Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing

Andreas Mayer & L Stirling Churchman

Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. Correspondence should be addressed to L.S.C. (churchman@genetics.med.harvard.edu) or A.M. (mayer@genetics.med.harvard.edu).

Published online 24 March 2016; doi:10.1038/nprot.2016.047

Many features of how gene transcription occurs in human cells remain unclear, mainly because of a lack of quantitative approaches to follow genome transcription with nucleotide precision *in vivo*. Here we present a robust genome-wide approach for studying RNA polymerase II (Pol II)–mediated transcription in human cells at single-nucleotide resolution by native elongating transcript sequencing (NET-seq). Elongating RNA polymerase and the associated nascent RNA are prepared by cell fractionation, avoiding immunoprecipitation or RNA labeling. The 3′ ends of nascent RNAs are captured through barcode linker ligation and converted into a DNA sequencing library. The identity and abundance of the 3′ ends are determined by high-throughput sequencing, which reveals the exact genomic locations of Pol II. Human NET-seq can be applied to the study of the full spectrum of Pol II transcriptional activities, including the production of unstable RNAs and transcriptional pausing. By using the protocol described here, a NET-seq library can be obtained from human cells in 5 d.

INTRODUCTION

Elucidating how genome transcription is regulated in human cells requires genomic tools that reveal the exact location of transcribing RNA polymerases. The application of genome-wide methods that map RNA Pol II density along genomes *in vivo* has uncovered new general features of gene transcription, and it has changed the view of how genomes are transcribed in living cells[1–3]. Originally, it was thought that the recruitment of Pol II to the gene promoter and transcription initiation were the major regulatory steps during RNA synthesis[4]. However, genomic methods that map Pol II occupancy *in vivo* have shown that Pol II is predominantly located in a region 50–100 nt downstream of the transcription start site (TSS) in the majority of active mammalian genes, indicating promoter-proximal pausing[2,3,5,6]. These genome-wide studies provided strong evidence that transcription elongation by Pol II is a key regulatory step in RNA production. In addition, genomic approaches also found that Pol II transcription in the sense orientation is usually accompanied by divergent antisense transcription at the majority of promoters of active mammalian genes[7–9]. Genome-wide Pol II profiling studies revealed that promoter-proximal pausing of Pol II in the sense direction typically co-occurs with promoter-proximal pausing in the upstream antisense orientation in mammalian cells, which uncovered new complexities of genome transcription[6,10–14].

Until now, two genome-wide strategies have mainly been applied to mapping RNA polymerase density across mammalian genomes, as described in more detail in the 'Comparison with other approaches' section below: chromatin immunoprecipitation (ChIP)–based approaches[5,15–20] and transcription run-on–based methods[8,21,22]. Here we provide a step-by-step protocol for NET-seq, which is an alternative method for mapping Pol II genome transcription with single-nucleotide precision and DNA strand specificity in unperturbed human cells. Human NET-seq is a robust Pol II profiling approach that does not require any genetic modification or labeling of the nascent RNA, and its application to different human cell types is straightforward; we have successfully applied the NET-seq approach to four different human cell types, as described in detail in the following section.

Development and overview of human NET-seq

NET-seq allows RNA polymerase transcription elongation to be studied with single-nucleotide precision *in vivo*. The NET-seq methodology was originally developed for budding yeast grown to mid-log phase[23,24]. In the original yeast NET-seq approach, native Pol II transcription elongation complexes were quantitatively purified by IP using a C-terminal 3× FLAG-tagged version of the Pol II subunit Rpb3 (refs. 23,24). A very similar approach was recently used to map the RNA polymerase density across the bacterial genome[25,26]. Although the NET-seq strategy was recently applied to *Drosophila* S2 and Kc cells[27,28], a NET-seq method was not available for mammalian cells. To provide a simple and robust genomic tool for visualizing Pol II transcription with single-nucleotide precision in human cells, we developed a human NET-seq approach[29]. Human NET-seq revealed new general aspects of Pol II transcription, including widespread transcriptional pausing across the expressed human genome and new promoter-proximal transcriptional activity in the antisense orientation[29].

Human NET-seq is a quantitative genomic approach. This requires a strategy to quantitatively (>95% as determined by western blot analysis) purify transcriptionally engaged RNA polymerase from human cells, as well as an efficient sequencing library preparation method (**Fig. 1**).

In contrast to the yeast NET-seq method, in which an epitope-tagged version of Pol II is purified by IP[23,24], we have found that the most efficient and simple way to quantitatively purify transcribing Pol II from human cells is by cellular fractionation. Here we exploit the high degree of stability of the Pol II transcription elongation complex, which is very stable even in the presence of large amounts of salts, detergents and urea[30–33]. Upon fractionation, the cytoplasm and the nucleoplasm are removed (eliminating the bulk of mature mRNA), and elongating Pol II associated with the nascent RNA is greatly enriched in the chromatin fraction

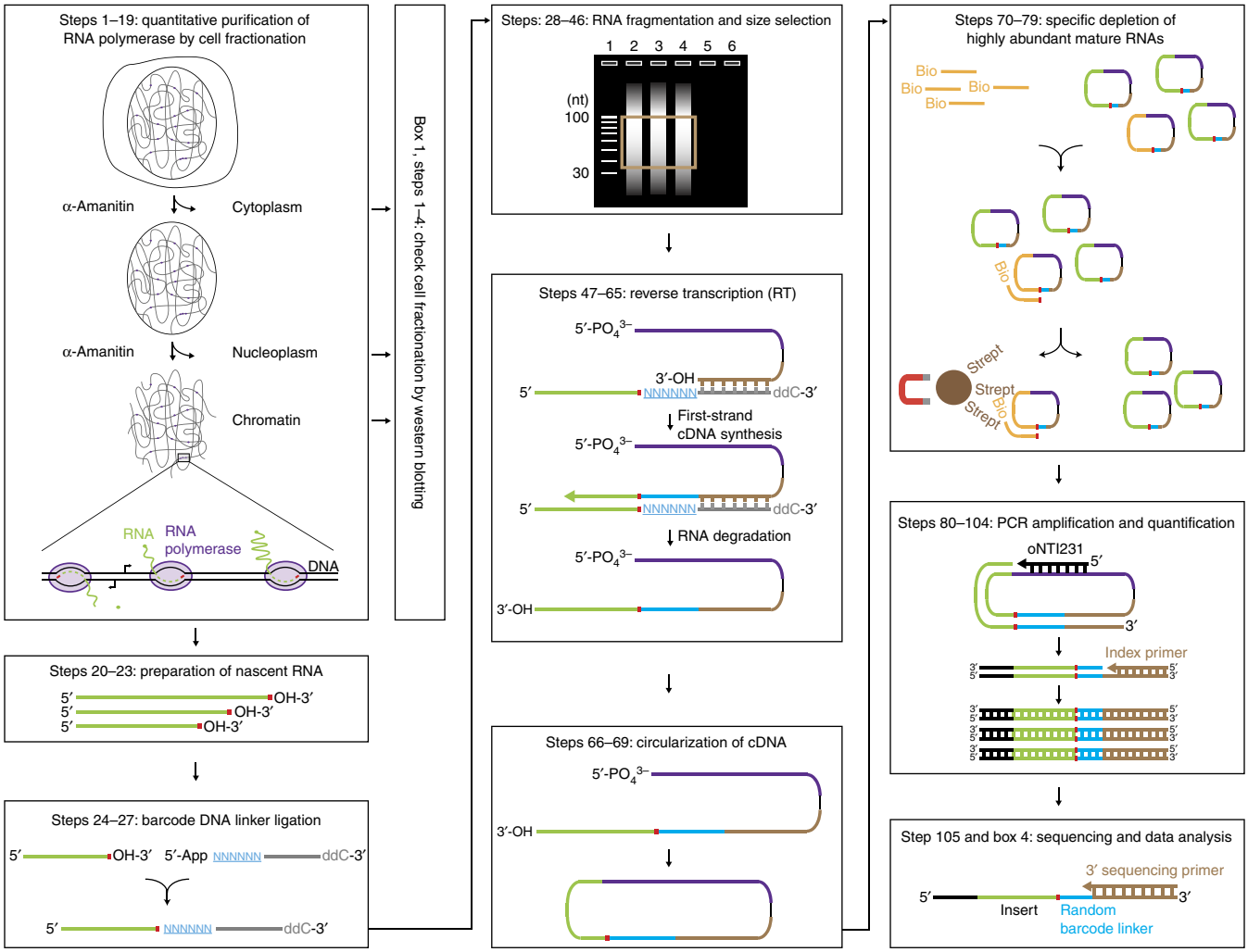

Figure 1 | Schematic overview of the human NET-seq protocol. Each panel describes a key experimental step of the human NET-seq approach. Strept, streptavidin; bio, biotin; ddC, dideoxycytosine; app, adenosine pyrophosphate.

(**Figs. 1** and **2**). Briefly, human cells are lysed in the presence of mild detergents. Nuclei are separated from the cytoplasm by centrifugation through a sucrose cushion. Next, nuclei are washed to remove cytoplasmic remnants. The nucleoplasm is separated from the chromatin in the presence of urea, salt and mild detergents. Urea removes chromatin-bound proteins, except the very stable elongating RNA polymerase and histone proteins[32,34,35]. As the histone proteins are not detached upon urea treatment, the native chromatin can be collected by low-speed centrifugation using a refrigerated benchtop centrifuge. No ultracentrifugation is required. The chromatin-associated nascent RNA is purified, and the remaining DNA is degraded by DNase treatment.

The cell fractionation approach described here is based on protocols published by the Schibler and Black laboratories[32,34,36]. However, the fractionation procedure has been modified and extensively optimized to ensure that:

(i) More than 95% of elongating Pol II, together with the nascent RNA, is captured in the chromatin fraction. This was achieved by a systematic optimization of cell lysis conditions, as well as of the buffers and solutions used for subcellular fractionation.

(ii) Cross-contamination between the subcellular fractions is minimized. This was realized by optimizing buffer and washing conditions during subcellular fractionation.

(iii) The risk of transcriptional run-on and RNA/protein degradation during sample preparation is reduced. This was accomplished by performing the cell fractionation in the presence of α-amanitin (a strong inhibitor of elongating Pol II (ref. 37)), RNase and protease inhibitors and by conducting all experimental steps on ice or at 4 °C. Furthermore, a fast and efficient cell lysis method is used, which leads to rapid depletion of the nuclear nucleotide triphosphate (NTP) pool and further reduces the risk of run-on transcription during cell fractionation.

Apart from its simplicity and robustness, other major advantages of this purification approach are that it avoids restarting Pol II *in vitro* to label and purify nascent RNA, and that it evades possible biases from epitope masking and cross-reactivity when antisera are used for Pol II purification.

The 3′ ends of the purified nascent RNA are then converted into a DNA sequencing library (**Fig. 1**). The human NET-seq library generation is based on the library preparation method that was originally

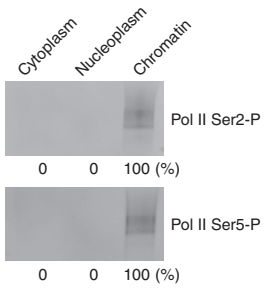

Figure 2 | Representative western blot showing the quantitative purification of elongating RNA Pol II by cell fractionation. Subcellular fractions have been obtained from ~1 × 10^7 HeLa S3 cells. Sample volumes have been adjusted to the cytoplasmic fraction (largest volume) so that the western blot signals between the different fractions can be compared. The subcellular fractions have been probed with antibodies directed against the Pol II CTD Ser2-phosphorylated (top; 3E10 antibody) and CTD Ser5-phosphorylated form (bottom; 3E8 antibody). The numbers below the blots represent the western blot signals in percentages, as obtained by image quantification using ImageJ 1.47v software[64]. The full-size blots of the western analysis are shown in **Supplementary Figure 2**. P, phosphorylated.

developed for budding yeast[23,24], but it includes two important modifications to minimize biases and to increase the efficiency by which the human nascent RNA is converted into a NET-seq library:

- *Barcode DNA linker ligation*. In contrast to the original yeast NET-seq library preparation method[23,24], human NET-seq uses a barcoded DNA linker with a random hexameric sequence at the 5′ end (**Table 1**). The use of this new DNA linker pool has two main advantages. First, mispriming artifacts that can arise during reverse transcription (RT) are strongly reduced[29]. We observed that because of the length of human nascent RNA the RT primer anneals to stretches of complementarity within the nascent RNA that are as short as six nucleotides, leading to frequent mispriming events[29]. By ligating a DNA linker pool that possesses a random hexameric sequence at the 5′ end, mispriming events during RT are markedly decreased[29]. Second, the random hexameric sequence serves as a molecular barcode that allows PCR duplicates and residual mispriming RT events to be identified and bioinformatically removed[29].
- *Depletion of abundant chromatin-associated mature RNAs*. Specific depletion of the most abundant chromatin-associated mature RNAs increases the fraction of sequencing reads that originate from nascent Pol II transcripts[29]. Specific depletion is performed by subtractive hybridization using a set of 20 biotinylated DNA oligos (**Table 2**) that anneal to the 3′ ends of the original mature RNA and are removed with streptavidin-coupled magnetic beads (**Fig. 1**). The specific subtractive hybridization strategy described here is based on the tailored rRNA depletion protocol of the ribosome profiling approach[38].
- Finally, the identity and the abundance of the 3′ ends of the original nascent RNA are determined by high-throughput single-end sequencing. This identifies the last nucleotide that was incorporated into the nascent RNA chain and thus the exact genomic position of the RNA polymerase active site. After these main experimental steps, NET-seq provides a quantitative measure of Pol II density across the human genome with single-nucleotide resolution.

Here we present a step-by-step protocol for performing NET-seq in cultured human cells, from cell lysis through to the quality control measures of the final NET-seq library (**Fig. 1**). The human NET-seq protocol was originally applied to HeLa S3 and HEK293T cells[29]. We have since successfully performed NET-seq on two other human cell types (K562 and MOLT-4 cells; A.M., data not shown), suggesting that the technique is robust and that it works for different human cell types. The detailed description of the experimental steps from cell lysis to the quality control measures of the final NET-seq library will allow our colleagues to adapt this approach in their own laboratories.

Advantages and applications of human NET-seq

NET-seq can be applied to address a broad range of biological questions. The most basic applications of the NET-seq approach can be derived from its main attributes. NET-seq maps Pol II density DNA strand specifically, and therefore it can be used to investigate how transcriptional directionality (sense and antisense transcription) is regulated *in vivo*.

Because of the single-nucleotide resolution, the human NET-seq approach provides a quantitative measure of Pol II occupancy at each nucleotide that is transcribed. Prominent peaks of Pol II density at specific nucleotides or within narrow regions reveal pausing sites, which are locations at which Pol II is detected with a higher probability. Thus, NET-seq is a powerful approach for studying how transcriptional pausing is regulated in the promoter-proximal region, as well as

TABLE 1 | DNA and RNA oligos required for NET-seq library preparation (5′–3′).

DNA oligos	
Barcode DNA linker	/5rApp/$(N)_6$CTGTAGGCACCATCAAT/3ddC
oLSC007 (RT primer)	/5Phos/ATCTCGTATGCCGTCTTCTGCTTG/iSp18/CACTCA/iSp18/TCCGACGATCATTGATGGTGCCTACAG
oNTI231 (reverse primer)	CAAGCAGAAGACGGCATACGA
oLSC006 (sequencing primer)	TCCGACGATCATTGATGGTGCCTACAG
RNA oligo	
oGAB11 (control oligo)	agucacuuagcgauguacacugacugug

$(N)_6$, random hexameric sequence (handmixed by Integrated DNA Technologies); 3ddC, 3′-dideoxycytidine; iSp18, internal 18-atom hexa-ethylenglycol spacer; 5Phos, 5′-phosphate; 5rApp: 5′-riboadenylate. Please note that the 5′-adenine nucleotide of the pre-adenylated barcode DNA linker is a ribonucleotide (rA).

TABLE 2 | Biotinylated DNA oligos for depletion of mature RNAs.

Gene	Transcript	DNA sequence (5′–3′)
RNVU1-1	snRNA	CAGGGGAAAGCGCGAACGCAGTCCCCCACTACC
RNVU1-7	snRNA	CAGGGGAAAGCGCGGACGCAGTCCCCCACT
RNU2-1	snRNA	TGCACCGTTCCTGGAGGTACTGCAATACCAGGTCGATG
RNU4-1	snRNA	GCCTACAGCAGTCTCCGTAGAGACTGTCAAAAATTGCCA
RNU5B-1	snRNA	AGCCTTGTCGGAACAAGGCCTCAAAAAATTAGCTTAAGAC
RNU5E-1	snRNA	AGCCTTGCCAAAGCAAGGCCTCAAAAAATTGGGTTAAG
SNORD3D	snoRNA	CCACTCAGACCGCGTTCTCTCCCTCTCACTCCCCAA
SNORD12B	snoRNA	CAGACAGAAACTGGCTTAGAAAAGTTAGATCAACATAGTCGATC
SNORD12C	snoRNA	CTAACTGGCAAAATATAAGACGTCAGCATTGTCGATCTGATG
SNORD27	snoRNA	CTCAGTAGTAAGATGACATCACTTGAAAGTTCAGCCATATGC
SNORD29	snoRNA	CTCAGGTGTTCATGTATTTTCACTGTCGGTCATAGTGAGC
SNORD31	snoRNA	TACCTTTCAGTCACACATTGATCAGACTGGGGCGGTAT
SNORD80	snoRNA	CATCAGATAGGAGCGAAAGACTTAATATTGCTCATCAGCG
SNORD81	snoRNA	TTGTCATCAAGTAATCAGTGAGAGAGTTCAAGTTGG
SNORD118	snoRNA	GGAGCAATCAGGGTGTTGCAAGACCTGATTACGCAGAG
RNA5S1	rRNA	CACCCGGTATTCCCAGGCGGTCTCCCATCCAAGTAC
RNA5-8S5	rRNA	AGCGACGCTCAGACAGGCGTAGCCCCGGGAGG
RNA28S5	rRNA	ACAAACCCTTGTGTCGAGGGCTGACTTTCAATAGATCGCA
MT-TM (ID: 4569)	Mitochondrial tRNA	GTACGGGAAGGGTATAACCAACATTTTCGGGGTATGGG
MT-TV (ID: 4577)	Mitochondrial tRNA	TGGTCAGAGCGGTCAAGTTAAGTTGAAATCTCCTAAGTG

DNA oligos possess a 5′ biotin-tetraethylene glycol (TEG) and are HPLC purified. The ID numbers listed for *MT-TM* and *MT-TV* are GeneIDs.

throughout gene-body regions of genes. In addition, the high spatial resolution allows Pol II transcriptional activities (e.g., sense and different types of antisense transcription[29]) that arise in close proximity to each other to be distinguished. Human NET-seq now provides the opportunity to investigate how these activities are co-regulated in normal cells and upon disease conditions or treatments.

Furthermore, NET-seq captures the RNA as it is being produced, and human NET-seq has been shown to detect unstable RNA species[29] such as promoter upstream transcripts[7] and upstream antisense RNAs[39], as well as enhancer-derived noncoding RNAs[40,41]. Human NET-seq can therefore be used to study these transcripts before they are turned over by cellular RNA degradation pathways[42–44]. Standard RNA-seq methods mainly capture steady-state mRNA levels of fully processed mature RNAs and thus struggle to detect unstable transcripts.

Application of human NET-seq revealed strong Pol II pauses at the boundaries of exons that are retained during splicing but not at exons that are subsequently removed by the spliceosome[29]. This indicates that Pol II recognizes exons with different processing fates where Pol II slows down at exons that are retained in the final mRNA. The determinants of Pol II pausing at exon boundaries and how Pol II transcription and splicing are coupled *in vivo* remain to be determined. NET-seq will help elucidate these links, and it will shed new light on the regulatory mechanisms that underlie these fundamental biological processes.

Although the NET-seq approach has been established and optimized for the analysis of Pol II transcription, the adaptation of the protocol to the RNA Pol I and RNA Pol III transcriptional systems should be straightforward. Using the human NET-seq protocol described here, we detect Pol I and Pol III transcription[29]. However, the NET-seq library preparation may require further adaptations to study Pol I and Pol III transcription in an unbiased way. For example, the extended secondary structures in Pol I and Pol III transcripts may pose a problem for the RT reaction.

Limitations of human NET-seq

To fully comprehend the potential applications of NET-seq and to extract meaningful conclusions, it is important to be cognizant of its limitations. NET-seq reveals the exact position of Pol II by determining the 3′ ends of Pol II–associated nascent RNA. Although Pol II involved in elongation can be localized with NET-seq, the exact genomic positions of Pol II that is part of the preinitiation complex (PIC) at promoter regions cannot be mapped because PIC formation occurs before nascent RNA is produced. In addition, Pol II involved in transcription initiation around the TSS of genes cannot be mapped, because the synthesized nascent transcripts are too short to produce mappable sequencing reads; minimum read lengths of ~18 nt are required for unique mapping to the human reference genome.

Purification of transcriptionally engaged RNA polymerase will co-purify nascent RNA and RNA processing intermediates from splicing and 3′-end cleavage. Thus, the NET-seq signal at the last nucleotide of introns and exons can either be because of elongating Pol II or splicing intermediates. Similarly, it is not clear whether the NET-seq signal that is obtained at the polyadenylation (pA) site, where the transcript is cleaved and polyadenylated, relates to transcriptionally engaged Pol II or to a 3′-end processing intermediate. Therefore, NET-seq signals at these single-nucleotide positions are computationally removed by discarding the corresponding sequencing reads.

Purification of transcriptionally engaged Pol II will also co-purify some fully processed mature RNAs. Abundant noncoding mature

RNAs, such as those associated with the spliceosome, are removed by specific depletion (PROCEDURE Steps 70–79). Although the vast majority of mature mRNAs are experimentally removed by discarding the cytoplasmic and nucleoplasmic fractions (PROCEDURE Steps 8–18), as determined by western blot analysis and subcellular RNA sequencing[29], some mature mRNAs are likely to be captured in the final NET-seq library. These mature mRNAs can be polyadenylated or nonpolyadenylated at their 3′ ends. As this protocol is designed to sequence the 3′ ends of RNAs, the polyadenylated transcripts are not a problem. First, the long polyadenylated tracks are technically challenging for RT reactions, PCRs and sequencing platforms to capture. Second, any sequencing reads that arise from polyadenylated mature RNAs do not align to the human reference genome because of polyadenine stretches, and they are removed bioinformatically. NET-seq reads that arise from nonpolyadenylated mature mRNAs align exactly to the pA sites corresponding to the pA site for the transcript (see the 'Experimental design' section for more information). The identification of this subset of mature mRNAs and the subsequent removal of the corresponding sequencing reads depends on the annotation of the pA sites, which has been strongly improved over the past years and comprehensive maps are available[12,45]. For genes in which the pA site information is lacking, these mature mRNAs might introduce low levels of background signals. These unspecific signals are restricted to a few single-nucleotide positions. Therefore, with the focus by NET-seq on 3′-end sequencing, we expect that these unspecific signals occur at a much lower level compared with background signals that are usually observed in IP-based genomic methods for which the unspecific pull-down of RNAs represents an inherent problem when the entire molecule is fragmented and sequenced.

Our conclusion that mature mRNAs have negligible impact, if any, on NET-seq results is based on two experimental results. First, we performed human NET-seq upon conditions in which productive Pol II transcription elongation is blocked, by inhibiting the Cdk9 subunit of the positive transcription elongation factor b (P-TEFb)[29]. P-TEFb represents a key player in the release of promoter-proximally paused Pol II (refs. 46,47). In agreement with previous studies[5,39], Cdk9 inhibition resulted in a genome-wide loss of productive elongating Pol II in both sense and antisense orientation[29], indicating that NET-seq quantitatively captures transcribing Pol II. If mature mRNAs were influencing the NET-seq signal, we would not expect a complete reduction in productive elongating Pol II levels upon 1 h of P-TEFb inhibition. Second, key observations of mammalian Pol II transcription made by human NET-seq, such as widespread promoter-proximal pausing and divergent transcription originating from bidirectional promoters, are consistent with observations by many laboratories using a broad range of genomic methods[5–9,18,48–50]. Nevertheless, mature RNAs are likely to introduce some background signal, which is important to consider when inspecting NET-seq data.

The human NET-seq strategy described here requires 1×10^7 cells as an input, and it therefore works best for cells that can be obtained in larger quantities. Current efforts aim to reduce the amount of cells that are required per NET-seq experiment.

Comparison with other approaches

Before the development of NET-seq, two main strategies existed to determine the genome-wide location of Pol II and to study gene transcription *in vivo*: Pol II ChIP and transcription run-on. The Proudfoot and Carmo-Fonseca laboratories[51] have also recently developed a variant of the NET-seq approach for mammalian cells, called mNET-seq.

Pol II ChIP. In this approach, chromatin is cross-linked and Pol II–bound DNA is purified by IP typically using antibodies directed against Pol II. The DNA is converted into a library that is either hybridized to microarrays (ChIP-chip[15,16,52–58]) or applied to high-throughput sequencing (ChIP-seq[5,17,18]). Although these approaches have revealed the genomic Pol II density in various model organisms, including yeast and humans, the spatial resolution is usually limited to >200 bp because of the fragmentation approaches typically used[59]. ChIP data usually suffer from high levels of background signals due to the unspecific pull-down of DNA that can arise from cross-reactive antibodies or the unspecific binding of DNA to beads used in IP experiments. Furthermore, these methods lack DNA strand specificity, and the transcriptional state of Pol II remains unclear. To clarify whether Pol II is transcriptionally engaged, ChIP methods need to be complemented by other approaches such as permanganate footprinting assays[60]. Recent improvements in the genome-wide ChIP approach could greatly increase spatial resolution and decrease nonspecific background signals[19,20]. Although these new ChIP approaches lack single-nucleotide resolution and DNA strand specificity, they provide an opportunity to determine the location of Pol II during preinitiation and initiation of transcription complementing NET-seq and transcription run-on data.

Transcription run-on. To date, two transcription run-on approaches have been developed to map genomic Pol II density: global run-on-sequencing (GRO-seq)[8] and precision nuclear run-on and sequencing (PRO-seq)[21,22]. In both methods, transcription is halted and then restarted *in vitro* in the presence of modified nucleotides that enable the purification of nascent RNA. The purified nascent RNA is then converted into a DNA sequencing library and subjected to high-throughput sequencing. These methods provide a DNA strand–specific quantitative measure of transcriptionally engaged Pol II, and they have provided new insights into gene regulation in various model organisms, including human cells and *Drosophila* cells. Similarly to human NET-seq, GRO-seq and PRO-seq require nascent RNAs of a minimum length (at least ~18 nt) so that the corresponding sequencing reads can uniquely align to the human reference genome. Therefore, these methods do not provide any information about the localization of Pol II during preinitiation or transcription initiation. Furthermore, transcription run-on–based methods are strongly dependent on an efficient restart of transcription under nonphysiological conditions; this process is sensitive to the experimental setup and the transcriptional status of Pol II (ref. 61). The pool of Pol II enzymes that are recovering from transcriptional pauses in which the nascent RNA is misaligned in the polymerase active site cannot be restarted *in vitro*; these polymerases therefore escape detection by nuclear run-on–based methods[27]. Thus, comparison of transcription run-on data with NET-seq data that maps all Pol II—paused, recovering and not paused—enables the identification of where Pol II is in the process of pause recovery[27].

mNET-seq. mNET-seq uses IP to purify Pol II (and its different C-terminal domain (CTD)-phosphorylated forms) with the associated nascent RNA from HeLa cells[51], and thus it is capable of mapping subpopulations of Pol II across the genome. The ability to analyze Pol II complexes in a phosphorylation-dependent manner comes with the known limitations of an IP experiment, such as cross-reactivity of antibodies and unspecific binding of RNAs to beads, which are likely to result in unspecific background signals. It also comes at the cost of additional procedural steps. mNET-seq requires that Pol II transcription complexes be released from the chromatin by a micrococcal nuclease (MNase) digest; this process is inefficient, and it may release a biased subset of transcriptionally engaged Pol II (ref. 51). Subsequently, the purified nascent RNA must be phosphorylated at its 5′ end before entering a sequencing library protocol. In contrast, the human NET-seq protocol quantitatively purifies human Pol II elongation complexes by cellular fractionation, and it uses an optimized library preparation method to reduce experimental biases (**Fig. 1**).

Experimental design

Each experimental step of the protocol described in the PROCEDURE has been optimized, and we recommend always including controls to monitor the success of key steps before sequencing a NET-seq library (**Figs. 2** and **3**). Below we describe the steps that need particular attention before adapting the human NET-seq methodology. It is important that each step of the library construction works with high efficiency so that the DNA sequencing library quantitatively reflects the original pool of nascent RNA molecules.

Cell fractionation and RNA purification. The cell fractionation approach used in this protocol has been optimized for an input of 1×10^7 cells. If more input is required, we recommend performing fractionation on parallel samples rather than increasing the amount of cells per sample. The use of more cells per sample reduces the cell lysis efficiency and increases the risk of cross-contamination between the different subcellular fractions.

To avoid run-on transcription, the fractionation is performed in the presence of α-amanitin, which is a very strong inhibitor of elongating Pol II (refs. 37,62). Furthermore, all steps are performed on ice or at 4 °C. To avoid the risk of RNA and protein degradation, the fractionation is conducted in the presence of RNase and protease inhibitors.

For RNA purification, it is important to use the miRNeasy kit instead of the RNeasy kit (Qiagen) to minimize RNA size biases.

Quality controls for cell fractionation. Cell fractionation represents the first fundamental step of the human NET-seq approach, and monitoring its success is crucial before proceeding with the NET-seq library preparation. We recommend performing western blot analysis by probing the subcellular fractions with antibodies directed against the elongating form of Pol II. We suggest using the well-characterized antibodies directed against the CTD Ser2- (3E10) and CTD Ser5 (3E8)-phosphorylated forms of Pol II, which were originally developed in the laboratories of Eick and Kremmer[63]. Approximately ≥95% of elongating Pol II should be captured in the chromatin fraction (**Fig. 2**).

When applying the cell fractionation approach to a new cell type, we also recommend performing the following control experiments. To check for potential cross-contamination between the subcellular fractions, we suggest probing the different fractions on a western blot with antibodies raised against subcellular marker proteins (chromatin fraction: FL-126 directed against histone 2B; nucleoplasmic fraction: C-18 raised against U1 snRNP70; cytoplasmic fraction: 6C5 recognizing GAPDH; please also see MATERIALS for more information). The cytoplasmic and nucleoplasmic protein markers should be absent from the chromatin fraction. In addition, the enrichment of nascent intron-containing RNAs in the chromatin fraction compared with the nucleoplasmic and cytoplasmic fractions can be monitored by subcellular RNA-seq or by RT coupled with quantitative real-time PCR (RT-qPCR) using specific primer pairs directed against intronic regions of the nascent RNA. We performed subcellular RNA-seq for HeLa S3 and HEK293T cells, revealing that intron-containing nascent RNA is greatly enriched in the chromatin fraction[29]. The quantity and quality of the purified RNA can be monitored using a NanoDrop spectrophotometer.

Barcode DNA linker ligation. Although other adapter and/or barcoding strategies can be used, we strongly recommend using the barcode DNA linker design described here and the truncated version of the T4 RNA ligase 2 (from New England BioLabs (NEB)) for NET-seq library construction. High-throughput sequencing requires libraries that contain specific sequences, such as a binding site for the sequencing primer. In the NET-seq protocol, the DNA linker provides the sequence in which the RT primer anneals, which in turn introduces the binding site for the sequencing primer. Human NET-seq uses a barcode DNA linker that is preadenylated, and which possesses a random hexameric sequence at its 5′ end (**Table 1**). The barcode DNA linker is ligated to the 3′-hydroxyl group of nascent RNA molecules by a truncated version of the T4 RNA ligase 2. This truncated enzyme cannot adenylate the 5′ ends of the RNA substrate, thus reducing background ligation. In addition, the DNA linker possesses a dideoxy cytosine at the 3′ end to prevent its extension by the reverse transcriptase during the first-strand cDNA synthesis. This keeps the RT directional, and it is important for the strand specificity of the NET-seq library.

RNA fragmentation. RNA fragmentation by partial alkaline hydrolysis should result in a homogeneous RNA pool, with most of the molecules in the range of 35–100 nt (**Fig. 3a**). We have observed that the optimal fragmentation time can vary with different batches of alkaline fragmentation solution, with the preferred fragmentation time usually between 10 and 40 min. We recommend that the optimal RNA fragmentation time be determined for each new batch of fragmentation solution. Even when the same batch of fragmentation solution is used, we suggest validating the fragmentation time every 4 months to minimize the risk of over-fragmentation or underfragmentation of the RNA pool, as well as to maximize the reproducibility of the human NET-seq data.

Specific depletion of highly abundant mature RNAs. Although the vast majority of fully spliced mature RNAs are removed with the cytoplasmic and nucleoplasmic fractions, some mature RNA species are present in the chromatin fraction. These RNAs include chromatin-associated mature RNAs, such as small nuclear RNAs (snRNAs), which are an integral part of spliceosomes. Although the fraction of chromatin-associated mature RNAs is further

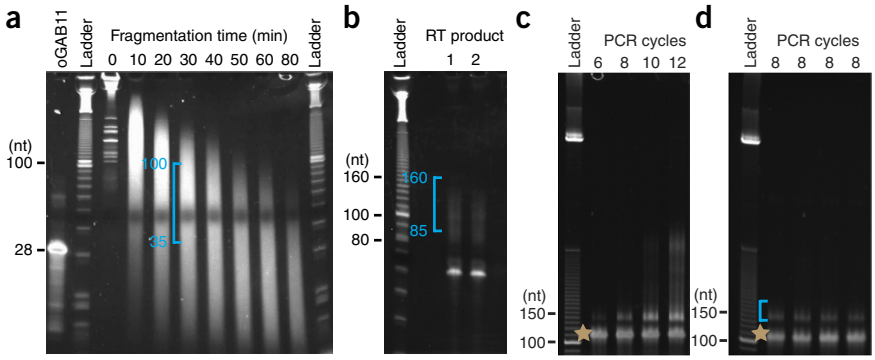

Figure 3 | Representative gels showing the size selections during NET-seq library preparation. (**a**) Optimization of RNA fragmentation time and size selection of fragmented RNA. The RNA was obtained from HeLa S3 cells. Approximately 1 µg of RNA was used per fragmentation reaction. The fragmented RNA was separated on a 15% (wt/vol) TBE-urea gel. The optimal fragmentation time was 30 min (lane 6). The blue bracket indicates the region that was excised from the gel. The fragmented RNA between 35 and 100 nt was selected. (**b**) Size selection of the cDNA after RT. The RT product was separated on two lanes of a 10% (wt/vol) TBE-urea gel. The blue bracket indicates the region that was excised from the gel. The cDNA in the range of 85–160 nt was selected. (**c**) Optimization of the final PCR amplification reaction. PCRs were stopped after 6, 8, 10 and 12 amplification cycles. The PCR products obtained from four different test amplifications were separated on an 8% (wt/vol) TBE gel. The PCR product that corresponds to the NET-seq library runs at ~150 nt. Note that PCR products in the higher molecular range were obtained by ten or more amplification cycles, indicating overamplification. For this NET-seq library, the optimal number of amplification cycles was eight (lane 3). The lower band (~120 nt, brown asterisk) corresponds to a PCR product obtained from empty circles that arise from unextended RT primers. (**d**) Purification of PCR products. The PCR products of four final PCRs using eight amplification cycles were separated on an 8% (wt/vol) TBE gel. The excised band is indicated by a blue bracket. The PCR product that arises from empty circles is labeled by a brown asterisk as in **c**.

reduced by urea treatment during the separation of the chromatin from the nucleoplasm, some fully processed RNAs are not completely removed and are part of the final NET-seq library, as described in the 'Limitations of human NET-seq' section above. NET-seq reads that are due to abundant noncoding mature RNA species, such as spliceosomal-associated RNAs, markedly reduce the amount of informative sequencing data; therefore, it is preferable to remove them before sequencing. Sequencing of our first NET-seq libraries generated from HeLa S3 cells identified the most abundant mature RNAs in the chromatin fraction, and biological replicates identified the same RNAs as being the most abundant in different HeLa S3 NET-seq libraries. On the basis of these sequencing results, we designed biotinylated DNA oligos complementary to the 3′ regions of the 20 most abundant mature RNAs in our NET-seq libraries (**Table 2**). To deplete abundant mature RNAs, these DNA oligos are hybridized to first-strand cDNA during library preparation; duplexes formed between the biotinylated oligos and their targets are removed by streptavidin-coupled magnetic beads. The same set of depletion oligos has been successfully applied during NET-seq library preparations from HEK293T cells. The set of depletion oligos might need to be amended for other cell types, as chromatin-associated mature RNAs can vary among different cell and tissue types.

PCR amplification. Before high-throughput sequencing on an Illumina platform, the NET-seq library is amplified by a limited number of PCR cycles using an Illumina index forward primer and a NET-seq–specific reverse primer (oNTI231; **Table 1**). To avoid overamplification, the minimum number of PCR cycles required to obtain a NET-seq library needs to be determined for each library preparation (see Step 88 of the PROCEDURE and **Fig. 3c** for more information). We typically test between 6 and 12 PCR amplification cycles (**Fig. 3c**). The NET-seq library is then amplified with the minimal amount of PCR cycles (**Fig. 3d**).

Quality controls for library preparation. The human NET-seq library construction requires several enzymatic reactions, such as ligation, RT, circularization and PCR amplification (**Fig. 1**). We have observed variations between different enzyme lots, and thus we suggest keeping track of the lot numbers for each enzyme. We also include a purified RNA oligonucleotide (oGAB11; **Table 1**)[24] sample, which is processed in parallel to the library sample (Steps 24–27, 30–69 and 80–88 of the PROCEDURE), as a control to monitor the success of each enzymatic reaction. In particular, we recommend using this control RNA oligonucleotide to determine the ligation efficiency. Ligation of the DNA linker to the 3′ end of the nascent RNA is one of the most critical steps in the NET-seq protocol, as only nascent RNA molecules that ligate with the DNA linker are captured in the NET-seq library. To ensure a high sensitivity of the NET-seq approach and to minimize biases, the ligation efficiency should be at least 90% (**Supplementary Fig. 1**). After the DNA linker has been ligated to oGAB11, the ligation product can be used to monitor the success of first-strand cDNA synthesis by SuperScript III reverse transcriptase; if successful, a distinct band should be observed at ~100 nt. This oGAB11 cDNA can then be used to determine the circularization efficiency by the CircLigase. Alternatively, circularization efficiency can be monitored by circularizing the RT primer oLSC007 (**Table 1**). The circularized product migrates more slowly than the linear molecule, and it can be visualized by gel electrophoresis. oGAB11 cDNA also serves as a positive control during PCR amplification by the Phusion DNA polymerase.

MATERIALS

REAGENTS

- HeLa S3 cells (American Type Culture Collection (ATCC), cat. no. CCL-2.2) or HEK293T cells (ATCC, cat. no. CRL-11268)
  ! CAUTION It is important to regularly check cell lines to ensure that they are authentic and are not infected with mycoplasma. Handle human cell lines according to the supplier's instructions. Work in a fume hood, use sterile equipment and wear gloves to minimize the risk of contamination.
- DMEM (Life Technologies, cat. no. 11995)
- FBS (Life Technologies, cat. no. 16000)

- Penicillin-streptomycin (10,000 U/ml; Life Technologies, cat. no. 15140) **! CAUTION** Penicillin-streptomycin is toxic. Handle the medium containing these antibiotics with care and dispose of waste according to institutional regulations.
- RNase/DNase-free H_2O (Life Technologies, cat. no. 10977-015)
- PBS, 1× (Life Technologies, cat. no. 10010-023)
- TBE buffer, 10× (Life Technologies, cat. no. 15581-044)
- α-Amanitin (Sigma-Aldrich, cat. no. A2263) **! CAUTION** α-Amanitin is toxic. Handle solutions containing α-amanitin with care and dispose of waste according to institutional regulations.
- DTT (0.1 M, part of the SuperScript III first-strand synthesis system; Life Technologies, cat. no. 18080-051) **! CAUTION** DTT is toxic and corrosive, and it is an irritant. Handle solutions containing DTT with care and dispose of waste according to institutional regulations.
- NP-40, molecular biology grade (Life Technologies, cat. no. 28324) **! CAUTION** NP-40 is an irritant. Handle solutions containing NP-40 with care and dispose of waste according to institutional regulations.
- Triton X-100, molecular biology grade (Sigma-Aldrich, cat. no. T9284) **! CAUTION** Triton X-100 is harmful, and it is an irritant. Triton X-100 is hazardous to the environment. Handle solutions containing Triton X-100 with care and dispose of waste according to institutional regulations.
- Tween 20, molecular biology grade (Sigma-Aldrich, cat. no. 274348)
- Sucrose, molecular biology grade (Sigma-Aldrich, cat. no. S0389)
- Glycerol, molecular biology grade (Sigma-Aldrich, cat. no. G5516)
- Urea, molecular biology grade (Sigma-Aldrich, cat. no. U6504)
- Sodium carbonate anhydrous, proteomics grade (VWR, cat. no. M138) **! CAUTION** Sodium carbonate is an irritant. Handle solutions containing sodium carbonate with care and dispose of waste according to institutional regulations.
- Sodium bicarbonate (VWR, cat. no. 3509)
- Sodium acetate, RNase-free (3 M; Life Technologies, cat. no. AM9740)
- NaCl, RNase-free (5 M; Life Technologies, cat. no. AM9760G)
- EDTA, RNase-free (0.5 M; Life Technologies, cat. no. AM9260G) **! CAUTION** EDTA is an irritant. Handle solutions containing EDTA with care and dispose of waste according to institutional regulations.
- Tris-HCl, RNase-free (1 M, pH 7.0; Life Technologies, cat. no. AM9850G)
- Tris-HCl, RNase-free (1 M, pH 8.0; Life Technologies, cat. no. AM9855G)
- HEPES, RNase-free (1 M, pH 7.5; Teknova, cat. no. H1035)
- NaOH solution (1.0 N; Sigma-Aldrich, cat. no. S2770) **! CAUTION** NaOH-containing solutions are corrosive. Handle solutions containing NaOH with care and dispose of waste according to institutional regulations.
- HCl, hydrochloric acid concentrate (1.0 mol for 1 liter of 1.0 N standard solution; Sigma-Aldrich, cat. no. 38282) **! CAUTION** HCl-containing solutions are corrosive. HCl is an irritant. Handle solutions containing HCl with care and dispose of waste according to institutional regulations.
- SUPERase.In (Life Technologies, AM2696)
- Protease inhibitor mix cOmplete, EDTA-free (Roche, 11873580001) **! CAUTION** This mix is an irritant. Handle solutions containing the protease inhibitor mix with care and dispose of waste according to institutional regulations.
- Benzonase nuclease (Sigma-Aldrich, cat. no. E1014)
- GlycoBlue (15 mg/ml; Life Technologies, cat. no. AM9515)
- Isopropanol, molecular biology grade (Sigma-Aldrich, cat. no. 278475) **! CAUTION** Isopropanol is highly flammable and volatile. Isopropanol is an irritant. Handle solutions containing isopropanol with care and dispose of waste according to institutional regulations.
- Ethanol, molecular biology grade (VWR, cat. no. V1016) **! CAUTION** Ethanol is highly flammable and volatile. Ethanol is an irritant. Handle solutions containing ethanol with care and dispose of waste according to institutional regulations.
- SDS (Sigma-Aldrich, cat. no. L3771) **! CAUTION** SDS is corrosive and flammable. SDS is an irritant. Handle solutions containing SDS with care and dispose of waste according to institutional regulations.
- Pol II CTD Ser2-P antibody (3E10; Active Motif, cat. no. 61083) **▲ CRITICAL** We highly recommend using this well-characterized Pol II CTD Ser2-P antibody for western blot control experiments (Please see Step 19 of the PROCEDURE and **Box 1** for more details.).
- Pol II CTD Ser5-P antibody (3E8; Active Motif, cat. no. 61085) **▲ CRITICAL** We highly recommend using this well-characterized Pol II CTD Ser5-P antibody for western blot control experiments (Please see Step 19 of the PROCEDURE and **Box 1** for more details.).
- Histone 2B antibody (FL-126; Santa Cruz Biotechnology, cat. no. sc-10808)
- GAPDH antibody (6C5; Life Technologies, cat. no. AM4300)
- U1 snRNP70 antibody (C-18; Santa Cruz Biotechnology, cat. no. sc-9571)
- Chloroform, molecular biology grade (Sigma-Aldrich, cat. no. 288306) **! CAUTION** Chloroform is volatile and toxic. Chloroform is an irritant. Handle solutions containing chloroform with care and dispose of chloroform waste according to institutional regulations.
- miRNeasy mini kit (50; Qiagen, cat. no. 217004) **! CAUTION** The RWT buffer is corrosive, and it is an irritant. The QIAzol lysis reagent is toxic and corrosive. Use personal protective equipment when handling this kit, and be sure to dispose of waste according to institutional regulations. **▲ CRITICAL** It is important to use the miRNeasy mini kit instead of the RNeasy mini kit to avoid RNA length biases.
- RNase-free DNase set (50; Qiagen, cat. no. 79254)
- Polyethylene glycol (PEG)8000, molecular biology grade (part of T4 RNA ligase 2, truncated; New England BioLabs (NEB), cat. no. M0242S) **▲ CRITICAL** Store PEG8000 solution at −20 °C.
- DMSO, molecular biology grade (Sigma-Aldrich, cat. no. D8418)
- T4 RNA ligase buffer (part of T4 RNA ligase 2, truncated; NEB, cat. no. M0242S)
- T4 RNA ligase 2, truncated (NEB, cat. no. M0242S) **▲ CRITICAL** We recommend keeping track of the enzyme lot, as the ligase activity can vary between different batches.
- Orange G (Sigma-Aldrich, cat. no. O3756)
- SYBR Gold nucleic acid gel stain (10,000× concentrate; Life Technologies, cat. no. S-11494) **! CAUTION** SYBR Gold nucleic acid gel stain is flammable. Nucleic acid stains are usually mutagenic. Use personal protective equipment when handling nucleic acid gel stains, and dispose of waste according to institutional regulations.
- RNA control ladder (0.1–2 kb; Life Technologies, cat. no. 15623-100)
- DNA control ladder (10 bp; Life Technologies, cat. no. 10821-015)
- TBE-urea (TBU) denaturing sample buffer, 2× (Life Technologies, cat. no. LC6876)
- dNTP mix (10 mM; Life Technologies, cat. no. 18427-013)
- First-strand buffer, 5× (part of SuperScript III first-strand synthesis system; Life Technologies, cat. no. 18080-051)
- SuperScript III first-strand synthesis system (Life Technologies, cat. no. 18080-051)
- CircLigase ssDNA ligase (100 U/μl; Epicentre, cat. no. CL4111K)
- CircLigase reaction buffer, 10× (part of CircLigase ssDNA ligase; Epicentre, cat. no. CL4111K)
- ATP, 1 mM (part of CircLigase ssDNA ligase; Epicentre, cat. no. CL4111K)
- $MnCl_2$, 50 mM (part of CircLigase ssDNA ligase; Epicentre, cat. no. CL4111K) **! CAUTION** $MnCl_2$ is toxic and hazardous to the environment. Handle solutions containing $MnCl_2$ with care, and dispose of waste according to institutional regulations.
- Dynabeads MyOne streptavidin C1 (Life Technologies, cat. no. 65001)
- Saline–sodium citrate (SSC), 20× (Life Technologies, cat. no. AM9763)
- Phusion high-fidelity (HF) DNA polymerase (2,000 U/ml; NEB, cat. no. M0530S)
- Phusion HF buffer, 5× (part of Phusion high-fidelity DNA polymerase; NEB, cat. no. M0530S)
- Qubit dsDNA high-sensitivity (HS) assay kit (Life Technologies, cat. no. Q32851)
- High sensitivity DNA analysis kit (Agilent Technologies, cat. no. 5067-4626)

EQUIPMENT

- Tissue culture dish, 15 × 2.5 cm (VWR, cat. no. 25383-103)
- Cell scrapers (GeneMate, cat. no. T-2443-2)
- Scalpels (Electron Microscopy Sciences, cat. no. 72042-11) **! CAUTION** Handle sharps with care. Dispose of sharps according to institutional regulations.
- RNase/DNase-free PCR tubes, 0.2 ml (Corning, cat. no. 3745)
- RNase/DNase-free microcentrifuge tubes, 0.5 ml (Life Technologies, cat. no. AM12350)
- RNase/DNase-free microcentrifuge tubes, 1.5 ml (Life Technologies, cat. no. AM12450)
- Microcentrifuge tube filter: Costar Spin-X centrifuge tube filters (Sigma-Aldrich, cat. no. CLS8162-96EA)
- RNase/DNase-free centrifuge tubes, 15 ml (VWR, cat. no. 89039-670)
- Needle, 20G (Becton Dickinson and Company (BD), cat. no. 305175) **! CAUTION** Handle sharps with care. Dispose of sharps according to institutional regulations.
- Needle, 22G (BD, cat. no. 305155) **! CAUTION** Handle sharps with care. Dispose of sharps according to institutional regulations.
- Syringe, 1 ml (BD, cat. no. 309628)

Box 1 | Checking cell fractionation by western blotting ● TIMING 5 h (plus overnight and ~2.5 h the next day)

1. Adjust the volumes of the nucleoplasmic (from PROCEDURE Step 18) and chromatin fractions (from PROCEDURE Step 19) to the volume of the cytoplasmic fraction (PROCEDURE Step 12) by adding 1× PBS. The volume of the cytoplasmic fraction is typically ~500 µl. Mix well.

▲ **CRITICAL STEP** It is important to completely resuspend the sticky chromatin pellet. We found that this can be accomplished in the presence of 250 U of Benzonase.

2. Add 50 µl of 2× SDS buffer to 50 µl of each subcellular fraction, and boil the samples at 95 °C for 5 min.

■ **PAUSE POINT** Samples can be stored for months at –20 °C.

3. Load 10 µl of each boiled sample per lane, and separate it by standard SDS-PAGE.

4. Probe the membrane overnight at 4 °C with antibodies directed against transcribing RNA Pol II: Ser2 CTD phosphorylated form (3E10, 1:1,000 dilution); Ser5 CTD phosphorylated form (3E8, 1:1,000 dilution; **Fig. 2** and **Supplementary Fig. 2**). We recommend using these Pol II–specific antibodies, as their CTD specificities have been extensively characterized[63].

Optionally, the membrane can also be probed with antibodies raised against the subcellular marker proteins GAPDH (cytoplasm), U1 snRNP70 (nucleoplasm) and histone 2B (chromatin). We recommend probing the subcellular fractions with these antibodies when a new cell type is used.

? TROUBLESHOOTING

- Cell counter Countess (Life Technologies, cat. no. C10227)
- Refrigerated centrifuge 5810R (VWR, cat. no. 89305-180)
- Refrigerated microcentrifuge 5424R (VWR, cat. no. 97058-914)
- TBE-urea gels, 15% (wt/vol) (Life Technologies, cat. no. EC68852BOX)
- TBE-urea gels, 10% (wt/vol) (Life Technologies, cat. no. EC68752BOX)
- TBE gels, 8% (wt/vol) (Life Technologies, cat. no. EC62152BOX)
- Mini-Cell polyacrylamide gel box, XCell SureLock (Life Technologies, cat. no. EI0001)
- Electrophoresis power supply (VWR, cat. no. 93000-744)
- Black gel box (LI-COR, cat. no. 929-97301)
- Shaker (BioExpress, cat. no. S-3200-LS)
- Vortexer (VWR, cat. no. 58816-121)
- Magnetic rack (Thermo Scientific, cat. no. 21359)
- Thermal cycler (Fisher Scientific, cat. no. E950030020)
- Thermomixer (Core Life Sciences, cat. no. H5000-HC)
- NanoDrop 2000 UV-visible spectrophotometer (Thermo Scientific, cat. no. ND-2000)
- 2100 Bioanalyzer (Agilent Technologies, cat. no. G2940CA)
- Qubit 2.0 fluorometer (Life Technologies, cat. no. Q32866)
- MiSeq, HiSeq or NextSeq next-generation sequencing platform (Illumina)

REAGENT SETUP

Protease inhibitor mix (50×) Dissolve one tablet of protease inhibitor in 1 ml of precooled RNase-free H_2O. Prepare it before use and store aliquots for up to 1 year at −20 °C. **! CAUTION** The protease inhibitor mix is an irritant. Handle solutions containing the protease inhibitor mix with care and dispose of waste according to institutional regulations.

α-Amanitin solution (1 mM) Dissolve 1 mg of α-amanitin in 1 ml of RNase-free H_2O. Prepare the solution before use and store aliquots for up to 1 year at −20 °C. **! CAUTION** α-Amanitin is toxic. Handle α-amanitin solution with care and dispose of waste according to institutional regulations.

Cytoplasmic lysis buffer Cytoplasmic lysis buffer is 0.15% (vol/vol) NP-40, 10 mM Tris-HCl (pH 7.0), 150 mM NaCl, 25 µM α-amanitin, 10 U SUPERase. In and 1× protease inhibitor mix. For one reaction mix, 3.8 µl of 10% (vol/vol) NP-40, 2.5 µl of 1 M Tris-HCl (pH 7.0), 7.5 µl of 5 M NaCl, 5 µl of 1× protease inhibitor mix (50×), 6.2 µl of 1 mM α-amanitin, 0.6 µl of SUPERase. In (20 U/µl) and 224.4 µl of RNase-free H_2O. Prepare this solution freshly before use with RNase-free reagents, and keep it on ice. **! CAUTION** α-Amanitin is toxic. Handle solutions containing α-amanitin with care, and dispose of waste according to institutional regulations. NP-40 and the protease inhibitor mix are irritants.

Sucrose buffer Sucrose buffer is 10 mM Tris-HCl (pH 7.0), 150 mM NaCl, 25% (wt/vol) sucrose, 25 µM α-amanitin, 20 U SUPERase. In and 1× protease inhibitor mix. For one reaction, mix 5 µl of 1 M Tris-HCl (pH 7.0), 15 µl of 5 M NaCl and 250 µl of 50% (wt/vol) filter-sterilized sucrose, 10 µl of 1× protease inhibitor mix (50×), 12.5 µl of 1 mM α-amanitin, 1.2 µl of SUPERase. In (20 U/µl) and 206.3 µl of RNase-free H_2O. Prepare this solution freshly before use with RNase-free reagents and keep it on ice. **! CAUTION** α-Amanitin is toxic. Handle solutions containing α-amanitin with care and dispose of waste according to institutional regulations. The protease inhibitor mix is an irritant.

Nuclei wash buffer Nuclei wash buffer is 0.1% (vol/vol) Triton X-100, 1 mM EDTA, in 1× PBS, 25 µM α-amanitin, 40 U SUPERase. In and 1× protease inhibitor mix. For one reaction, mix 2 µl of 0.5 M EDTA solution, 10 µl of 10% (vol/vol) Triton X-100, 20 µl of 1× protease inhibitor mix (50×), 25 µl of 1 mM α-amanitin, 2.5 µl of SUPERase. In (20 U/µl) and 940.5 µl of 1× PBS. Freshly prepare this solution before use with RNase-free reagents, and keep it on ice. **! CAUTION** α-Amanitin is toxic. Handle solutions containing α-amanitin with care and dispose of waste according to institutional regulations. Triton X-100 is harmful and is an irritant. Triton X-100 is hazardous to the environment. Handle solutions containing Triton X-100 with care and dispose of waste according to institutional regulations. The protease inhibitor mix is an irritant.

Glycerol buffer Glycerol buffer is 20 mM Tris-HCl (pH 8.0), 75 mM NaCl, 0.5 mM EDTA, 50% (vol/vol) glycerol, 0.85 mM DTT, 25 µM α-amanitin, 10 U SUPERase. In and 1× protease inhibitor mix. For one reaction, mix 5 µl of 1 M Tris-HCl (pH 8.0), 3.8 µl of 5 M NaCl, 0.5 µl of 0.25 M EDTA, 125 µl of 100% (vol/vol) filter-sterilized glycerol, 2.1 µl of 0.1 M filter-sterilized DTT, 5 µl of 1× protease inhibitor mix (50×), 6.2 µl of 1 mM α-amanitin, 0.6 µl of SUPERase. In (20 U/µl) and 101.8 µl of RNase-free H_2O. Freshly prepare this solution before use with RNase-free reagents and keep it on ice. **! CAUTION** α-Amanitin is toxic. Handle solutions containing α-amanitin with care and dispose of waste according to institutional regulations. DTT is toxic and corrosive. DTT, EDTA and the protease inhibitor mix are irritants.

Nuclei lysis buffer Nuclei lysis buffer is 1% (vol/vol) NP-40, 20 mM HEPES (pH 7.5), 300 mM NaCl, 1 M urea, 0.2 mM EDTA, 1 mM DTT, 25 µM α-amanitin, 10 U SUPERase. In and 1× protease inhibitor mix. For one reaction, mix 25 µl of 10% (vol/vol) NP-40, 5 µl of 1 M HEPES (pH 7.5), 0.5 µl of 0.1 M EDTA, 15 µl of 5 M NaCl, 25 µl of 10 M filter-sterilized urea, 2.5 µl of 0.1 M filter-sterilized DTT, 5 µl of 1× protease inhibitor mix (50×), 6.2 µl of 1 mM α-amanitin, 0.6 µl of SUPERase. In (20 U/µl) and 165.2 µl of RNase-free H_2O. Freshly prepare this solution before use with RNase-free reagents and keep it on ice. **! CAUTION** α-Amanitin is toxic. Handle solutions containing α-amanitin with care and dispose of waste according to institutional regulations. DTT is toxic and corrosive. DTT, EDTA, NP-40 and the protease inhibitor mix are irritants.

Chromatin resuspension solution Chromatin resuspension solution is 25 µM α-amanitin, 20 units SUPERase. In and 1× protease inhibitor

mix Roche in 1× PBS. Mix 5 µl of 1 mM α-amanitin, 4 µl of 1× protease inhibitor mix (50×), 0.5 µl of SUPERase.In (20 U/µl) and 190.5 µl of 1× PBS. This volume is sufficient for three reactions. Freshly prepare this solution before use with RNase-free reagents, and store it on ice. ! CAUTION α-Amanitin is toxic. Handle solutions containing α-amanitin with care, and dispose of waste according to institutional regulations. The protease inhibitor mix is an irritant.

Alkaline fragmentation solution (2×) For a 5-ml stock solution, mix 0.6 ml of 0.1 M Na_2CO_3 and 4.4 ml of 0.1 M $NaHCO_3$. Store 500-µl aliquots in air-tight screw-cap tubes at room temperature (22 °C) for up to 4 months. ! CAUTION Sodium carbonate is an irritant.

RNA precipitation solution For one reaction (562 µl), mix 60 µl of 3 M sodium acetate (pH 5.5), 2 µl of GlycoBlue (15 mg/ml) and 500 µl of RNase-free H_2O. Freshly prepare the solution before use and store it on ice.

SUPERase. In/DTT mix For one reaction (1.3 µl), mix 0.5 µl of SUPERase. In (20 U/µl) and 0.8 µl of 0.1 M DTT. Freshly prepare the solution before use and store it on ice. ! CAUTION DTT is toxic and corrosive. DTT is also an irritant.

DNA soaking buffer For one reaction (668 µl), mix 6.7 µl of 1 M Tris-HCl (pH 8.0), 40 µl of 5 M NaCl, 1.3 µl of 0.5 M EDTA and 620 µl of RNase-free H_2O. Freshly prepare the buffer before use and keep it at room temperature. ! CAUTION EDTA is an irritant.

HCl (1 N) Dilute hydrochloric acid concentrate by adding DNase-free H_2O to a final volume of 1 liter. Store 1 N HCl for up to 1 year at room temperature. ! CAUTION HCl-containing solutions are corrosive and cause irritation. Handle solutions containing HCl with care and dispose of waste according to institutional regulations.

SDS buffer (2×) SDS buffer is 100 mM Tris-HCl (pH 7.0), 4% (wt/vol) SDS and 20% (vol/vol) glycerol. Mix 5 ml of 1 M Tris-HCl (pH 7.0), 10 ml of 20% (wt/vol) SDS, 10 ml of 100% (vol/vol) filter-sterilized glycerol and 25 ml of filter-sterilized and deionized H_2O. The volume will be enough to process at least 50 subcellular fractions. Prepare SDS before use and store it at room temperature for up to several months. ! CAUTION SDS is corrosive and flammable. SDS is an irritant. Handle solutions containing SDS with care and dispose of waste according to institutional regulations.

Depletion oligo mix Resuspend each of the 20 biotinylated DNA oligos (**Table 2**) with 10 mM Tris-HCl (pH 8.0) so that the final concentration of each depletion oligo solution is 200 µM. Next, combine 5 µl of each depletion oligo solution and mix. The final concentration of each DNA oligo in the depletion oligo mix will be 10 µM. The volume of the depletion oligo mix will be 100 µl. This volume is enough for 50 depletion experiments when two depletion reactions are performed per sample. Prepare the mix before use and store it indefinitely at −20 °C.

Bind/wash buffer (2×) Bind/wash buffer is 5 mM Tris-HCl (pH 7.0), 2 M NaCl, 1 mM EDTA and 0.2% (vol/vol) Triton X-100. Mix 5 µl of 1 M Tris-HCl (pH 7.0), 400 µl of 5 M NaCl, 2 µl of 0.5 M EDTA, 20 µl of 10% Triton X-100 and 573 µl of RNase-free H_2O. The volume will be enough to perform seven specific depletion experiments using two depletion reactions per sample. Store it at room temperature for several months. ! CAUTION EDTA and Triton X-100 are irritants. Triton X-100 is harmful and hazardous to the environment. Handle solutions containing Triton X-100 with care and dispose of waste according to institutional regulations.

DNA loading buffer (6×) Dissolve 6 g of sucrose and 30 mg of Orange G in 20 ml of RNase-free H_2O. Bring the buffer to a final volume of 25 ml with RNase-free H_2O. Store it, protected from light, at room temperature for up to 1 year.

Gel staining solution To stain one TBE or TBE-urea gel, add 5 µl of SYBR Gold nucleic acid gel stain to 50 ml of 1× TBE buffer and mix. Prepare this solution immediately before use, and store it, protected from light, at room temperature until required. ! CAUTION SYBR Gold nucleic acid gel stain is flammable. Nucleic acid stains are usually mutagenic. Use personal protective equipment when handling nucleic acid gel stains and dispose of waste according to institutional regulations.

PROCEDURE

Quantitative purification of RNA polymerase by cell fractionation ● TIMING 45 min

▲ CRITICAL Cell fractionation has been optimized for HeLa S3 and HEK293T cells. If you are using other cell types, protocol optimizations might be required.

▲ CRITICAL Cell fractionation is performed on ice or at 4 °C, with buffers freshly prepared on the same day. All buffers are precooled on ice before use. Use RNase-free reagents and equipment.

1| Grow HeLa S3 or HEK293T cells in a 15 × 2.5 cm dish in DMEM containing 10% (vol/vol) FBS, 100 U/ml penicillin and 100 µg/ml streptomycin until they are 90% confluent.

▲ CRITICAL STEP Grow human cells according to the Cell Culture Guidelines of the Encyclopedia of DNA Elements (ENCODE) project (https://www.encodeproject.org/documents/60b6b535-870f-436b-8943-a7e5787358eb/). This includes keeping track of the growth time, passage number and cell density.

2| Wash the cells twice with 10 ml of PBS buffer.

▲ CRITICAL STEP Unless otherwise stated, perform all washing steps on ice. A typical wash takes <1 min.

3| Scrape cells into 1 ml of PBS buffer.

4| Determine the cell number by counting using a cell counter (see Equipment for details). Use 1×10^7 cells as input for each cell fractionation. In addition to the experimental sample, also process in parallel a control for western blot analysis (described in **Box 1**; see also **Figs. 1**,**2** and **Supplementary Fig. 2**).

5| Wash each sample of 1×10^7 cells with 10 ml of PBS.

6| Collect cells by centrifugation at 500*g* for 2 min at 4 °C.

7| Remove the supernatant by aspiration.

▲ CRITICAL STEP It is important to remove the supernatant completely at this step. If the supernatant is not completely removed, the cytoplasmic lysis buffer (Step 8) will be diluted, which affects the cell lysis efficiency.

8| Add 200 µl of cytoplasmic lysis buffer and transfer it to an RNase-free 1.5-ml microcentrifuge tube. By using a cut 1,000-µl pipette tip, pipette the sample up and down ten times.

9| Incubate the cell lysate on ice for 5 min.

10| By using a cut 1,000-µl pipette tip, layer the cell lysate onto 500 µl of sucrose buffer.

11| Collect cell nuclei by centrifugation at 16,000*g* for 10 min at 4 °C.

12| Remove the supernatant.
▲ **CRITICAL STEP** The supernatant represents the cytoplasmic fraction. The cytoplasmic fraction is only retained for the western blot control sample (**Box 1**).

13| Wash nuclei with 800 µl of nuclei wash buffer.

14| Collect washed nuclei by centrifugation at 1,150*g* for 1 min at 4 °C.
▲ **CRITICAL STEP** To remove cytoplasmic mature RNAs, it is important to remove the supernatant completely.

15| Add 200 µl of glycerol buffer. By using a cut 1,000-µl tip, resuspend the washed nuclei by pipetting up and down. Transfer the suspension to a new 1.5-ml RNase-free microcentrifuge tube.

16| Add 200 µl of nuclei lysis buffer, mix by pulsed vortexing and incubate the mixture on ice for 2 min.

17| Centrifuge the mixture at 18,500*g* for 2 min at 4 °C.

18| Remove the supernatant completely.
▲ **CRITICAL STEP** It is important to completely remove the supernatant containing nucleoplasmic RNAs. The nucleoplasmic fraction is retained only for the western blot control sample (**Box 1**).

19| Resuspend the chromatin in 50 µl of chromatin resuspension solution.
▲ **CRITICAL STEP** It is important to resuspend the sticky chromatin pellet before the preparation of the nascent RNA. This step increases the RNA yield. This step is performed for the experimental sample, as well as for the control sample. The experimental sample is used for the preparation of the nascent RNA (Step 20). The control sample is used for the western blot analysis, as described in **Box 1**.

Preparation of nascent RNA ● TIMING 1 h

20| Add 700 µl of QIAzol lysis reagent (part of miRNeasy mini kit, Qiagen) to the resuspended chromatin from Step 19.

21| Mix thoroughly by slowly pipetting up and down using a 1-ml syringe with a 22G needle. Alternatively, the chromatin pellet can also be solubilized by gentle vortexing.
! **CAUTION** Be aware that the QIAzol lysis reagent contains phenol. Handle solutions with care and according to institutional regulations.
▲ **CRITICAL STEP** Mix very carefully until the solution is homogeneous. Mix it slowly to avoid spilling the sample.

22| Prepare RNA using the miRNeasy mini kit according to the manufacturer's instructions, including the optional on-column DNase treatment using the RNase-free DNase set (Qiagen).

23| Determine the quantity and the quality of the prepared RNA using a NanoDrop 2000. The RNA yield is usually in the range of 20–30 µg. The expected absorbance (A_{260}/A_{280}) ratio is 2.1. Please also see the ANTICIPATED RESULTS for more information.
■ **PAUSE POINT** The isolated RNA can be stored for months at –80 °C.

Barcode DNA linker ligation ● TIMING 3.5 h

24| Denature the RNA sample from Step 23 and 5 µl of the ligation control oligonucleotide oGAB11 (10 µM; **Table 1**) for 2 min at 80 °C in a Thermomixer.

25| Prepare DNA linker ligation mix for each RNA sample and for oGAB11 in 0.2-ml RNase-free PCR tubes, as described in the table below. Prepare three ligation reactions per RNA sample and for the oGAB11 control; the RNA input per ligation reaction is 1 µg.
▲ **CRITICAL STEP** Before adding the truncated T4 RNA ligase 2, it is important to mix the ligation reaction until it is homogeneous. Next, add the ligase and mix again. Incompletely mixed samples will negatively affect the ligation efficiency.

Component	Amount per reaction (µl)		Final
	RNA sample	oGAB11 control	
PEG8000 (50% (vol/vol))	8.0	8.0	20% (vol/vol)
DMSO	2.0	2.0	10% (vol/vol)
T4 RNA ligase buffer (10×)	2.0	2.0	1×
Barcode DNA linker (1 µg)	1.0	1.0	
RNA sample (1 µg)	6.0	–	
oGAB11 (10 µM)	–	1.0	0.5 µM
RNase-free H_2O	–	5.0	
Truncated T4 RNA ligase 2	1.0	1.0	200 U

26| Incubate the ligation mixes for 3 h at 37 °C in a thermal cycler.

27| Add 0.7 µl of EDTA (0.5 M) to each ligation mix to stop the ligation reaction.

RNA fragmentation ● TIMING 45 min (plus overnight and ~3.5 h the next day)
28| Add 20 µl of 2× alkaline fragmentation solution to each ligation reaction and mix. Please note that Steps 28 and 29 are not performed for the oGAB11 ligation control.

29| Fragment the RNA at 95 °C in a thermal cycler for the appropriate time.
▲ **CRITICAL STEP** The fragmentation time needs to be adjusted whenever a new batch of alkaline fragmentation solution is applied. An overfragmentation or underfragmentation of the nascent RNA pool can lead to systematic biases. In a typical experiment RNA is fragmented between 10 and 40 min at 95 °C (**Fig. 3a**). The optimal fragmentation time is when most RNA molecules are in the required size range, usually between 35 and 100 nt (**Fig. 3a**, lane 6).

30| Add fragmented RNA sample (from Step 29) to 562 µl of RNA precipitation solution. Add the nonfragmented oGAB11 ligation control (from Step 27) to 562 µl of RNA precipitation solution. Mix well.

31| Precipitate the fragmented RNA and the nonfragmented oGAB11 ligation control by adding 750 µl of isopropanol. Mix well. Precipitate the mixture for ≥1 h at –80 °C.
■ **PAUSE POINT** The precipitation can be left at –80 °C overnight.

32| Centrifuge at 20,000*g* for 30 min at 4 °C to pellet the RNA.

33| Remove the supernatant and wash the pellet with 750 µl of 80% (vol/vol) ice-cold ethanol.

34| Spin the sample at 20,000*g* for 2 min at 4 °C, and then remove the supernatant.

35| Air-dry the RNA pellet for 10 min at room temperature.

36| Resuspend the RNA pellet in 10 µl of RNase-free H_2O. Add 10 µl of 2× TBU denaturing sample buffer to each RNA sample and mix it.

37| Prepare the RNA control ladder and the oGAB11 control. Add 1.0 µl of RNA control ladder or oGAB11 to 9 µl of RNase-free H_2O. Add 10 µl of 2× TBU denaturing sample buffer to each control sample and mix it.

38| Denature the RNA sample and RNA control samples (oGAB11 ligation control, ladder and oGAB11) for 2 min at 80 °C. Cool the samples on ice for 3 min.

39| Prerun a 15% (wt/vol) polyacrylamide TBE-urea gel at 200 V for 15 min in 1× TBE.

40| Separate the fragmented RNA samples and the RNA control samples (including the oGAB11 ligation control) by PAGE at 200 V for 65 min.

41| Stain the gel in 50 ml of gel staining solution for 5 min at room temperature on a shaker. Protect the gel from light during staining by the use of a black gel box.

42| Visualize the fragmented RNA and the oGAB11 ligation control under blue or UV light (**Fig. 3a** and **Supplementary Fig. 1**). For the fragmented RNA, excise the region between 35 and 100 nt (**Fig. 3a**). For the oGAB11 ligation control, excise the narrow band at ~55 nt (**Supplementary Fig. 1**, lane 4).

43| Extract the RNA from the gel slice by rapid gel extraction, as described in **Box 2**.

44| Precipitate RNA by adding 2 µl of 15 mg/ml GlycoBlue and 50 µl of 3 M sodium acetate (pH 5.5). Mix well. Add 750 µl of isopropanol and mix well. Incubate the precipitations at –80 °C for ≥1 h.
■ **PAUSE POINT** The RNA precipitations can be stored at –80 °C overnight.

45| Recover the RNA as described in Steps 32–35.

46| Resuspend the size-selected RNA and the oGAB11 ligation control in 10 µl of precooled RNase-free H_2O.

RT ● TIMING 4 h (plus overnight and ~1 h the next day)
47| Prepare the RT reaction mix tabulated below in a 0.2-ml RNase-free PCR tube and store it on ice.

Component	Amount per reaction (µl)	Final
First-strand buffer	3.3	1×
dNTPs, 5× (10 mM)	0.8	0.5 mM
Reverse primer oLSC007 (10 µM)	0.5	0.3 µM

48| Add 4.6 µl of RT reaction mix to 10 µl of RNA sample and to the oGAB11 control RNA sample.

49| Incubate the sample mixture for 2 min at 80 °C in a thermal cycler, and then cool it on ice for 3 min.

50| Add 1.3 µl of SUPERase. In/DTT mix and mix well.

51| Add 0.8 µl of SuperScript III (200 U/µl) and mix.

52| Incubate the mixture for 30 min at 48 °C in a thermal cycler.

53| Add 1.8 µl of 1 N NaOH, mix well and incubate the reaction for 20 min at 98 °C.

54| Neutralize the reaction by adding 1.8 µl of 1 N HCl; mix well and put the reaction on ice.

55| Add 20 µl of 2× TBU denaturing sample buffer to each cDNA sample and the oGAB11 cDNA control, and then mix.

56| Prepare the DNA control ladder. Add 1.0 µl of DNA control ladder to 9 µl of RNase-free H_2O. Add 10 µl of 2× TBU denaturing sample buffer to each cDNA sample and mix.

57| Denature the cDNA sample, oGAB11 cDNA control and DNA control ladder for 3 min at 95 °C in a Thermomixer. Cool the samples on ice for 3 min.

Box 2 | Rapid gel extraction ● TIMING 20 min

This protocol for rapidly extracting RNA or cDNA from a polyacrylamide TBE-urea gel is similar to the gel extraction protocol described by the Weissman laboratory[24,38].

1. Pierce the bottom of a 0.5-ml RNase-free microcentrifuge tube with a 21G needle.
2. Put the pierced 0.5-ml tube in a 1.5-ml RNase-free microcentrifuge tube.
3. Place the gel slice (from PROCEDURE Step 43 or 62) into the inner pierced 0.5-ml tube.
4. Centrifuge the mixture at 20,000*g* for 4 min at room temperature.
5. Add 200 µl of RNase-free H_2O and mix.
6. Incubate the sample for 10 min at 70 °C in a Thermomixer.
7. Vortex the mixture for 30 s at a medium intensity setting.
8. Cut the tip off of a 1,000-µl pipette tip and transfer the gel slurry into a microcentrifuge tube filter.
9. Centrifuge the mixture at 20,000*g* for 3 min at room temperature.
10. Combine eluates from three original ligation reactions (PROCEDURE Step 43) or from two RT reactions (PROCEDURE Step 62). The expected volume is ~600 or 400 µl, respectively.

58| Prerun a 10% (wt/vol) polyacrylamide TBE-urea gel at 200 V for 15 min in 1× TBE.

59| Separate the cDNA sample, the oGAB11 cDNA control and the DNA ladder by PAGE at 200 V for 65 min.

60| Stain the gel in 50 ml of gel staining solution for 5 min at room temperature on a shaker. Protect the gel from light during staining by the use of a black gel box.

61| Visualize the gel under blue/UV light and excise the cDNA between 85 and 160 nt (**Fig. 3b**).
? TROUBLESHOOTING

62| Extract the cDNA from the gel slice by rapid gel extraction, as described in **Box 2**. Combine eluates from two lanes (expected volume ~400 µl), add 25 µl of 5 M NaCl and mix.

63| Precipitate the cDNA by adding 2 µl of 15 mg/ml GlycoBlue and 750 µl of isopropanol. Mix well. Incubate the precipitations at –20 °C for ≥1 h.
■ **PAUSE POINT** The cDNA precipitation can be stored at –20 °C overnight.

64| Recover cDNA as described in Steps 32–35.

65| Resuspend cDNA and oGAB11 control cDNA in 15 µl of precooled RNase-free H_2O each.
■ **PAUSE POINT** The cDNA can be stored indefinitely at –20 °C.

Circularization of cDNA ● TIMING 1.5 h

66| Prepare circularization mix as tabulated below and store it on ice.

Component	Amount per reaction (µl)	Final
CircLigase 10× reaction buffer	2.0	1×
ATP (1 mM)	1.0	50 µM
$MnCl_2$ (50 mM)	1.0	2.5 mM

67| Add 4 µl of circularization mix to 15 µl of cDNA sample and the oGAB11 cDNA control sample in a 0.2-ml RNase-free PCR tube, and then mix well.
? TROUBLESHOOTING

68| Add 1 µl of CircLigase (100 U/µl) and mix.

69| Incubate the CircLigase reaction for 60 min at 60 °C and for 10 min at 80 °C in a thermal cycler.
■ **PAUSE POINT** Circularized cDNA can be stored indefinitely at –20 °C.

Specific depletion of highly abundant mature RNAs ● TIMING **1.5 h**

70| Prepare two specific depletion reactions per sample, but not for the oGAB11 control sample. Prepare depletion reactions in 0.2-ml DNase-free PCR tubes. The depletion DNA oligo pool that has been successfully used for HeLa S3 and HEK293T cells is given in **Table 2**. Please note that the oGAB11 control sample is not subjected to Steps 70–79. Store the oGAB11 control sample on ice until it is further processed at Step 80.

Component	Amount per reaction (μl)	Final
Circularization reaction (from Step 69)	5.0	
Depletion DNA oligo pool	1.0	1 μM (per depletion oligo)
SSC, 20×	1.0	2×
DNase-free H_2O	3.0	

71| Perform subtractive hybridization in a thermal cycler as given below.

	Temperature	Time
Denature	99 °C	90 s
Annealing	99–37 °C in 0.1 °C steps	1 s (per 0.1 °C step)
Final annealing	37 °C	15 min

72| Prepare Dynabeads MyOne streptavidin C1 (10 mg/ml) for specific depletion, as described in **Box 3**.

73| Transfer 10 μl of depletion reaction directly from the 0.2-ml PCR tube in the thermal cycler (from Step 71) to the washed and equilibrated beads in the Thermomixer. Immediately mix by pipetting.

74| Incubate the mixture in the Thermomixer for 15 min at 37 °C with mixing at 1,000 r.p.m.

75| Transfer the tubes from the Thermomixer into a magnetic rack and leave them for 1 min. Transfer the supernatant into a new 1.5-ml microcentrifuge tube.
▲ CRITICAL STEP The supernatant needs to be transferred carefully. Any remaining magnetic beads in the supernatant will have a negative impact on subsequent steps.

76| Add 2 μl of GlycoBlue, 6 μl of 5 M NaCl and 74 μl of DNase-free H_2O to each depletion reaction and mix. Add 150 μl of isopropanol and mix.

77| Incubate the circularized and depleted DNA for ≥1 h at –20 °C.
■ PAUSE POINT Precipitation can be left at –20 °C overnight.

Box 3 | Preparation of streptavidin-coupled magnetic beads for specific depletion ● TIMING 10 min

The following protocol describes how to prepare the Dynabeads MyOne streptavidin C1 (10 mg/ml) for specific depletion of the 20 most abundant chromatin-associated mature RNAs that are captured in NET-seq libraries obtained from HeLa S3 and HEK293T cells. Unless otherwise stated, perform all steps at room temperature.

1. Resuspend the beads by gentle vortexing.
2. Transfer 37.5 μl of beads per depletion reaction to a DNase-free 1.5-ml microcentrifuge tube.
3. Place the tube on a magnetic rack for 1 min and withdraw all of the supernatant from the tube.
4. Tube from the magnetic rack and resuspend beads in 37.5 μl of 1× bind/wash buffer.
5. Repeat this washing procedure (steps 3 and 4) two more times.
6. Place the tube on a magnetic rack for 1 min and withdraw the supernatant from the tube.
7. Remove the tube from the magnetic rack and resuspend the beads in 15 μl of 2× bind/wash buffer.
8. Transfer 10 μl of the resuspended and washed beads to a new tube.
9. Place the tube in a Thermomixer at 37 °C to equilibrate for 15–30 min.

78| Recover DNA as described in Steps 32–35.

79| Resuspend two DNA pellets per sample in 10 µl of ice-cold H_2O.
■ **PAUSE POINT** DNA can be stored indefinitely at –20 °C.

PCR amplification of the cDNA sequencing library ● **TIMING 4 h (plus overnight and ~1 h the next day)**
80| Prepare PCR mix for four pilot PCR amplification reactions for both the cDNA sample and the oGAB11 control cDNA sample. Mix well and store it on ice.

Component	Amount for 4 reactions (µl)	Final
Phusion HF buffer, 5×	15.2	1×
dNTPs (10 mM)	1.5	0.2 mM
Forward primer (Illumina index primer, 100 µM)	0.4	0.5 µM
oNTI231 (reverse primer, 100 µM)	0.4	0.5 µM
DNase-free H_2O	57.6	
Phusion DNA polymerase (2 U/µl)	0.9	1.8 U

81| For each PCR, put 19 µl of PCR master mix in a 0.2-ml RNase-free PCR tube.

82| Add 1 µl of circularized cDNA and mix it well.

83| Perform PCR pilot amplifications as tabulated below. Remove one PCR tube for each sample at the end of the extension step after 6, 8, 10 and 12 amplification cycles.

Cycle number	Denature	Anneal	Extend
1	98 °C, 30 s		
2–14	98 °C, 10 s	60 °C, 10 s	72 °C, 5 s

84| Add 3.4 µl of 6× DNA loading dye to each tube and mix well.

85| Prepare DNA control ladder. Add 1.0 µl of DNA control ladder to 9 µl of DNase-free H_2O. Add 2 µl of 6× DNA loading dye and mix well.

86| Separate the PCR products and the DNA control ladder by TBE gel electrophoresis on an 8% (wt/vol) TBE gel at 180 V for 55 min.

87| Stain the gel in 50 ml of gel staining solution for 5 min at room temperature on a shaker. Protect the gel from light during staining by the use of a black gel box.

88| Visualize the gel under blue/UV light and identify the optimal PCR amplification cycle for each cDNA sequencing library (**Fig. 3c**). The NET-seq library runs at ~150 nt (**Fig. 3c**). The optimal PCR amplification cycle is characterized by a clear band at ~150 nt and the absence of PCR products at the higher-molecular-weight range (**Fig. 3c**, lane 3).
? TROUBLESHOOTING

89| Perform four PCR amplification reactions per sample with the optimal amplification cycle, as described in Steps 80–88. The oGAB11 control sample is not subjected to Steps 89–105.
? TROUBLESHOOTING

90| For each sample, excise the band that contains the PCR product from the gel (**Fig. 3d**).
▲ **CRITICAL STEP** Excise the broad band at ~150 nt. Avoid contamination from the lower band that runs at ~120 nt, representing PCR product from empty circles (**Fig. 3d**, brown asterisk). Empty circles are circularized cDNA molecules that arise from unextended RT primers, and hence they do not contain any information about the original nascent RNA.

91| Pierce the bottom of a 0.5-ml DNase-free microcentrifuge tube with a 21G needle.

92| Put the pierced 0.5-ml tube in a 1.5-ml DNase-free microcentrifuge tube.

93| Place the gel slices into the inner pierced 0.5-ml tube.

94| Centrifuge the tube at 20,000*g* for 4 min at room temperature.

95| Add 670 µl of DNA soaking buffer and mix.

96| Incubate the tube at room temperature at 1,500 r.p.m. overnight in a Thermomixer.

97| By using a cut 1,000-µl pipette tip, transfer the gel slurry into a microcentrifuge tube filter.

98| Centrifuge at 20,000*g* for 3 min at room temperature. Transfer the eluate to a new 1.5-ml microcentrifuge tube.

99| Precipitate the NET-seq library by adding 2 µl of 15 mg/ml GlycoBlue and 680 µl of isopropanol. Mix well. Incubate the precipitations at –20 °C for ≥1 h.
■ **PAUSE POINT** The DNA precipitations can be stored at –20 °C overnight.

100| Recover DNA as described in Steps 32–35.

101| Resuspend the NET-seq library in 10 µl of Tris-HCl (10 mM, pH 8.0).
■ **PAUSE POINT** The DNA sequencing library can be stored indefinitely at –20 °C.

Quantification and characterization of the NET-seq library ● **TIMING 1 h**
102| Prepare a 1:5 dilution of the NET-seq library by adding 1 µl of the NET-seq library to 4 µl of Tris-HCl (10 mM, pH 8.0); mix well.

Box 4 | Guide to the bioinformatics analysis of human NET-seq data

The following protocol describes the general steps for the computational analysis of human NET-seq data. For more information, please see Mayer *et al.*[29]. The bioinformatics analysis of human NET-seq data also uses custom scripts that will be shared upon request. We are currently working on converting these scripts for use on a generic server, and we intend to make the converted scripts available on our GitHub site in due course (https://github.com/churchmanlab).
1. Use sequencing reads that pass the default Illumina quality filter (BCL2FASTQ2).
2. Remove the six 5′-end nucleotides that correspond to the random molecular barcode from sequencing reads.
▲ **CRITICAL STEP** The information of the barcode sequence should remain associated with the corresponding sequencing read. This is important for the identification of reads that arise from RT mispriming events (step 4).
3. Align both sets of sequencing reads (containing and not containing the barcode sequence) to the human reference genome using the STAR aligner (v.2.4.0)[65].
▲ **CRITICAL STEP** Perform the alignment without providing the transcriptome information to avoid any biases favoring annotated regions.
4. Identify and remove reads that arise from mispriming events during RT. Reads that originate from RT mispriming can be identified by aligning the reads still containing the barcode sequence and that map without mismatches in the six 5′-end nucleotides to the human reference genome. Therefore, these reads seem to have 'barcode sequences' that perfectly match the genomic sequence adjacent to the aligned read and are discarded.
5. Record the position of the 5′ end of the sequencing read that corresponds to the 3′ end of the original nascent RNA, with a custom script that applies the HTSeq package[66].
6. Identify and remove sequencing reads that are due to PCR duplication—i.e., reads that align to the same genomic position and contain the same barcode sequence. These reads are computationally removed using a custom script.
7. Identify and remove sequencing reads due to splicing intermediates—i.e., reads that align exactly to the 3′-end nucleotide of introns and the 3′-end nucleotide of exons. These reads are computationally removed using a custom script.
8. (Optional) Identify and remove sequencing reads that arise from chromatin-associated mature RNAs. Sequencing reads that arise from polyadenylated mature RNAs do not align to the human reference genome and are discarded. Sequencing reads that originate from nonpolyadenylated mature RNAs or chromatin-associated RNA that is cleaved but not yet polyadenylated align exactly to the pA site of the corresponding gene—i.e., the site where the transcript is cleaved and polyadenylated. Reads that align exactly to the pA site can be computationally removed using a custom script. Alternatively, you can exclude NET-seq signals right at pA-sites from downstream analyses.
9. Visualize human NET-seq data in a genome browser such as the Integrative Genomics Viewer (IGV)[67].

103| Use 1 µl of the diluted NET-seq library for quantification with the Qubit fluorometer using the Qubit dsDNA HS assay kit. Prepare the sample and perform the measurement according to the manufacturer's protocol.

104| Use 1 µl of the diluted NET-seq library for characterization on the Agilent Bioanalyzer; use the high-sensitivity DNA analysis kit according to the manufacturer's instructions.

105| Sequence the human NET-seq library from the 3′ end on the Illumina platform using oLSC006 (**Table 1**) as a custom sequencing primer. Perform sequencing according to the manufacturer's instructions. NET-seq libraries are typically sequenced on MiSeq, HiSeq or NextSeq next-generation sequencing platforms. A reasonable coverage is obtained with 100–200 million reads. A brief guide to the bioinformatics analysis of human NET-seq data is given in **Box 4**. A more detailed description of the computational analysis of human NET-seq data can also be found in the supplemental information of the original human NET-Seq publication[29].

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 3**.

TABLE 3 | Troubleshooting table.

Step	Problem	Possible reason	Solution
61	No cDNA smear is observed after the reverse transcription	The DNA linker ligation is inefficient	Mix the viscous ligation reaction properly (Step 25). Replace components of the ligation reaction every 4 months
		The reverse transcriptase enzyme has lost activity	Replace the SuperScript III enzyme and other RT reaction reagents every 4 months, and note the enzyme lot
67	A brown precipitate forms after adding the RT product to the circularization mix	The reason is unclear	Replace the isopropanol precipitation (Step 63) by an ethanol precipitation to reduce the amount of salt that co-precipitates with the cDNA. In addition, reduce the amount of $MnCl_2$ to 50%
88	A DNA smear that migrates slower in an 8% (wt/vol) TBE gel appears (**Fig. 3c**; 10 and 12 PCR cycles)	PCR overamplification	Reduce the number of PCR cycles
89	Strong PCR band at ~120 nt on an 8% (wt/vol) TBE gel (**Fig. 3c**,**d**)	PCR product that originates from empty circles. Empty circles arise from unextended RT primers that get circularized by the CircLigase enzyme	Precisely excise the RT product (Step 61) and use less RT primer Precisely excise the final PCR product at around 150 nt (**Fig. 3d**) Run the 8% (wt/vol) TBE gel a little bit longer than 55 min
Box 1 (step 4)	Western blot signals of the cytoplasmic marker protein GAPDH in the chromatin fraction	Reduced cell lysis efficiency because of an excess of input per cell fractionation reaction	Reduce the amount of input. The amount of cells per fractionation reaction should not exceed 2×10^7 HeLa S3 or HEK293T cells
		The supernatant after washing the cell nuclei (Step 18) is not completely removed	Completely remove the supernatant after washing the cell nuclei
	Strong western blot signals for elongating Pol II (CTD Ser5-P and Ser2-P) in the cytoplasmic or nucleoplasmic fractions	Treatment during cell lysis is too harsh, leading to dissociation of transcribing Pol II	Reduce the amount of cell lysis buffer that is added to the cell pellet. Note that too little lysis buffer can negatively affect the cell lysis efficiency

● TIMING

Steps 1–19, quantitative purification of RNA polymerase by cell fractionation: 45 min
Steps 20–23, preparation of nascent RNA: 1 h

Steps 24–27, barcode DNA linker ligation: 3.5 h
Steps 28–31, RNA fragmentation: 45 min, overnight RNA precipitation
Steps 32–46, complete RNA fragmentation: ~3.5 h
Steps 47–63, RT: 4 h, overnight cDNA precipitation
Steps 64 and 65, complete RT: ~1 h
Steps 66–69, circularization of cDNA: 1.5 h
Steps 70–79, specific depletion of highly abundant mature RNAs: 1.5 h
Steps 80–96, PCR amplification of the cDNA sequencing library: ~4 h
Steps 97–101, complete NET-seq library preparation: ~1 h
Steps 102–105, quantification and characterization of the NET-seq library: 1 h
Box 1, steps 1–4, checking cell fractionation by western blotting: 5 h, plus overnight and ~2.5 h the next day
Box 2, rapid gel extraction: 20 min
Box 3, preparation of streptavidin-coupled magnetic beads for specific depletion: 10 min

ANTICIPATED RESULTS

RNA extraction (PROCEDURE Steps 20–23) from the chromatin prepared from 1 × 10^7 HeLa S3 and HEK293T cells typically results in a total RNA yield of 20–30 µg with an A_{260}/A_{280} ratio of 2.1, as determined by NanoDrop spectrophotometer measurement. The efficiency of the DNA linker ligation is usually >95%, as monitored by PAGE of a control ligation using the RNA oligonucleotide oGAB11 (**Table 1**; **Supplementary Fig. 1**, PROCEDURE Steps 24–27 and 37–42). Most of the RNA after fragmentation by partial alkaline hydrolysis (PROCEDURE Steps 28–42) typically runs between 10 and 150 nt, as revealed by PAGE (**Fig. 3a**). The cDNA that is obtained upon RT (PROCEDURE Steps 47–61) runs between 80 and 160 nt, as observed by PAGE (**Fig. 3b**). Gel electrophoresis of the PCR product after the final amplification (PROCEDURE Steps 80–89) reveals a broad band at ~150 nt (**Fig. 3c**,**d**). The total yield of a typical NET-seq library from HeLa S3 and HEK293T cells is usually 20–30 ng, as determined by Qubit fluorometer (PROCEDURE Step 103) and Bioanalyzer (PROCEDURE Step 104) measurements. The average fragment length of the human NET-seq library is ~150 nt (**Fig. 4**).

High-throughput sequencing of a typical NET-seq library (PROCEDURE Step 105 and **Box 4**) obtained from HeLa S3 cells has the following characteristics: ~50% of the sequencing reads map uniquely to the annotated human reference genome. Of those reads, 76% align to Pol II–transcribed genes, 15% align to Pol I–transcribed genes (also included are reads that map to spacer regions of

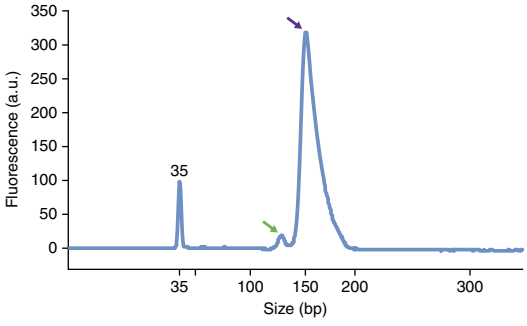

Figure 4 | Bioanalyzer electropherogram of a representative human NET-seq library. The NET-seq library was generated from 3 µg of RNA obtained from HeLa S3 cells. Approximately 1.1 ng of the final NET-seq library was loaded onto a high-sensitivity DNA microfluidic chip. The profile shows a prominent peak around 150 nt (violet arrow) and a smaller peak around 120 nt (green arrow). The less prominent peak corresponds to background that arises from unextended RT primer leading to a 119-nt PCR product (see also **Fig. 3c**,**d**). This background DNA also results in sequencing data. However, these sequencing reads will contain sequences of the 3′ and 5′ linker regions that are part of the RT primer and are computationally removed. The background DNA should be lower than 10%, as compared with the amount of the NET-seq library. The peak at 35 nt represents the manufacturer's internal standard. a.u., arbitrary units.

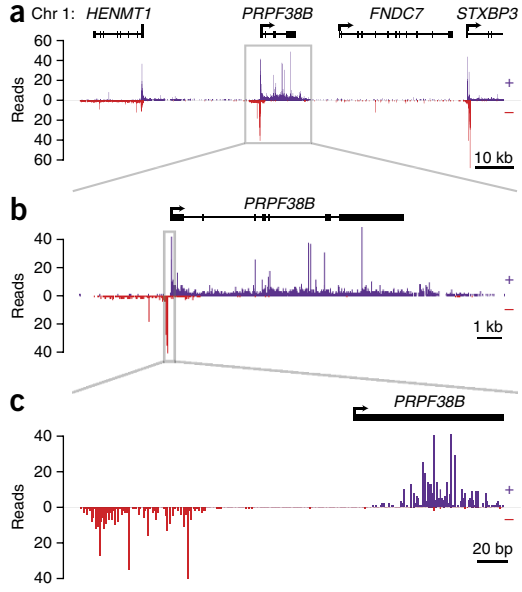

Figure 5 | DNA strand–specific Pol II occupancy at a representative region of the human genome, as determined by human NET-seq. (**a**–**c**) Pol II density for HeLa S3 cells is shown at different zoom levels for multiple genes (**a**), for a single gene (**b**) and for the promoter-proximal region of the gene shown in **b** (**c**). The Pol II density at the positive (+) and negative (–) DNA strand is shown in violet and red, respectively. The transcription start site and the direction of transcription is indicated by a black arrow. Exonic and intronic regions are shown as black boxes and lines, respectively. Adapted with permission from Mayer *et al.*[29]. *HENMT1*, HEN1 methyltransferase homolog 1; *PRPF38B*, pre-mRNA processing factor 38B; *FNDC7*, fibronectin type III domain containing 7; *STXBP3*, syntaxin binding protein 3.

rRNA genes, indicating nascent Pol I transcripts), 8% align to Pol III–transcribed genes and <1% align to mitochondrial genes. At protein-coding genes, the highest sequencing coverage is obtained at the promoter-proximal region (between the TSS and +1 kb): >50% of active genes have a coverage of >1 read per kb per million uniquely aligned reads[29]. Typical results of a human NET-seq experiment are shown in **Figure 5**.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS We thank J. di Iulio and S. Maleri for help developing this approach. We thank J. di Iulio, H. Landry and A. Snavely for critical comments on the manuscript. This work was supported by US National Institutes of Health National Human Genome Research Institute (NHGRI) grant R01HG007173 to L.S.C.; a Damon Runyon Dale F. Frey Award for Breakthrough Scientists (to L.S.C.); and a Burroughs Wellcome Fund Career Award at the Scientific Interface (to L.S.C.). A.M. was supported by Long-Term Postdoctoral Fellowships of the Human Frontier Science Program (HFSP) (LT000314/2013-L) and the European Molecular Biology Organization (EMBO) (ALTF858-2012).

AUTHOR CONTRIBUTIONS A.M. performed experiments and prepared figures. A.M. and L.S.C. wrote the manuscript.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

1. Margaritis, T. & Holstege, F.C.P. Poised RNA polymerase II gives pause for thought. *Cell* **133**, 581–584 (2008).
2. Jonkers, I. & Lis, J.T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16**, 167–177 (2015).
3. Adelman, K. & Lis, J.T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
4. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 599–577 (1997).
5. Rahl, P.B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
6. Jonkers, I., Kwak, H. & Lis, J.T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**, e02407 (2014).
7. Preker, P. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008).
8. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
9. Seila, A.C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
10. Sigova, A.A. *et al.* Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **110**, 2876–2881 (2013).
11. Seila, A.C., Core, L.J., Lis, J.T. & Sharp, P.A. Divergent transcription: a new feature of active promoters. *Cell Cycle* **8**, 2557–2564 (2009).
12. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **488**, 101–108 (2013).
13. Pelechano, V. & Steinmetz, L.M. Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14**, 880–893 (2013).
14. Jensen, T.H., Jacquier, A. & Libri, D. Dealing with pervasive transcription. *Mol. Cell* **52**, 473–484 (2013).
15. Buck, M.J. & Lieb, J.D. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349–360 (2004).
16. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. & Young, R.A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
17. Raha, D., Hong, M. & Snyder, M. ChIP-Seq: a method for global identification of regulatory elements in the genome. *Curr. Protoc. Mol. Biol.* **91**, 21.19.1–21.19.14 (2010).
18. Descostes, N. *et al.* Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife* **3**, e02105 (2014).
19. Rhee, H.S. & Pugh, B.F. ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr. Protoc. Mol. Biol.* **100**, 21.24.1–21.24.14 (2012).
20. He, Q., Johnston, J. & Zeitlinger, J. ChIP-nexus enables improved detection of *in vivo* transcription factor binding footprints. *Nat. Biotechnol.* **33**, 395–401 (2015).
21. Kwak, H., Fuda, N.J., Core, L.J. & Lis, J.T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).
22. Wang, I.X. *et al.* RNA-DNA differences are generated in human cells within seconds after RNA exits polymerase II. *Cell Rep.* **6**, 906–915 (2014).
23. Churchman, L.S. & Weissman, J.S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2012).
24. Churchman, L.S. & Weissman, J.S. Native elongating transcript sequencing (NET-seq). *Curr. Protoc. Mol. Biol.* **98**, 14.4.1–14.4.17 (2012).
25. Larson, M.H. *et al.* A pause sequence enriched at translation start sites drives transcription dynamics *in vivo*. *Science* **344**, 1042–1047 (2014).
26. Vvedenskaya, I.O. *et al.* Interactions between RNA polymerase and the 'core recognition element' counteract pausing. *Science* **344**, 1285–1289 (2014).
27. Weber, C.M., Ramachandran, S. & Henikoff, S. Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol. Cell* **53**, 819–830 (2014).
28. Ferrari, F. *et al.* 'Jump start and gain' model for dosage compensation in drosophila based on direct sequencing of nascent transcripts. *Cell Rep.* **5**, 629–636 (2013).
29. Mayer, A. *et al.* Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**, 541–554 (2015).
30. Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C. & Cramer, P. Architecture of the RNA polymerase–Spt4/5 complex and basis of universal transcription processivity. *EMBO J.* **30**, 1302–1310 (2011).
31. Kireeva, M.L., Komissarova, N., Waugh, D.S. & Kashlev, M. The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J. Biol. Chem.* **275**, 6530–6536 (2000).
32. Wuarin, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell. Biol.* **14**, 7219–7225 (1994).
33. Cai, H. & Luse, D.S. Transcription initiation by RNA polymerase II *in vitro*. *J. Biol. Chem.* **262**, 298–304 (1987).
34. Bhatt, D.M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290 (2012).
35. Khodor, Y.L. *et al.* Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev.* **25**, 2502–2512 (2011).
36. Pandya-Jones, A. & Black, D.L. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15**, 1896–1908 (2009).
37. Lindell, T.J., Weinberg, F., Morris, P.W., Roeder, R.G. & Rutter, W.J. Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science* **170**, 447–449 (1970).
38. Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M. & Weissman, J.S. The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.* **7**, 1534–1550 (2012).
39. Flynn, R.A., Almada, A.E., Zamudio, J.R. & Sharp, P.A. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl. Acad. Sci. USA* **108**, 10460–10465 (2011).
40. Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
41. De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).

42. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. & Sharp, P.A. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363 (2013).
43. Brannan, K. *et al.* mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. *Mol. Cell* **46**, 311–324 (2012).
44. Ntini, E. *et al.* Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20**, 923–928 (2013).
45. Ozsolak, F. *et al.* Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* **143**, 1018–1029 (2010).
46. Kim, J.B. & Sharp, P.A. Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *J. Biol. Chem.* **276**, 12317–12323 (2001).
47. Peterlin, B.M. & Price, D.H. Controlling the elongation phase of transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).
48. Scruggs, B.S. *et al.* Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol. Cell* **58**, 1101–1112 (2015).
49. Danko, C.G. *et al.* Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol. Cell* **50**, 212–222 (2013).
50. Laitem, C. *et al.* CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II–transcribed genes. *Nat. Struct. Mol. Biol.* **22**, 396–403 (2015).
51. Nojima, T. *et al.* Mammalian NET-Seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**, 526–540 (2015).
52. Mayer, A. *et al.* Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278 (2010).
53. Kim, H. *et al.* Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat. Struct. Mol. Biol.* **17**, 1279–1286 (2010).
54. Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* **39**, 1512–1516 (2007).
55. Muse, G.W. *et al.* RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39**, 1507–1511 (2007).
56. Steinmetz, E.J. *et al.* Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol. Cell* **24**, 735–746 (2006).
57. Bataille, A.R. *et al.* A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol. Cell* **45**, 158–170 (2012).
58. Tietjen, J.R. *et al.* Chemical-genomic dissection of the CTD code. *Nat. Struct. Mol. Biol.* **17**, 1154–1161 (2010).
59. Mahony, S. & Pugh, B.F. Protein–DNA binding in high-resolution. *Crit. Rev. Biochem. Mol. Biol.* **50**, 269–283 (2015).
60. Gilmour, D.S. & Fan, R. Detecting transcriptionally engaged RNA polymerase in eukaryotic cells with permanganate genomic footprinting. *Methods* **48**, 368–374 (2009).
61. Core, L.J. *et al.* Defining the status of RNA polymerase at promoters. *Cell Rep.* **2**, 1025–1035 (2012).
62. Brueckner, F. & Cramer, P. Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase II translocation. *Nat. Struct. Mol. Biol.* **15**, 811–818 (2008).
63. Chapman, R.D. *et al.* Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**, 1780–1782 (2007).
64. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
65. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2012).
66. Anders, S., Pyl, P.T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
67. Robinson, J.T. *et al.* Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).