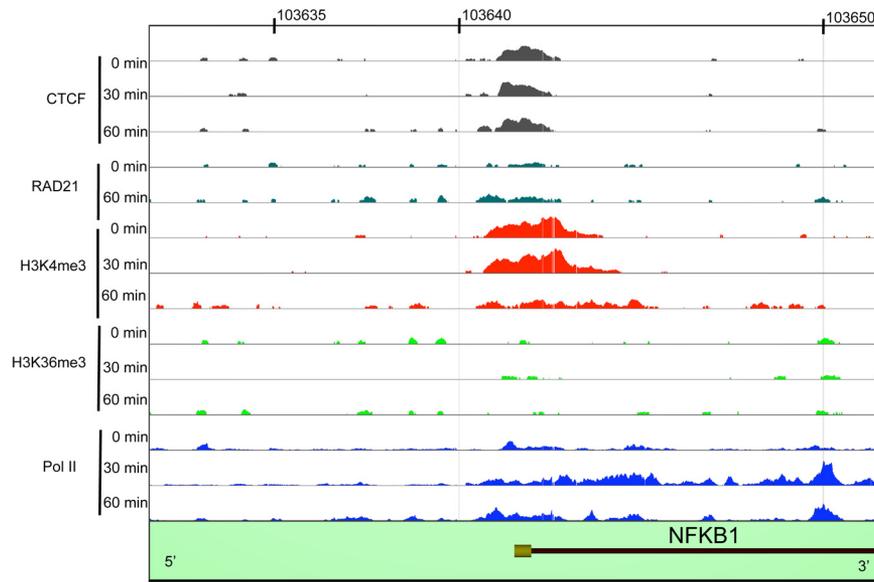


Figure S5 (continued).

G

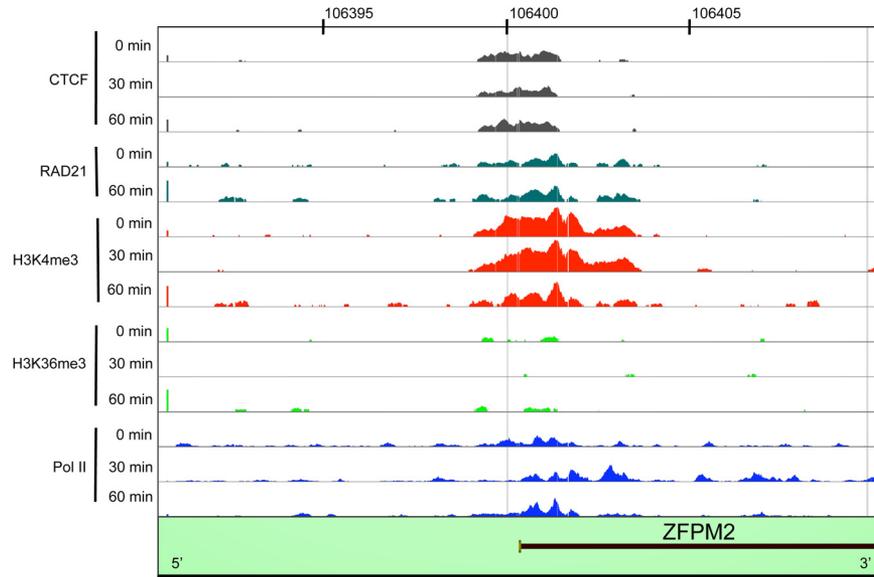


Figure S5 (continued).

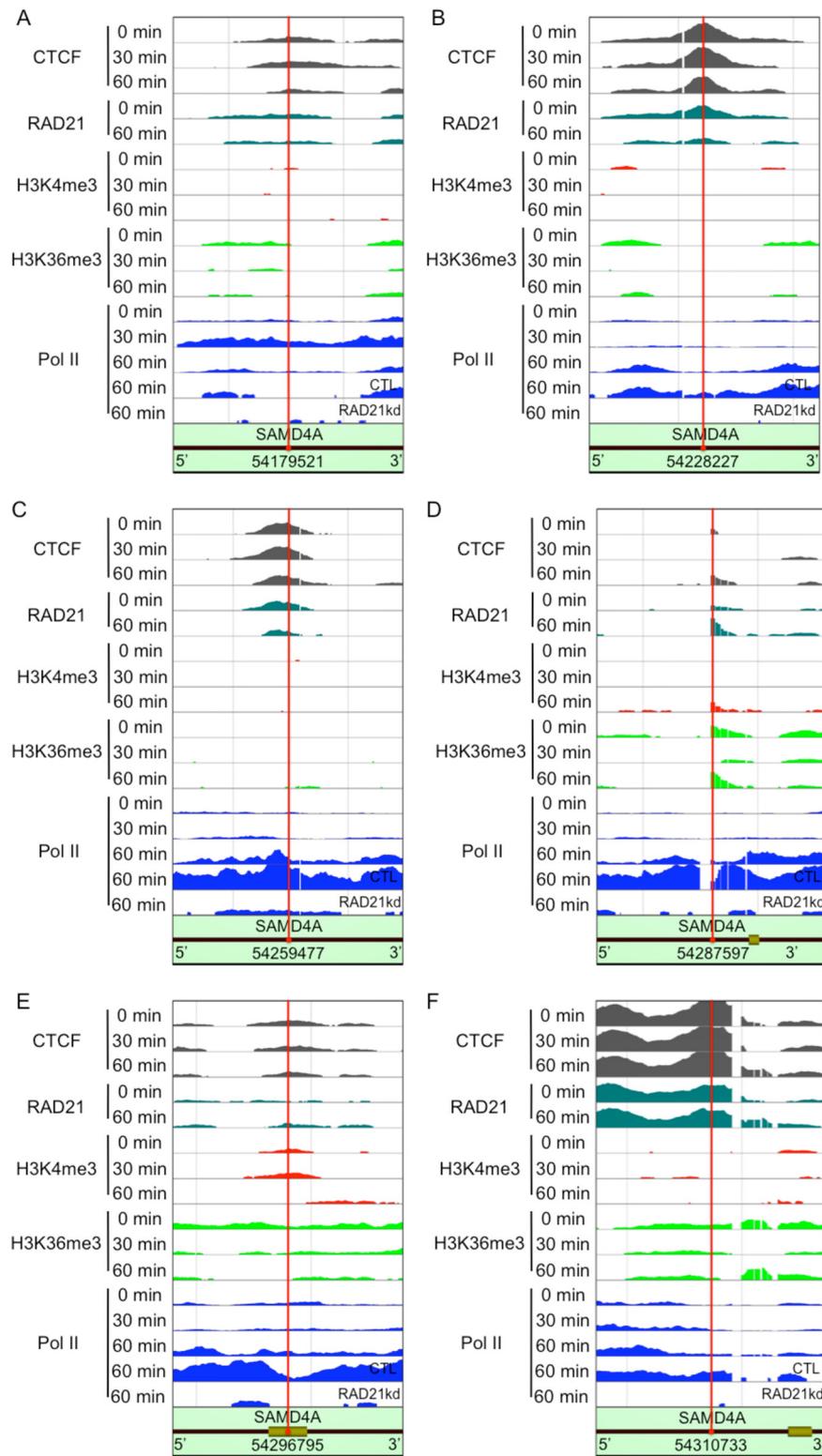


Fig. S6. Movement and stalling of Pol II in the body of genes analyzed using ChIP-chip. In *SAMD4A*, *EXT1*, *ALCAM*, *NFKB1*, and *ZFPM2*, 35 CTCF/RAD21 binding sites were observed. Images were prepared as described in Fig. S5. Six panels (A–F) show representative images of all of the CTCF/RAD21 binding sites and the nearby distributions of Pol II in *SAMD4A*. Gene name, genomic localization, and direction are shown in the lower column. CTL, control siRNA treated; RAD21kd, RAD21 knockdown by siRNA.

Table S1. Sequences of FISH probes used

SAMD4A

	Sequence	Start	End	Strand
1a-1	AAAC(T)GAAATTTGA(T)CTGCCAAAAGG(T)TGTCATTAC(T)CACTCCAGCC(T)GGCA	54136518	54136569	+
1a-1	GCCT(T)CCAAGATGC(T)AATTTCAAGGGG(T)GGGGCAGTTAA(T)GGTGAATAT(T)ATTT	54136756	54136810	+
1a-1	CTGG(T)GGCTAACTCAA(T)CACAAAGACAC(T)GCAGTGTAGT(T)CCAGTAGAA(T)AACA	54136957	54137010	+
1a-1	TTGT(T)GGTGAACATT(T)CAATGTACTG(T)GTCAGTACTG(T)GGCCAAATG(T)TCTG	54137213	54137264	+
1a	C(T)AATTACTCTAA(T)CCTGCCATTCC(T)AAGCCAGGTGA(T)GGCCCTTTGA(T)T	54138070	54138119	+
1a	AGAGTG(T)CTGAGGGGGT(T)AGAGTATTC(T)CTCCAAGTCA(T)CAGCCTTGGT(T)T	54138250	54138299	+
1a	(T)GGTTGCCAGA(T)GCCCTCGGAG(T)TTTCTTCT(T)CTCTAAGGAC(T)TAGCAC	54138392	54138441	+
1a	(T)CCTCTGCTT(T)CTCCTTTGTT(T)CTACATTGCT(T)CCCTCAATTT(T)CTCTCC	54138446	54138495	+
1a	GCA(T)TGATTCACAG(T)TATCTAATCCA(T)GGCCAGCCAG(T)CCCACATTCC(T)	54138589	54138638	+
1b	GTCC(T)GAGATCCAA(T)GCCACTCCCTA(T)TTCATCCCCT(T)CAGGGGTACC(T)G	54219630	54219679	+
1b	C(T)CTATCTGCC(T)CACAAAGGC(T)CGCACCTGTA(T)TTTCAGCCA(T)CCCTAT	54219687	54219736	+
1b	GCCA(T)CCACTTGGGT(T)TCTGGGAGAC(T)CTACTAACC(T)GGCATGGTCC(T)	54219747	54219796	+
1b	GTCA(T)GATCTTAGAAA(T)CCTATGTGAC(T)GCCTTGGGC(T)CCAGACATCAC(T)	54219797	54219846	+
1b	(T)CCTGATCTCA(T)GACCAGCATT(T)TTCTCAATC(T)TTCTTGGAAAC(T)ATCCC	54219852	54219901	+
7	ATTC(T)TCACTGTGT(T)CTTGCTTT(T)CTTCTTTG(T)CGCTTTTTCC(T)GGCC	54302708	54302757	+
7	TGCC(T)AATTCCTCAC(T)ACTGTAACC(T)GCCTAGGTGC(T)CAGTAAGTATC(T)	54302779	54302828	+
7	GCTG(T)GATGGCCTG(T)TTCTTTCTG(T)CTCACCCCT(T)CTCTACTAC(T)GTAA	54302831	54302880	+
7	CTGC(T)CCTCTGTGC(T)CCAGGGCAC(T)CAGGGATGACC(T)TATCAGAGGCT(T)T	54302890	54302939	+
7	C(T)CAGCTGTCT(T)GAGGCCAAGG(T)GGACAGTACC(T)TGCTCTCAA(T)ATCAAT	54303206	54303255	+
1a	GGAA(T)CAATGCCCTTT(T)ACTTGGGACA(T)CTTCCAAG(T)GGTGATTAT(T)TC	54138628	54138677	-
1a	CAGGGA(T)GAAGTACT(T)CCTTTCTG(T)CTGTTCTAG(T)CTCTTCTGG(T)TGC	54138874	54138923	-
1a	CTTG(T)GGGAATGACA(T)GAGTAGCCCCA(T)GGACAGAAACC(T)TCTCAGCCC(T)	54138990	54139039	-
1a	CTTACA(T)TCTCTCAG(T)TCATGCCTG(T)CAGGCTAAA(T)GTCTGGATCC(T)T	54139289	54139338	-
1a	TACT(T)CAAGTGACAC(T)GGGGACCAG(T)ATCAAGGACA(T)CACCTGAGA(T)CT	54139341	54139390	-
1b	TTA(T)GACTGCCGCAA(T)ATCCCCTTT(T)AGGAGGAAA(T)GCAGAGCCAT(T)GA	54218373	54218422	-
1b	AATAAT(T)ATGCTGTCA(T)AACATGCAG(T)ACCTGGGT(T)AGGACTGGCC(T)TA	54218436	54218485	-
1b	(T)GAACATGTGCC(T)GAGGTAGGTAT(T)TCATTAGTGTG(T)GACTCAGGAA(T)CT	54218509	54218558	-
1b	TTCTTA(T)CCCACAATG(T)TTGGGGTTG(T)GGCCTTTAA(T)GTTTGGGAT(T)GTG	54218570	54218619	-
1b	ACTTGC(T)CAGTCCATG(T)CGGCCTCCTG(T)CTCTTCTGT(T)GATTCCAT(T)T	54218723	54218772	-
7	TTATTA(T)CCTCAGTTT(T)AGGAGGCAC(T)GAGTCTGGG(T)AGAGGGGCA(T)G	54304865	54304914	-
7	CGTGTG(T)AGCTGGACA(T)GCAGGGAGCA(T)CACCAACCAA(T)GAGAGGGAC(T)G	54305126	54305175	-
7	CGCC(T)CCTAGCCCACC(T)TGATGCATA(T)TCTTAATGGCA(T)CTGGCTAAGG(T)	54305195	54305244	-
7	G(T)TTGTAAGTCC(T)CAATGACGTT(T)CTGCCACATGA(T)TAACATCCT(T)CTGC	54305261	54305310	-
7	TTCTCC(T)GGTATTAGG(T)CCTTAAAGA(T)TCAGCATCAC(T)GGGAGACAAC(T)	54305325	54305374	-
<i>EDN1</i>				
Intron	Sequence	Start	End	Strand
3	C(T)CA(T)AGCCAGAGGGC(T)CTCCAATAAT(T)TCCAATGTG(T)CCATTTTCA(T)TCTC	12401546	12401597	+
3	CTCC(T)TGTTGGTTTT(T)GGTGTGGTT(T)GATATTGTT(T)GGATTTTGG(T)CCTC	12401445	12401493	+
3	AATG(T)GTGCTTGT(T)ATGACTGCTCCGACAGATGA(T)GAACTAGTG(T)CCAG	12401136	12401412	+
3	CTGC(T)ACAACTCAC(T)CCTGCACAAA(T)GGCTTCAACAC(T)TTGAGCCTAGG(T)TTTT	12401169	12401223	+
3	ATTC(T)AACCTCTA(T)ATCATCACTT(C)GCCTCTCAGT(C)CCACCCTCCA(T)GAGA	12401095	12401149	+

(T) in the sequence denotes amine-C6-dT. Start and end positions in the genomic sequence, and strand sense are indicated.

SAMD4A is located on chromosome 14; start and end positions in the genomic sequence, and strand sense are indicated. (T) in the sequence denotes amine-C6-dT. The location of probes 1a, 1b and 7 is shown in Fig. 3A. Results for sense probes are given in [Tables S2](#) and [S3](#). All antisense probes gave <0.02 foci per nucleus (average of at least 400 nuclei).

EDN1 is located on chromosome 6; start and end positions in the genomic sequence, and strand sense are indicated. (T) in the sequence denotes amine-C6-dT.

Table S2. Size, intensity, number, and percentage of cells with 0, 1, or 2 *SAMD4A* or *EDN1* foci

Intron	Min	Focus					
		size (pixels ± SD)	Relative intensity (% ± SD)	Number per nucleus	Distribution (%)		
					0	1	2
<i>SAMD4A</i>							
1a	0	7 ± 0	48 ± 21	0.04	96.5	3	0.5
	30	9 ± 1	42 ± 15	0.73	50	27.5	22.5
	52.5	9 ± 2	51 ± 18	0.27	79.5	14	6.5
	75	10 ± 3	51 ± 24	0.15	87.5	10	2.5
1b	0	8 ± 0	61 ± 96*	0.03	97.5	2.5*	0
	30	10 ± 4	42 ± 18	0.11	89.5	10	0.5
	52.5	9 ± 2	40 ± 14	0.40	66	28	6
	75	9 ± 3	39 ± 18	0.24	77.5	21.5	1
7	0	9 ± 4	65 ± 39*	0.02	98	2*	0
	30	9 ± 4	35 ± 20	0.08	93.5	5.5	1
	52.5	9 ± 4	47 ± 22	0.13	87.5	12	0.5
	75	8 ± 2	35 ± 11	0.58	57.5	27	15.5
<i>EDN1</i>							
2	0	10 ± 4	22 ± 13	0.40	69	22	9
	30	10 ± 4	31 ± 13	0.32	74	20	6
	52.5	10 ± 3	35 ± 13	0.31	77	15	8
	75	9 ± 2	39 ± 20	0.33	72	23	5

Values were obtained by single labeling from images like that in Fig. 3B. Intensities were normalized relative to fluorescent reference beads (Fig. 3B, inset), and are averages per pixel. For each time and region, 400 cells were analyzed in two independent experiments for *SAMD4A* and 200 cells from one experiment for *EDN1*. Pixels were 103 × 103 nm. *, Few foci were detected, and most were "rogue" (background) foci, as intensities were higher than the average. The fraction of cells with two *SAMD4A* foci at 30 min is higher than expected if we assume that alleles fire independently; this is consistent with the presence of "responding" cells in the population in which both alleles are likely to fire.

For *SAMD4A* at 0 min, the number of foci per nucleus given by the sense probes was not significantly higher than values given by the antisense probes ($P > 0.05$; unpaired, two-tailed, equal variance t test); however, at all other times, sense probes gave significantly more foci ($P < 0.05$). Similarly, significantly more foci per nucleus were observed after 30 min or more, compared with 0 min ($P < 0.05$). For *EDN1*, comparing any two time points, no time point yielded significantly more foci than the other ($P > 0.05$).

Each *SAMD4A* or *EDN1* focus probably contains only one transcript, although we cannot formally prove this. This would be consistent with what is seen with most genes, even highly active ones, in bacteria and yeast (2). Thus, if foci contain variable numbers of transcripts, we would expect them to exhibit a wider range of intensities than that seen. Furthermore, comparison of intensities given by a single focus, bead, and 50-mer (below) indicates that a typical focus contains only three of the five 50-mers in the probe set; if there were > 5 , then there would be > 1 transcript/focus. Moreover, *EDN1* has a different expression pattern but yields foci with similar intensities to *SAMD4A*. Other things being equal, this suggests that the two genes are probably transcribed by the same number of polymerases. Then, the chances are that this number is 1, because, if it were 2 or more, we would have to assume that (by chance) the two genes just happened to be transcribed by the same number of polymerases. In the case of *SAMD4A*, we would also have to assume that polymerases initiate and traverse *SAMD4A* in closely packed groups usually containing roughly the same number of polymerases (as foci have similar intensities). Therefore, the simplest interpretation of this data is that each focus contains only one transcript.

But is the detection system sensitive enough to detect a single transcript? If a single fluorescent 50-mer can be visualized, then there is a good chance that a mixture of five such 50-mers could be used to detect a single RNA molecule. The following experiment indicates that a single 50-mer can be visualized. Five labeled 50-mers, targeting region 1a of *SAMD4A* (Fig. 3A) and tagged with Alexa 488, were mixed with fluorescent reference beads, absorbed to glass slides, images collected, and intensities analyzed. Foci (marking single 50-mers) were seen at the expected frequencies. The experiment was now repeated by adding one or two linkers, linking either the first and second and/or the third and fourth 50-mers; after hybridization, absorption, and imaging, fewer foci were seen. The intensity of each focus was measured and normalized relative to that of the reference beads and the fraction of foci in bins of different intensities determined. When one or two linkers are added and hybridized, the median intensity increases as the linker brings together different 50-mers. Theoretically, the ratio of mean intensity without linker: +1 linker: +2 linkers is 1: 1.25: 1.66. Analysis of the result showed a consistent shift of intensity distribution, and the ratio of mean intensity was 1: 1.28: 1.61 (i.e., similar to the expected value). This confirms that a single 50-mer can be detected.

Table S3. Fraction of cells with the pattern indicated after double labeling using probes against intron 1 (probe 1a; green; Alexa488), and intron 7 (probe 7; red; Cy3)

Time (min)	Fraction of cells with pattern indicated (%)						Yellow focus (any combination) [expected]
							
30	63	29	6	1	0	1	0 [0.2]
120	70	13	1	14	0	2	0 [0.7]
135	57	18	4	18	1	2	0 [1.5]
150	55	14	6	18	3	4	0 [2.1]
165	70	12	4	10	2	2	0 [0.9]

Values were obtained as in Fig. 3C. The expected percentage of yellow foci is the probability of a focus being red multiplied by the probability of a focus being green. Average intensity and size of individual foci are similar at different times (as in Table S2; not shown). A total of 200 cells were analyzed at each time. At all times, most cells possess no foci, and the next largest group possesses foci of one or other color; few cells possess both red and green foci, and no yellow foci were seen. This demonstrates cotranscriptional degradation (presumably after splicing), as foci containing intron 7 (red) never contain intron 1a (green), which must therefore have been removed. With the exception of the 30-min sample (when few polymerases have reached intron 7 and we would not expect to see any yellow foci), the experimentally determined value for the percentage of yellow foci at each of the other four times was significantly different from the expected (random) value ($P < 0.05$; unpaired, two-tailed, equal-variance *t* test). Therefore, our failure to detect any yellow foci at any time is highly significant. The following provided a positive control showing that yellow foci could be detected (Fig. 3C, *inset*). Probes 1a (green) and 1a-1 (red) target RNA sequences in intron 1 of *SAMD4A* lying $\approx 1,000$ nucleotides apart (Table S1) and were hybridized with cells 30 min after stimulation. Of 97 nuclei analyzed, 54 had at least one focus (either red or green), four had one green and one red focus (i.e., signal on separate alleles), and seven had at least one yellow focus (i.e., where $>75\%$ pixels of one color overlapped pixels containing the other color, and *vice versa*).