Published in final edited form as:

Methods. 2019 April 15; 165–160: 165–176. doi:10.1016/j.ymeth.2019.02.003.

Use of the Nuclear Walk-on Methodology to Determine Sites of RNA Polymerase II Initiation and Pausing and Quantify Nascent RNAs in Cells

Christopher B. Ball^a, Kyle A. Nilson^b, and David H. Price^a

^aDepartment of Biochemistry, University of Iowa, Iowa City, IA, 52242

^bCenter for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Abstract

Transcription by RNA polymerase II (Pol II) is controlled during initiation, elongation, and termination by a large variety of transcription factors, the state of chromatin modifications, and environmental conditions. Herein we describe experimental approaches for the examination of Pol II transcription at semi-global and genome-wide scales through analysis of nascent Pol II transcripts. We begin with a description of the nuclear walk-on (NWO) assay, which involves rapid isolation of nuclei in the presence of EDTA, followed by extension of about a quarter of the nascent transcripts with ³²P-CTP. Labeled nascent transcripts are then analyzed by denaturing PAGE and phosphorimaging followed by densitometry analysis to quantify the signal on the gel. A parallel reaction containing α -amanitin to inhibit Pol II reveals transcription due to Pol I and Pol III, which can be subtracted to yield a profile of Pol II transcription. We then describe how to use the NWO as a front end for PRO-Seq and PRO-Cap methods, which permit the genome-wide characterization of Pol II transcription at nucleotide resolution and provide precise information about sites of transcription initiation and pausing. We discuss strategies for optimizing sequencing methods that capture nascent Pol II transcripts, methods of bias reduction, and approaches for normalizing these and other sequencing datasets using spike-in controls.

Graphical Abstract

Corresponding Author: David H. Price. 260 EMRB, Department of Biochemistry, University of Iowa, 500 Newton Road, Iowa City, IA 52242. 319-335-7910. david-price@uiowa.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Pol II; promoter-proximal pausing; nuclear walk-on; nascent transcript; PRO-Seq; spike-in control

1. Introduction

Pol II is the RNA polymerase responsible for the synthesis of messenger RNA, long non-coding RNA, micro RNA, and most small nuclear RNAs. Transcription by Pol II occurs in three phases. From start to finish, these phases are initiation, elongation, and termination. Initiation of eukaryotic gene transcription follows the assembly of a preinitiation complex that consists of general transcription factors and Pol II at a core promoter region. The assembly of this complex is influenced by the concerted functions of core promoter elements, enhancers, transcription factors, the Mediator complex, and chromatin. Immediately following initiation at genes in metazoans, the general transcription factors are exchanged for the pausing factors DSIF and NELF. These pausing factors decrease the rate of the early elongating polymerase and thereby induce promoter-proximal pausing. Promoter-proximal pausing is an important mechanism of transcriptional control that has been implicated in both timely response to environmental stimuli and maintenance of the promoter chromatin state [1–3]. Clearance from the paused state occurs by release of Pol II into productive elongation by P-TEFb or by termination from the pause via a mechanism that is yet unclear. To release Pol II into productive elongation, P-TEFb phosphorylates DSIF, which results in the dissociation of NELF. Phospho-DSIF becomes a productive elongation factor, retaining its association with Pol II. Other elongation factors, such as the PAF1 complex, are recruited to productively elongating Pol II upon loss of NELF [4]. Cellular machinery that mediates capping, splicing, and polyadenylation of the nascent RNA, as well as negotiation of nucleosomes, is also recruited to elongating Pol II. Once synthesis of the nascent messenger is complete, Pol II is terminated. Each of the above stages of transcription is regulated by a multitude of different factors and each is potentially rate-limiting [5–7]. Defects in many different transcription regulatory mechanisms are associated with disease states, including cancer [8–10]. Moreover, many different pathogenic viruses hijack the host transcriptional machinery during their life cycle [11, 12]. There is clear potential for therapeutic design in elucidating mechanisms of transcriptional control.

Characterization of the Pol II transcription cycle has relied on numerous different experimental approaches, including functional biochemical assays, structural studies, and global bioinformatics analyses. The advent of next-generation sequencing and chromatin immunoprecipitation and sequencing (ChIP-Seq) permitted the first genome-wide characterizations of Pol II occupancy, which is interpreted as a readout of Pol II transcription [13]. In Pol II ChIP-Seq data, promoter-proximal pausing is observed at the vast majority of metazoan genes, and the signal that falls over gene bodies, downstream of the pause, captures the level of productive elongation. Variations on ChIP-Seq, such as ChIP-Exo, provide more precise determinations of Pol II and transcription factor occupancy [14–16]. Permanganate footprinting offers an alternative to ChIP that maps the position of engaged Pol II through permanganate oxidation of thymine in single-stranded DNA in association with Pol II [17]. However, methods that profile the nascent transcriptome by sequencing of

nascent RNA offer the highest resolution views of Pol II transcription. These methods include nascent elongating transcript (NET)-Seq [18, 19], global run-on (GRO)-Seq [20], and precision nuclear run-on (PRO)-Seq [21, 22]. These methods directly measure Pol II transcription rather than just measuring mRNAs as is the case for standard RNA-Seq methods. Each of these methods varies in its approach to capturing the nascent transcriptome. While NET-Seq involves extraction of nascent RNA from immunoprecipitated Pol II elongation complexes, both GRO-Seq and PRO-Seq involve labeling of nascent RNA during a nuclear run-on reaction. During GRO-Seq, nascent transcripts are extended in the presence of Sarkosyl and a mixture of nucleoside triphosphates (NTPs) that contains BromoU (BrU). Transcripts that contain BrU can then be immunoprecipitated to prepare for sequencing. During PRO-Seq, nascent transcripts are extended in the presence of Sarkosyl and biotinylated NTPs. Although biotinylated NTPs are not chain-terminating, the efficiency of incorporating more than one is extremely low. This constitutes an advantage over GRO-Seq in that the 3' ends of the nascent transcripts are accurately determined, whereas in GRO-Seq the nascent transcripts are extended by 10–100 nucleotides. Biotinylated nascent transcripts can be isolated with streptavidin beads and prepared for sequencing. Each of these methods can be used in different capacities to study sequence determinants of Pol II transcription, such as sites of transcription initiation and pausing. Moreover, these methods can be employed to study how different transcription factors and environmental factors impact Pol II transcription at extremely high resolution.

Herein we discuss experimental approaches to the examination of Pol II transcription through the analysis of nascent Pol II transcripts. We start with a detailed description of the NWO assay, a method that permits rapid and quantitative analysis of Pol II pausing and productive elongation. We highlight important aspects of cell culture, and carefully detail the isolation of nuclei, which must be performed quickly to prevent engaged Pol II from elongating during the isolation process. The nuclei isolation protocol that was developed for the NWO is an applicable front end to GRO-Seq and PRO-Seq nuclear run-on reactions. We apply this front end to PRO-Seq and a variant protocol called PRO-Cap, and discuss aspects of PRO-Seq and PRO-Cap library preparation, as well as how the NWO and in vitro transcription can be utilized to monitor the efficiencies of library preparation steps and optimize experimental conditions. Emphasis is given to the utilization of unique molecular identifiers, which permit the computational removal of PCR duplicates that are formed during library preparation. In addition, approaches to library size-selection and bias reduction at various steps of library preparation are highlighted. The methods section ends with a discussion of the utility of spike-in controls, and how these controls can be used to normalize sequencing data.

2. Description of Methods

2.1. Materials

Chemicals and reagents—2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES, Millipore 391338)

potassium chloride (KCl, Sigma P9541)

sodium phosphate dibasic, anhydrous (Na_2HPO_4 , RPI, S23145)

potassium phosphate monobasic (KH_2PO_4 , Sigma, P5655)

1 mM ethylenediaminetetraacetic acid (EDTA, Sigma ED3SS)

IGEPAL CA-360 (Sigma I3021)

phenylmethane sulfonyl fluoride (PMSF, Sigma P7626)

spermine (Sigma S1141)

spermidine (Sigma S2501)

dithiothreitol (DTT, Calbiochem 3860-OP)

SUPERase-In (Thermo Scientific, AM2694)

sucrose (Sigma S0389)

Roche cOmplete EDTA-free protease inhibitor cocktail (Sigma 1187358001)

magnesium chloride (MgCl_2 , Sigma M9272)

glycerol (JT Baker 4043-01)

α -amanitin (Sigma A2263)

α - ^{32}P -CTP (3000 Ci/mmol, Perkin Elmer BLU008H001MC)

Sarkosyl (IBI Scientific IB07080)

Trizma base (Sigma T1503)

sodium chloride (NaCl , Sigma S3014)

torula yeast RNA (Sigma R6625)

TRIzol LS (Thermo Scientific 10296028)

urea (Sigma 51456)

Dynabeads™ M-280 streptavidin beads (11206D)

Stocks—500 mM HEPES, pH 7.8

4 M / 500 mM KCl

500 mM EDTA, pH 8.3

10% v/v IGEPAL CA-360

isopropanol-saturated PMSF

500 mM spermine

spermidine

1 M DTT

2 M sucrose

100 mM MgCl₂

50% glycerol v/v

1 mg/mL α -amanitin

10% Sarkosyl w/v

1 M Trizma base, pH 8.0

1 M NaCl

10 M urea 10X PBS

Solutions

Reagents to be added right before use:

DTT: 1 M stock

PMSF: Isopropanol-saturated solution (always 1:1000 dilution)

RNase inhibitor: SUPERase-In

Protease inhibitor: cOmplete EDTA-free protease inhibitor cocktail (1 tablet/ 50 mL)

10X PBS:

80 g NaCl

2 g KCl

11 g Na₂HPO₄ (anhydrous)

2 g KH₂PO₄

Dissolve above reagents in 1 L H₂O

Lysis Buffer:

20 mM HEPES, pH 7.6 (from pH 7.8 stock)

100 mM KCl

1 mM EDTA

1% v/v IGEPAL CA-360

1 mM spermine
1 mM spermidine
320 mM sucrose
Complete with 1 mM DTT, PMSF, 0.004 U/ μ L SUPERase-In, and protease inhibitor

Sucrose pad:

20 mM HEPES, pH 7.6 (from pH 7.8 stock)
0.1% v/v IGEPAL CA-360
0.1 mM EDTA
1 mM spermine
1 mM spermidine
1 M sucrose
Complete with 1 mM DTT, PMSF, 0.004 U/ μ L SUPERase-In

Nuclei storage buffer:

20 mM HEPES, pH 7.6 (from pH 7.8 stock)
5 mM MgCl₂
25% v/v glycerol
Complete with 5 mM DTT

5X NWO Master Mix:

25 mM HEPES, pH 7.6 (from pH 7.8 stock)
6.25 mM MgCl₂
750 mM KCl
Complete with 6.25 mM DTT and 2 U/ μ L SUPERase-In
Complete 5X NWO Master Mix with water or 8 μ g/mL α -amanitin to make 4X Master Mix.

3X NWO Reaction Mix:

20 mM HEPES, pH 7.6 (from pH 7.8 stock)
5 mM MgCl₂
150 mM KCl
0.25 μ M α -³²P-CTP
1.5% w/v Sarkosyl
Complete with 5 mM DTT

Stop solution

Methods. Author manuscript; available in PMC 2020 April 15.

50 mM Trizma base, pH 8.0
100 mM NaCl
20 mM EDTA
1% w/v Sarkosyl
200 µg/mL torula yeast RNA

2.2. The Nuclear Walk-on Assay

2.2.1. Culturing and treating cells for nuclei isolation—Great consideration must be given to cell growth conditions and drug treatments prior to isolating nuclei. Pol II transcription is impacted by changes in growth conditions, discrepancies in which can give rise to very different measures of transcription even between experimental replicates. This subsection focuses on aspects of culturing and treating adherent cells, as the nuclei isolation procedure discussed later was developed for adherent cell lines. A 150 cm² rectangular cell culture flask, or T-150, can yield a sufficient amount of nuclei for about ten NWO reactions. The density at which the cells are grown prior to harvest of nuclei must be consistent between experimental conditions unless the effect of cell density is being studied. Ideally, for a transformed cell line, cells should be sufficiently confluent to provide ample nuclei in log phase growth at the time of harvest (80–90 percent confluent – 10–20 million cells). Care should be taken to make sure that cells are not overgrown, recovering from being overgrown, or grown under nutrient-depleted conditions. To diminish lag phase effects and the effects of new serum on gene expression, routinely plan to harvest cells at least 48 hours after plating. Plating equal numbers of cells is critical for accurate comparisons of transcription between conditions. If the planned experiment requires isolating nuclei from many separate flasks, then pool enough cells from flasks that are designated for sub-culturing, and plate cells from the pool. When working with a contact-inhibited cell line it is important to culture cells in the contact-inhibited state for a sufficient period of time so as to ensure that the growth phase of the cells is homogenous. Medium can be replenished as necessary, but this should be consistent between the cultures, because it will affect Pol II transcription.

Although many labs include the antibiotics penicillin and streptomycin in cell culture media, with proper sterile technique, these antibiotics are not necessary and should be avoided if possible. Furthermore, utilization of these antibiotics has been linked to significant changes in gene expression [23]. In addition, care should be taken ensure that the cells are not infected with mycoplasma, which is not responsive to conventional antibiotics and significantly changes the cellular metabolic state as well as the transcription of specific genes [24].

There are several variables to consider when planning to test how different cell treatments impact Pol II transcription. Treatments of cells with compounds or siRNAs should always be performed alongside appropriate controls. When treating cells with a compound, for instance, a vehicle control should be included. If the compound is in a non-aqueous solvent (e.g. DMSO), then the final concentration of DMSO present in the control and experimental conditions should be the same. An effort should be made to minimize the final concentration

of DMSO in the cell medium to less than 1%. When treating cells, it is best to add the compound in a large volume, to ensure even distribution over the cells. This can be achieved by pipetting the necessary volume of the compound into the bottom of a 15 mL conical tube. At the time of treatment, 5–10 mL of medium can be drawn from the flask in which the cells are growing, mixed with the compound in the conical tube, and gently re-dispensed into the cell culture flask. As discussed in the next sections, treatments of cells should be staggered such that enough time is allowed for harvesting of nuclei.

2.2.2. Preparation of Reagents for Nuclei Isolation—This subsection sets the stage for nuclei isolation and describes how buffer components permit the rapid isolation of nuclei and quickly lock engaged Pol II into place on chromatin. Four buffers must be assembled in preparation for nuclei isolation. These are 1X phosphate-buffered saline (PBS), a cell lysis buffer, a sucrose pad buffer, and a nuclei storage buffer. The amount of each buffer that must be prepared depends on the number and growth area of the flasks from which nuclei are being harvested. First, prepare as many 50 mL conical tubes containing approximately 40 mL of PBS as there are flasks of cells. Also label an equal number of 15 mL conical tubes (one for each flask) with details regarding the cell treatment. Place these conical tubes on ice. The 15 mL conical tubes will hold aliquots of the lysis buffer and be used as vessels for the lysed cells. Next, prepare the lysis buffer, sucrose pad buffer, and nuclei storage buffer. Remember to complete these buffers with DTT, PMSF, SUPERase-In, and the protease inhibitor cocktail tablet immediately before use, where relevant. Essential components of the lysis buffer are the detergent IGEPAL CA-360, which rapidly lyses the cells, and EDTA, which chelates magnesium to prevent further elongation by Pol II. In general, 10 mL of lysis buffer per 150 cm² of growth area is enough for cell lysis, and one can prepare $10 \text{ mL} \times (\text{growth area}/150 \text{ cm}^2) \times (\# \text{ of flasks} + 1)$ mL of lysis buffer for the experiment. The lysis buffer should be aliquoted upon its preparation into the pre-designated 15 mL conical tubes. The sucrose pad buffer contains 1 M sucrose, which will allow the passage of relatively dense nuclei during later centrifugation steps, but not small molecules like NTPs. A 10 mL sucrose pad can be used per 10 mL of lysed cells, and one can prepare $10 \text{ mL} \times (\# \text{ of flasks} + 1)$ of sucrose pad buffer for the experiment. For each flask of cells, a round-bottom, 50 mL polycarbonate tube should be placed on ice and labeled. Add to each of these tubes 10 mL of the prepared sucrose pad. Finally, prepare the nuclei storage buffer. We routinely re-suspend nuclei from a T150 flask of cells in 300 μL of storage buffer. The volume of storage buffer to prepare depends on the chosen resuspension volume and the number of nuclei preparations in the experiment.

During this time, one should also prepare the following necessary equipment and tools. A refrigerated centrifuge with a swinging bucket rotor that supports the 50 mL round-bottom polycarbonate tubes and $22,500 \times g$ should be powered on and cooled to 4°C. Cell scrapers (one per flask) should be set aside. Prepare a waste vessel into which medium and PBS from the flasks can be emptied. Prepare a work station for aspiration of the lysis buffer and sucrose pad from the polypropylene tubes containing the pelleted nuclei. Finally, prepare 1 mL glass Dounce homogenizers (ours were made by Wheaton) with tight pestles for resuspension of the nuclei. Fill the Dounces with nuclease-free water and a pestle and place

them on ice. Set aside a large aliquot of nuclease-free water on ice for rinsing of the Dounce between nuclei resuspensions.

2.2.3. Rapid Isolation of Nuclei—This section covers physical steps of nuclei isolation through resuspension and storage. The approach discussed herein is optimized for speed of cell lysis, which is key to quickly locking engaged Pol II into place and avoiding potential effects on transcription that come with lengthy cell manipulations. All nuclei isolation steps should be coordinated with cell treatments, to ensure that time of treatment is not a variable that could influence results. Note that the steps in section 2.2.2. can be performed between the last cell treatment and first isolation of nuclei, if this time window is at least 1 hour.

Our nuclei isolation protocol requires about 20 seconds from the time of decanting cell media from the flask to placing a suspension of lysed cells on ice. Due to the time that is required to access the tissue culture incubator and prepare each lysate, we recommend that treatments are staggered by 1.5 minutes. At the appropriate time, remove the cap from the flask and pour off the media. Immediately wash the cells with 40 mL of ice-cold PBS previously aliquoted into 50 mL conical tubes. To prevent cells from being dislodged, we recommend flipping over the cell culture flask so that the PBS pours onto the inner roof of the flask rather than on top of the cells, and then gently turning the flask back over to wash the cells. Decant the PBS and then pour in an aliquot of the lysis buffer from the designated 15 mL conical tube. Using a cell scraper, scrape the cells to help dislodge them from the surface of the plate, making sure that lysis buffer covers the area being scraped. Pour the suspension of cells back into the same 15 mL conical tube. Optionally, one can check the cells by phase contrast microscopy after addition of the lysis buffer. Loss of refractivity and darkening of nuclei is an indication that lysis has occurred.

Next, take the suspensions of lysed cells, and, using a serological pipette, gently lay them over the 10 mL sucrose pad in the designated polycarbonate tube. It is important to avoid significant mixing of the sucrose pad and cell lysate. Centrifuge the samples for 10 minutes at $22,500 \times g$. Carefully remove the tubes from the centrifuge and place them on ice. One should observe a white disc of pelleted nuclei at the bottom of the tube. Aspirate the supernatant using a Pasteur pipette connected to a vacuum source. Small molecules like NTPs will not have passed through the sucrose pad. Therefore, it is important that the tip of the pipette is very close to the top of the liquid at all times during the aspiration. This will ensure the removal of NTPs and prevent their downward migration towards the nuclear pellet. This is crucial, as contaminating NTPs in nuclei preps can result in extension of nascent transcripts and will compete with α - ^{32}P -CTP that is added during the nuclear walk-on reaction. Tilt the tube to remove the last bit of supernatant without touching the nuclear pellet.

To re-suspend nuclei, begin by adding 200 μL of storage buffer to each pellet. This ensures that the pellets do not dry out. Using a cut 200- μL pipette tip, gently break up the nuclear pellet and transfer as many pieces of the pellet as possible to a 1 mL Dounce that has been emptied of water and is on ice. With another 100 μL of storage buffer, rinse the bottom of the polycarbonate tube and collect any remaining pieces of the nuclear pellet to transfer to the Dounce. Then, using a tight pestle, homogenize the nuclei. These nuclei can be aliquoted

and stored at -80°C . Repeat this step for each pellet of nuclei and rinse the Dounce between uses.

2.2.4. The Nuclear Walk-On: Execution and Analysis of Data—With isolated nuclei, one can now perform a NWO assay. The NWO is a rapid, robust, and inexpensive way of asking how a tested experimental condition impacts Pol II pausing and productive elongation. In this method, engaged Pol II in isolated nuclei is allowed to incorporate α - ^{32}P -CTP to the 3' end of its associated nascent transcript when C is the next transcribed nucleotide. This reaction is performed in the presence of Sarkosyl, which relieves the effects of transcription factors and chromatin and prevents initiation from any free Pol II or preinitiation complexes. Labeled nascent transcripts are then isolated and resolved by denaturing PAGE. A paired α -amanitin treatment for each sample reveals signals from Pol I and III, which can be subtracted. Profile analysis quantifies the levels of paused and productively elongating Pol II. In addition, quantification of endogenous snRNAs on the ethidium bromide stained gel can be used to calculate correction factors for variable loading of nuclei. This subsection details the steps of the NWO assay from buffer preparation through data analysis.

Preparing for a NWO assay first involves attaining reagents and assembling reaction buffers. Among the buffers that must be prepared is a concentrated master reaction buffer that provides physiological salt and inhibits RNase. This buffer is prepared with and without α -amanitin to inhibit Pol II. One must also prepare a concentrated NWO mix that provides Sarkosyl and α - ^{32}P -CTP. In general, we use 15 μL of nuclei per reaction. This volume is brought to 20 μL with a 4X master mix, and to 30 μL with addition of a 3X NWO mix. Initially, the 5X master mix is prepared. An aliquot of the 5X master mix is then made 4X with the addition of H_2O or α -amanitin. At 4X concentration, the concentration of α -amanitin should be 8 $\mu\text{g}/\text{mL}$. In fresh, RNase-free 1.5 mL tubes, combine 15 μL of each nuclei sample with 5 μL of the 4X master mix with or without amanitin (two reactions per sample). Make sure that the thawed nuclei are re-suspended prior to pipetting as nuclei quickly settle at the bottom of the tube. The addition of the 4X master mix maintains the concentration of HEPES, magnesium chloride, and DTT, provides potassium chloride at 150 mM and SUPERase-In at 0.4 U/ μL , and does or does not introduce α -amanitin at 2 $\mu\text{g}/\text{mL}$.

Next, prepare the 3X NWO mix. Make sure to exercise the proper precautions when handling ^{32}P and properly dispose of radioactive waste. Warm the 20 μL samples and the 3X NWO mix to 37°C . With reference to a time table, add 10 μL of the 3X NWO mix to each of the samples. The final reaction condition is 20 mM HEPES, 5 mM MgCl_2 , 150 mM KCl, 5 mM DTT, 0.0833 μM α - ^{32}P -CTP, 0.5% Sarkosyl, and 0.133 U/ μL SUPERase-In. Additions can be staggered by as little as 30 seconds. Upon adding the 3X NWO mix, quickly cap the tube and vortex or flick the tube to mix the sample. Upon the addition of Sarkosyl, the sample becomes viscous and sticky due to the release of free DNA from chromatin. Pipetting the sample can result in its irrecoverable loss. Incubate the samples at 37°C for 10 minutes following the addition of the NWO mix. In theory, 25% of engaged polymerases will incorporate a labeled cytosine. This would be the case if the identity of the next transcribed nucleotide were random. However, a recent study has shown that promoter-proximal pausing in mammalian cells tends to occur at sites where C is the next transcribed

nucleotide [25]. Therefore, we would contend that the NWO labels more than 25% of the nascent transcripts associated with paused Pol II. To stop the reaction, add 90 μ L of stop solution. The EDTA quickly quenches the reaction and the addition of torula yeast RNA permits visualization of the precipitated RNA pellet during later steps.

After stopping every reaction, use TRIzol LS reagent to extract RNA from the samples per the manufacturer's instructions. Exercise caution with TRIzol reagent, as it contains phenol, which is corrosive. Contact with skin and eyes should be avoided. Upon recovery of an aqueous phase, add a 3X volume of 0.5 M ammonium acetate in 95% ethanol to each sample, thoroughly mix them, and incubate the samples at -80°C for 10 minutes. Afterwards, spin the samples in a refrigerated centrifuge at 13,000 rpm for 15 minutes. Carefully remove and discard the supernatant, and wash the pellet in 175 μ L of 70% ethanol. Gently maneuver the tube to wash all surfaces with the ethanol solution. This removes salt, an excess of which will retard migration of RNA on the gel. Spin the samples for 10 minutes at 13,000 rpm and discard the supernatant. Let the samples air-dry for 5 minutes. Afterwards, add 10 μ L of an RNA loading buffer that contains 0.25X Tris-borate EDTA (TBE), 9.35 M urea, 0.0075% w/v bromophenol blue, and 0.03% w/v xylene cyanol. Incubate the samples at 70°C for no more than 5 minutes, occasionally flicking them to resuspend the RNA.

During precipitation and centrifugation steps, cast a gel and prepare gel running buffer. We use large 20×20 cm gels (1 mm thickness) to allow ample separation of transcripts. We routinely cast 6% gels with 37.5:1 acrylamide:*bis*-acrylamide, 1X TBE, and 6 M urea, and use 1X TBE as a running buffer. Remove the comb, wash out the wells, and assemble the gel running apparatus just before loading the RNA samples. Using a gel-loading tip, carefully load each sample into a well of the gel. We run the minus/plus α -amanitin conditions side-by-side. After loading the samples, run the gel at constant wattage (10 watts) until the bromophenol blue has run two thirds of the way through the gel. Set a large Pyrex casserole dish onto a rocker and fill it with about one liter of water that contains 20 μ L of 10 mg/mL ethidium bromide. Disassemble the gel apparatus and check the gel with a Geiger counter. High counts at the bottom of the gel indicates free label. Using a razor blade, remove and dispose of this section of the gel before proceeding. Gently transfer the gel to the dish on the rocker and stain the gel for 15 minutes with gentle rocking. Carefully pour off the solution, add 1 liter of fresh water, and destain the gel with gentle rocking for another 10 minutes. Visualize the ethidium bromide stained gel with UV light making sure that the bands to be used for normalization are not saturated. Ribosomal RNA, small nuclear RNA, and transfer RNA all survive nuclei isolation and can be observed through ethidium bromide staining, and intensities of selected bands corresponding to these RNA species can be quantified to calculate correction factors for signal from radiolabeled nascent transcripts.

Transfer the gel to Whatman paper and dry the gel with heat under vacuum to completion. Expose the gel to a blanked phosphor screen. In our experience, the results of a NWO experiment that used fresh α - ^{32}P -CTP can be examined after an overnight exposure. Visualize the pattern of labeled transcripts via phosphorimaging. If one wishes to do so, ^{32}P -end-labeled or cold RNA markers can be incorporated into the experiment to tracks the sizes of labeled nascent transcripts. Densitometry analysis software such as the Fuji Film Multi

Gauge or GE ImageQuant packages are used to generate a profile for each lane, which is then corrected using the amount of stained RNA in that lane. In practice, we quantify several snRNAs and determine an average correction factor for each lane. After exporting the quantitative profile data into Excel, the intensities for each lane are multiplied by the correction factor. The + α -amanitin profiles are then subtracted from the paired profile without α -amanitin to remove signal associated with Pol I and Pol III. Figure 1A summarizes in bullet points the NWO protocol. We provide as an example the results of a NWO assay in which nuclei were harvested from control and Flavopiridol-treated cells. Flavopiridol inhibits the Cdk9 subunit of P-TEFb, thereby blocking pause-release. This results in increased occupancy of paused Pol II. Figure 1B shows the visualization of the ethidium-bromide stained gel and the phosphorimage. A profile analysis of the lanes with applied correction factors is shown. Note the increase in paused Pol II signal and the decrease in signal from productively elongating Pol II. In addition to examining the effects of Flavopiridol, we have used the NWO to study many other behaviors of Pol II, including effects of oxidative stress on Pol II elongation [26], knockdown of the elongation factor NELF [26], inhibition of XPB by triptolide [26], and treatment with the Cdk7 inhibitor THZ1 [27].

Overall, the results of a NWO can be attained in two days and provide quantitative information about levels of Pol II pausing and productive elongation in cells. The NWO is primarily useful for determining effects on levels of Pol II pausing and the lengths of paused Pol II nascent transcripts. Changes in levels of productive elongation can also be evaluated, although sensitivity of these measurements is reduced and long nascent Pol II transcripts run at the top of the gel. In addition to these measurements, the NWO assay can inform planning of expensive next generation sequencing experiments. We next move on to discuss how the NWO front end can be applied to PRO-Seq and a variant protocol, PRO-Cap.

2.3. Applications of the NWO Front End to PRO-Seq and PRO-Cap, and General Strategies for Improving Sequencing Data and Reducing Bias

PRO-Seq is a method that involves (i) a nuclear-run on with biotinylated NTPs, (ii), capture of biotinylated nascent transcripts with streptavidin beads, and (iii) massively parallel sequencing of the captured nascent transcripts. If performed properly, PRO-Seq can pin the location of the Pol II active site genome-wide at nucleotide resolution. A variant of PRO-Seq, called PRO-Cap, involves a negative selection against uncapped transcripts and provides enormous insight into sites of transcription initiation and pausing. The PRO-Seq and PRO-Cap protocols were developed by the Lis lab and have been detailed by Mahat et. al [21]. The section will focus on adaptation of the NWO front end for PRO-Seq and PRO-Cap, utilization of radiolabeled transcripts from NWOs and in vitro transcription to monitor the efficiency of library preparation steps, and useful, broadly applicable modifications to the protocol that can improve data, eliminate artifacts, and reduce bias.

2.3.1. Utilization of the NWO Front End for PRO-Seq and PRO-Cap—In section 2.2, we provided a protocol for rapid isolation of nuclei that should lock Pol II into place precisely where it was engaged on chromatin. We consider this an especially important aspect of an experiment that aims to examine sequence determinants of Pol II transcription

or asks how a cell treatment impacts Pol II occupancy. Nuclei that are isolated following our protocol can be used for PRO-Seq or PRO-Cap without any modification to the published PRO-Seq/Cap protocols [21] from the nuclear run-on step onward, except for a change in the volume of storage buffer used to suspend the nuclei. In addition, the same front end can be used for GRO-Seq or other methods that involve labeling of nascent Pol II transcripts by nuclear run-on.

2.3.2. Monitoring Library Preparation Steps with Radiolabeled Pol II

Transcripts—Ideally, one would like to be able to track and quantify the efficiencies of the many steps involved in PRO-Seq and PRO-Cap library preparation. We have utilized radiolabeled transcripts generated by NWO and in vitro transcription to monitor key steps. Detection of isolated biotinylated nascent transcripts requires the RNA to be ^{32}P -labeled because the low level of RNA generated cannot be observed using standard RNA stains such as ethidium bromide. The experiments below are examples of how the NWO and in vitro transcription can be utilized to monitor and optimize the efficiency of library preparation.

To measure the efficiency of biotinylated NTP incorporation and follow nascent transcripts through binding and enrichment with streptavidin beads, we labeled Pol II transcripts with α - ^{32}P -CTP by NWO and followed with addition of all four biotinylated NTPs. RNA was then isolated by TRIzol extraction and incubated with paramagnetic M-280 streptavidin beads in 1 M sodium chloride for 30 minutes at room temperature, with rotation. Afterwards, the unbound fraction was taken, and the beads were subjected to washing in a high salt or a low salt plus detergent condition. Equal fractions of the bound RNA were either set aside for analysis or taken once more through the process of bead-binding and washing. Biotinylated RNAs were released from streptavidin beads by TRIzol extraction. Ultimately, we collected two bound and two unbound fractions. These were each analyzed by denaturing PAGE and phosphorimaging (Fig. 2A). We quantified the efficiency of biotinylated NTP incorporation and found that about 35% of labeled transcripts incorporated biotin-11-NTPs. Importantly, biotinylated nascent transcripts that were bound by the streptavidin beads were not lost between successive rounds of binding, washing, and extraction of the bound material. This strategy could be used to further optimize nuclear run-on reaction conditions and the conditions for enrichment of biotinylated RNAs. A similar approach could also be used to follow the enrichment of nascent transcripts labeled with BrU with anti-BrU antibodies [20].

We have also used radiolabeled RNAs generated by in vitro transcription to compare the efficiency of RNA adapter ligation by T4 RNA ligase 1 (Rnl1) and T4 RNA ligase 2 K227Q (Rnl2 K227Q). In brief, in vitro transcription was carried out using a DNA template immobilized on M-280 beads. We have described in detail this readily manipulable in vitro transcription system in a previous methods paper [28]. Transcripts were pulse-labeled for 30 seconds following addition of nucleotides with limiting α - ^{32}P -CTP. The array of transcripts that are synthesized reflect Pol II stopping at positions on the template where C is the next transcribed base. After 30 seconds, the reaction was quenched. We then we isolated elongation complexes and either extracted the RNA or allowed Pol II in the elongation complexes to incorporate biotinylated CTP, and then isolated RNA. Labeled transcripts that were and were not extended with biotinylated CTP were then used to examine the efficiency of 3' RNA adapter ligation by Rnl1 and Rnl2 K227Q (Fig. 2B). Importantly, while Rnl1

ligates two RNA molecules, Rnl2 K227Q can ligate the 3' end of an RNA molecule to a 5' adenylated DNA molecule. Adenylation activates the 5' end of the oligo and allows ligation to occur in the absence of added ATP. For Rnl1, an RNA adapter of the sequence /5Phos/rGrArUrCrGrUrCrGrGrArCrUrGrUrArGrArArCrUrCrUrGrArArC/3InvdT/ was used. For Rnl2 K227Q, an adenylated DNA adapter of the sequence /5rApp/NNNNTGGAATTCTCGGGTGCCAAGG/3ddC/ was used. Ligation reactions were performed for 1 or 4 hours and the reaction products were analyzed by urea-PAGE and phosphorimaging. Addition of biotinylated CTP is reflected by a decrease in transcript mobility on the gel that corresponds to approximately four nucleotides and in this in vitro reaction, incorporation was efficient. Our findings indicate that both Rnl1 and Rnl2 K227Q can act on biotinylated transcripts and that increasing the time of the ligation reaction from 1 hour to 4 hours significantly increases the fraction of transcripts that underwent ligation to an adapter. Rnl1 and Rnl2 are similarly efficient in ligation of adapters to biotinylated transcripts. Since adenylated adapters are very expensive, we use Rnl1. This experimental scheme could be applied to radiolabeled, biotinylated nascent transcripts that are generated by a NWO and better reflect the pool of nascent transcripts that one aims to prepare into a PRO-Seq library. Together, these data illustrate the utility of the NWO and in vitro transcription in optimizing steps of sequencing library preparation.

2.3.3. The Utility of Unique Molecular Identifiers, Approaches to Reducing Bias, and Considerations for PCR Amplification and Size-Selection of Libraries

Preparation of RNA and DNA sequencing libraries generally involves PCR amplification of the library. This step introduces PCR duplicates, which should be excluded from downstream analyses. PRO-Seq and PRO-Cap capture nascent transcripts. It is important to recognize that transcription from highly utilized promoters may generate many transcripts with the same 5' and 3' ends. It would be ideal to retain such independently generated identical RNAs during quantitative analyses of initiation and pausing. As shown in Figure 3, this issue can be addressed by unique molecular identifiers (UMIs), which are short stretches of random nucleotides that can be added to the ends of the adapters that are ligated onto the inserts. These UMIs are sequenced, and two fragments that share the same 5' and 3' end and same UMIs can be considered PCR duplicates and computationally removed. UMIs can be incorporated onto one adapter for single-end sequencing or both adapters for paired-end sequencing. The level of redundancy that can be captured by UMIs is dictated by their length. For example, in a paired-end sequencing experiment, ligating 5' and 3' adapters that each contain 4 nucleotide-long UMIs would permit the identification of 4^8 , or 65,536 unique nascent RNAs with the same 5' and 3' ends. For PRO-Seq and PRO-Cap, we have written a deduplication algorithm called dedup (<https://github.com/P-TEFb/dedup>) that accepts as input paired-end, aligned sequencing data in the form of a SAM file, removes PCR duplicates based on UMIs, and returns a BED file (Fig. 3B). Dedup can be parameterized for PRO-Seq and PRO-Cap to trim off the 3' nucleotide corresponding to the biotinylated nucleotide that was added during the nuclear run-on reaction. In addition to being useful for removing PCR duplicates, UMIs increase the sequence diversity between the ends of the insert RNAs and adapters, which reduces ligation bias [29].

Library preparation is burdened by the inherent biases of all enzymes involved. Ligation bias and PCR bias are thought to significantly impact the results of sequencing experiments. Algorithms now exist that correct for these biases computationally [30]. However, bias reduction should fully be practiced during the actual steps of library preparation. For example, PCR can bias the content of the sequencing library. Amplification of GC-rich DNA is challenging to DNA polymerases, which can lead to an under-representation of GC-rich molecules in sequencing libraries. We and others have found the KAPA HiFi polymerase (Kapa HiFi HotStart ReadyMix, KAPA Biosystems, KK2601) out-performs other commercial DNA polymerases in the amplification of GC-rich DNA and is relatively uninhibited by amplification of even very long PCR products [29, 31]. We have recently moved to use KAPA HiFi for PCR amplification of all sequencing libraries.

Analysis of PCR products during the preparation of sequencing libraries is required to minimize the number of cycles of PCR needed. In Figure 4A, PRO-Seq and PRO-Cap libraries were subjected to eight, 4-fold serial dilutions from the libraries were each subjected to 24 cycles of PCR, as described by Mahat et. al [21], to determine the number of PCR cycles needed for library amplification. The products of the PCR reaction were separated by TAE PAGE and visualized by ethidium bromide staining and UV light. A 25-base ladder was included on the gel. From the results shown the number of PCR cycles needed to amplify the PRO-Seq and PRO-Cap libraries were 8 and 12, respectively. The gel shows a zero-insert product at 126 base-pairs that arises due to ligation of the 5' and 3' adapters. The zero-insert product can be a significant component of sequencing libraries following PCR amplification. The library of interest runs as a smeared pattern above the zero-insert product and is indicated for both the PRO-Seq and PRO-Cap datasets. Note that as the primers are depleted, one begins to observe a smear of products that runs high on the gel. These products are called daisy-chains, and they are thought to be formed by the annealing of the ends of multiple ssDNA molecules when primer is absent. Daisy chains form following the final denaturation step in the PCR reaction and do not reflect further amplification of the library. One should make certain that primers are not depleted upon completion of the chosen number of cycles for library amplification, as a depletion of primers would indicate that the reaction is no longer in exponential phase.

A size-selection should be performed to remove unused primers and the zero-insert product from the PCR-amplified library. Both can negatively impact sequencing runs. Free primer has been linked to index hopping, which can result in the mis-assignment of reads in multiplexed libraries [32]. Contaminating zero-insert product would take up space on the flow cell and reduce the amount of data yielded by the run and could potentially interfere with detection of fluorescent bases during the sequencing run by reducing cluster diversity. The zero-insert product runs very close to the library of interest. To effectively remove the zero-insert product and unused primers, we size-select our libraries using a Sage Science Blue Pippin. The Blue Pippin uses high resolution agarose gel electrophoresis with fluorescent standards to direct DNA molecules of a user-specified size range into a collection reservoir, while excluding products that are smaller and larger. Shown in the top panel of Figure 4B is an electropherogram of a PCR-amplified library before size selection that was generated using an Agilent 2100 Bioanalyzer System. Arrows indicate the zero-insert product (ZI) and unused primer (UP). In the lower panel is an electropherogram of the

pooled and size-selected library. This profile is representative of ten individual libraries that were pooled for size-selection and multiplexed sequencing. Both the zero-insert product and unused primers were efficiently removed using the Blue Pippen.

2.4. Why to Normalize Sequencing Data with Spike-in Controls

Many approaches are used to normalize next-generation sequencing datasets for proper comparison. Data is often normalized such that the same amount of total signal is present in each dataset. This approach falsely assumes that the cells from which material was isolated produce the same total amount of the DNA or RNA species of interest between conditions. Consider that exposure of cells to certain conditions may result in there being globally more or less transcription per cell compared to the untreated control. This would have tremendous implications for a PRO-Seq or PRO-Cap experiment. Preparing equimolar amounts of sample from these two conditions for sequencing and normalizing to total signal would preclude the observation of globally up or down-regulated transcription (Fig. 5A). Only the inclusion of a spike-in control permits a reliable quantification of these differences.

The spike-in provides the same type of molecule that is being measured in the samples of interest. The molecules that are contributed by the spike-in must be differentiable from those that come from experimental samples. Libraries are sequenced in roughly equimolar amounts and the fraction of reads that originate from the spike-in can be used to normalize the reads from the experimental sample as will be described below. Many studies have used ERCC Spike-in controls for quantitative normalization of RNA-Seq datasets. The ERCC spike-in is a pool of synthetic RNA molecules of known sequence and concentration that are added to cellular samples during RNA extraction. These RNAs have Poly(A) tails and can be processed alongside cellular RNA for sequencing using Poly(A) or total RNA-Seq pipelines. Reads that map to the spike-in RNAs are quantified, and their relative levels with respect to each other and the reads that map to the genome of interest can be used for normalization and as a measure of quality control [33–35]. For PRO-Seq and PRO-Cap, the ERCC spike-in is not ideal because the spike-in transcripts cannot be carried through library preparation (spike-in RNAs are not biotinylated). Instead, cells can be spiked in at the time of cell lysis or nuclei can be spiked in just prior to the nuclear run-on [36, 37](Fig. 5B). Engaged Pol II in the spike-in nuclei is allowed to incorporate biotinylated nucleotides during a nuclear run-on reaction alongside Pol II the experimental nuclei. In this manner, the spike-in sample controls for all steps of library preparation.

2.4.1. The Spike-in Control: Choice of Sample and Integration into

Experimental Workflow—A key aspect of the spike-in control is that it must contribute DNA or RNA molecules that can be traced back to their origin. If one were testing how depletion of a transcription factor impacts Pol II occupancy in a human cell line, then spiking in insect cells, for example, would be appropriate. Cells of *Drosophila* or *Spodoptera* origin could be chosen, as they are relatively common in laboratories and genomic sequences are available for both. Determining the number of spike-in reads requires that the organism from which the spike-in sample was derived has a genome sequence available for alignment.

PRO-Seq and PRO-Cap experiments begin with isolation of nuclei from cell cultures that should contain equal numbers of cells. An approach to achieve this was discussed in section 2.2.1. Introduction of spike-in cells at the nuclei isolation step is appropriate for PRO-Seq and PRO-Cap (Fig. 5B). Alternatively, spike-in nuclei could be added just before the nuclear run-on. It is important that one be able to process the spike-in sample in the same way as the samples of interest, and it is critical that each sample receive the same amount of spike-in sample. This requires that the spike-in sample be completely homogenous. One should aim for the number of spike-in reads to comprise 1–3% of the total reads in a sequenced sample. This provides sufficient data for normalization and does not dramatically reduce depth of coverage for the samples of interest.

2.4.2. Normalizing Sequencing Data with Spike-in Reads—Upon retrieval of sequencing data, reads are mapped to both the genome of the experimental cell line and the spike-in cell line. This can be done by mapping reads to one genome and reserving the unmapped reads for mapping to the other genome, but a better method is to map both at the same time to a concatenated genome and later parse out the reads based on the genome to which they aligned. This allows the mapping program to make decisions about the best match for each sequence. Sequencing libraries are pooled in approximately equimolar ratios for sequencing. Errors in quantification of library concentrations during pooling and other issues lead to a different number of total mapped reads for each library. PRO-Seq datasets with spike-in controls are normalized by correcting for the total number of reads in each library, as well as the fraction of reads from the spike-in. Table 1 provides examples for how the corrections are made to human PRO-Seq datasets. First a library size correction factor is calculated and used to correct both the human and spike-in reads. A spike-in correction factor is then calculated that takes into account the competition between human and spike-in RNA in generating libraries (Fig. 5C). The initial mapped reads are multiplied by both factors, which together correct for differences in library size and spike-in reads. Note, the small library (#2) has the biggest size correction factor. Library 4 had the most spike-in reads and this led to a downward correction for the number of human reads.

3. Concluding Remarks

Here we discussed approaches to studying Pol II transcription through examination of nascent transcripts. We provided a method for rapidly isolating nuclei from cells that quickly locks Pol II into place. This is a useful front end not only for our NWO method, but also GRO-Seq, PRO-Seq, and PRO-Cap, each of which involve labeling of nascent transcripts in nuclear run-on reactions. We detailed a protocol for the NWO assay, which permits the semi-global analysis of levels of Pol II pausing and productive elongation. Both the NWO assay and in vitro transcription provide radiolabeled Pol II transcripts, which are useful tools for monitoring and optimizing library preparation steps for PRO-Seq and PRO-Cap. Finally, we broadly discussed the utility of UMIs and considered strategies for reducing ligation and PCR bias. In the end, we provide an approach to the utilization of spike-in controls for normalization of PRO-Seq and PRO-Cap datasets, though the approach is broadly applicable to other next generation sequencing methods.

Acknowledgments

Funding: This work was supported by the National Institute of General Medical Science grants [R01-GM35500 to DHP, R01-GM113935 to DHP and Donal Luse (both just replaced by R35-GM126908 to DHP)]; the National Institute of Allergy and Infection Diseases grant [R21-AI130453 to DHP and Jeff Meier]

References

1. Gilchrist DA, et al., Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell*, 2010 143(4): p. 540–51. [PubMed: 21074046]
2. Gilchrist DA, et al., NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev*, 2008 22(14): p. 1921–33. [PubMed: 18628398]
3. Mahat DB, et al., Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Mol Cell*, 2016 62(1): p. 63–78. [PubMed: 27052732]
4. Vos SM, et al., Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature*, 2018 560(7720): p. 607–612. [PubMed: 30135578]
5. Zhou Q, Li T, and Price DH, RNA polymerase II elongation control. *Annu Rev Biochem*, 2012 81: p. 119–43. [PubMed: 22404626]
6. Jonkers I and Lis JT, Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol*, 2015 16(3): p. 167–77. [PubMed: 25693130]
7. Peterlin BM and Price DH, Controlling the elongation phase of transcription with P-TEFb. *Mol Cell*, 2006 23(3): p. 297–305. [PubMed: 16885020]
8. Lin CY, et al., Transcriptional amplification in tumor cells with elevated c-Myc. *Cell*, 2012 151(1): p. 56–67. [PubMed: 23021215]
9. Hanahan D and Weinberg RA, Hallmarks of cancer: the next generation. *Cell*, 2011 144(5): p. 646–74. [PubMed: 21376230]
10. Gonda TJ and Ramsay RG, Directly targeting transcriptional dysregulation in cancer. *Nat Rev Cancer*, 2015 15(11): p. 686–94. [PubMed: 26493648]
11. Nilson KA and Price DH, The Role of RNA Polymerase II Elongation Control in HIV-1 Gene Expression, Replication, and Latency. *Genet Res Int*, 2011 2011: p. 726901. [PubMed: 22567366]
12. Chen CP, et al., Kaposi's Sarcoma-Associated Herpesvirus Hijacks RNA Polymerase II To Create a Viral Transcriptional Factory. *J Virol*, 2017 91(11).
13. Johnson DS, et al., Genome-wide mapping of in vivo protein-DNA interactions. *Science*, 2007 316(5830): p. 1497–502. [PubMed: 17540862]
14. Shao W and Zeitlinger J, Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet*, 2017 49(7): p. 1045–1051. [PubMed: 28504701]
15. Rhee HS and Pugh BF, Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*, 2011 147(6): p. 1408–19. [PubMed: 22153082]
16. Rhee HS and Pugh BF, ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr Protoc Mol Biol*, 2012 Chapter 21: p. Unit 21 24.
17. Gilmour DS and Fan R, Detecting transcriptionally engaged RNA polymerase in eukaryotic cells with permanganate genomic footprinting. *Methods*, 2009 48(4): p. 368–74. [PubMed: 19272453]
18. Churchman LS and Weissman JS, Native elongating transcript sequencing (NET-seq). *Curr Protoc Mol Biol*, 2012 Chapter 4: p. Unit 4 14 1–17.
19. Churchman LS and Weissman JS, Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, 2011 469(7330): p. 368–73. [PubMed: 21248844]
20. Core LJ, Waterfall JJ, and Lis JT, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 2008 322(5909): p. 1845–8. [PubMed: 19056941]
21. Mahat DB, et al., Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc*, 2016 11(8): p. 1455–76. [PubMed: 27442863]

22. Kwak H, et al., Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, 2013 339(6122): p. 950–3. [PubMed: 23430654]
23. Ryu AH, et al., Use antibiotics in cell culture with caution: genome-wide identification of antibiotic-induced changes in gene expression and regulation. *Sci Rep*, 2017 7(1): p. 7533. [PubMed: 28790348]
24. Olarerin-George AO and Hogenesch JB, Assessing the prevalence of mycoplasma contamination in cell culture via a survey of NCBI's RNA-seq archive. *Nucleic Acids Res*, 2015 43(5): p. 2535–42. [PubMed: 25712092]
25. Tome JM, Tippens ND, and Lis JT, Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat Genet*, 2018 50(11): p. 1533–1541. [PubMed: 30349116]
26. Nilson KA, et al., Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. *Nucleic Acids Res*, 2017 45(19): p. 11088–11105. [PubMed: 28977633]
27. Nilson KA, et al., THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing. *Mol Cell*, 2015 59(4): p. 576–87. [PubMed: 26257281]
28. Cheng B and Price DH, Isolation and functional analysis of RNA polymerase II elongation complexes. *Methods*, 2009 48(4): p. 346–52. [PubMed: 19409997]
29. van Dijk EL, Jaszczyszyn Y, and Thermes C, Library preparation methods for next-generation sequencing: tone down the bias. *Exp Cell Res*, 2014 322(1): p. 12–20. [PubMed: 24440557]
30. Wang JR, Quach B, and Furey TS, Correcting nucleotide-specific biases in high-throughput sequencing data. *BMC Bioinformatics*, 2017 18(1): p. 357. [PubMed: 28764645]
31. Quail MA, et al., Optimal enzymes for amplifying sequencing libraries. *Nat Methods*, 2011 9(1): p. 10–1. [PubMed: 22205512]
32. Costello M, et al., Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics*, 2018 19(1): p. 332. [PubMed: 29739332]
33. Jiang L, et al., Synthetic spike-in standards for RNA-seq experiments. *Genome Res*, 2011 21(9): p. 1543–51. [PubMed: 21816910]
34. Chen K, et al., The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Mol Cell Biol*, 2015 36(5): p. 662–7. [PubMed: 26711261]
35. Loven J, et al., Revisiting global gene expression analysis. *Cell*, 2012 151(3): p. 476–82. [PubMed: 23101621]
36. Dukler N, et al., Nascent RNA sequencing reveals a dynamic global transcriptional response at genes and enhancers to the natural medicinal compound celastrol. *Genome Res*, 2017 27(11): p. 1816–1829. [PubMed: 29025894]
37. Parua PK, et al., A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature*, 2018 558(7710): p. 460–464. [PubMed: 29899453]

Highlights:

- Profiling of nascent transcripts captures Pol II occupancy at high resolution
- Rapid isolation of nuclei is critical for quickly locking Pol II into place
- The nuclear walk-on is a robust method for quantifying Pol II transcription
- The nuclear walk-on is readily adapted as a front end for PRO-Seq and PRO-Cap
- Spike-in controls are important tools for normalization of sequencing data

Figure 1.

Overview of the Nuclear Walk-On Assay. (A) Workflow of the NWO protocol. (B) A representative NWO comparing levels of Pol II pausing and productive elongation in cells that were treated with a DMSO control or the P-TEFb inhibitor Flavopiridol for 1 hour. Shown are the ethidium bromide stained gel (left), phosphorimage (middle), and Pol II profile after correcting for loading (right). Sizes of RNAs are indicated: 75 nt (tRNA), 160 nt (5.8S rRNA), 330 nt (7SK), 1800 nt (18S rRNA). Reactions were performed with or without α -amanitin and size markers (in nucleotides) are indicated.

Figure 2.

Utilization of the nuclear walk-on assay and in vitro transcription to quantify and track efficiencies of PRO-Seq and PRO-Cap library preparation steps. (A) Schematic of a modified NWO assay that was used to monitor biotinylated NTP incorporation and follow biotinylated Pol II nascent transcripts through streptavidin bead selection (top). Phosphorimage results are shown in the lower panel. Bound (B) and unbound (U) fractions from the first (1) and second (2) rounds of affinity selection are indicated. (B) Schematic of experiment performed to quantify the efficiency of 3' adapter ligation by RNA ligase 1 and

RNA Ligase 2 K227Q (top). Phosphorimage results are shown in the lower panel. Ligation times are indicated. The difference in size of the Rnl1 and Rnl2 KQ ligation products is due to a difference in the length of the adapters used for their respective reactions. Rnl1 – RNA ligase 1, Rnl2 KQ – RNA ligase 2 K227Q.

Figure 3.

Use of unique molecular identifiers (UMIs). (A) UMIs on the RNA adapters are indicated by the short stretches of Ns at the 3' end of the 5' adapter and 5' end of the 3' adapter. The biotinylated nascent transcript insert is shown between. (B) Dedup, a program written by our lab, takes paired-end, genome-aligned sequencing data in the form of a SAM file and removes all but one of the identical reads, thereby eliminating PCR duplicates. Inserts that possess the same UMIs but different 5' and 3' ends are not removed. The UMIs are indicated by the colored bases. Depictions of the data before (left) and after (right) deduplication are shown.

Figure 4.

PCR amplification and size-selection of PRO-Seq and PRO-Cap libraries. (A) TAE-PAGE analysis of PCR products formed during amplification of libraries. A fraction of each reverse transcribed library was serially diluted and subjected to 24 cycles of PCR. The numbers at the top of the gel indicate the number of cycles of PCR that would be required to achieve the same level of amplification from the remaining, undiluted library. Unused primers, the zero-insert product, libraries, and daisy chains are indicated. (B) Size selection of sequencing libraries using the Blue Pippen. Shown are electropherograms generated using an Agilent

Bioanalyzer 2100 for a library that was PCR amplified and not size-selected (top) and a pooled library that was carried through size-selection (bottom). The peaks corresponding to unused primers (UP) and the zero-insert product (ZI) are indicated. The peaks at the far left and right edges of the electropherograms come from added DNA standards, and the blue dashed lines indicate a region that was specified for calculation of the molar concentration of DNA molecules that fell within the size-range of interest (145–600 bp).

Figure 5.

Utility and application of spike-in controls for PRO-Seq and PRO-Cap. (A) Rationale for using spike-in controls. The left graph depicts the actual transcriptional output per cell under conditions in which transcription is globally up or downregulated relative to a control. The right graph depicts the failure to detect differential transcription when the same data are normalized to total mapped reads. (B) A rudimentary workflow for a PRO-Seq for PRO-Cap experiment. Spike-ins can be integrated into the workflow at the time of cell lysis or nuclei isolation. (C) Total reads from two equally sized PRO-Seq libraries that were generated from equal numbers of human and spike-in cells are shown. In the case where transcription is globally upregulated in human cells, the human RNAs out compete the spike-in RNAs. The opposite is true when transcription is globally downregulated in the human cells.

Table 1.

Normalization of sequencing data using spike-in reads. This table contains hypothetical data for four separate PRO-Seq libraries that examine transcription under various conditions in a human cell line. Insect spike-in cells were included. Shown are the total mapped reads following deduplication (UMIs were utilized), human reads, and spike-in reads. All reads are in millions. First, to correct for differences in library size, the average total reads are calculated for all samples, and this average is divided by the total reads in each sample to obtain a library size correction factor. Next, the human and spike-in reads in each sample are multiplied by their library size correction factor. To calculate spike-in correction factors for human libraries, the average of the size-corrected spike-in reads is divided by the size-corrected spike-in reads in each sample. The human reads are then multiplied by their spike-in correction factors.

	Initial mapped reads			Library size corrected			Spike-in corrected	
	Total	Human	Spike-in	Correction	Human	Spike-in	Correction	Human
1	20	19.8	0.2	0.875	17.325	0.175	1.125	19.49
2	10	9.9	0.1	1.750	17.325	0.175	1.125	19.49
3	20	19.9	0.1	0.875	17.413	0.088	2.250	39.18
4	20	19.6	0.4	0.875	17.150	0.350	0.563	9.65
Avg.	17.5					0.197		