



Published in final edited form as:

Cell Rep. 2020 November 10; 33(6): 108373. doi:10.1016/j.celrep.2020.108373.

## Rapid and Scalable Profiling of Nascent RNA with fastGRO

Elisa Barbieri<sup>1,4</sup>, Connor Hill<sup>1,2,4</sup>, Mathieu Quesnel-Vallières<sup>3</sup>, Avery J. Zucco<sup>1</sup>, Yoseph Barash<sup>3</sup>, Alessandro Gardini<sup>1,5,\*</sup>

<sup>1</sup>The Wistar Institute, Gene Expression and Regulation Program, 3601 Spruce Street, Philadelphia, PA 19104, USA

<sup>2</sup>Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Boulevard, Philadelphia, PA 19104, USA

<sup>3</sup>Department of Genetics, Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Boulevard, Philadelphia, PA 19104, USA

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead Contact

### SUMMARY

Genome-wide profiling of nascent RNA has become a fundamental tool to study transcription regulation. Unlike steady-state RNA-seq, nascent RNA profiling mirrors real-time activity of RNA polymerases and provides an accurate readout of transcriptome-wide variations. Some species of nuclear RNAs (i.e., large intergenic noncoding RNAs [lincRNAs] and eRNAs) have a short half-life and can only be accurately gauged by nascent RNA techniques. Furthermore, nascent RNA-seq detects post-cleavage RNA at termination sites and promoter-associated antisense RNAs, providing insights into RNA polymerase II (RNAPII) dynamics and processivity. Here, we present a run-on assay with 4-thio ribonucleotide (4-S-UTP) labeling, followed by reversible biotinylation and affinity purification via streptavidin. Our protocol allows streamlined sample preparation within less than 3 days. We named the technique fastGRO (fast Global Run-On). We show that fastGRO is highly reproducible and yields a more complete and extensive coverage of nascent RNA than comparable techniques can. Importantly, we demonstrate that fastGRO is scalable and can be performed with as few as  $0.5 \times 10^6$  cells.

### In Brief

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\*Correspondence: [agardini@wistar.org](mailto:agardini@wistar.org).

#### AUTHOR CONTRIBUTIONS

E.B. and A.G. designed the project. E.B. and C.H. performed and designed all experiments and analyzed NGS data. M.Q.-V. performed computational splicing analysis. Y.B. provided support for the splicing analysis and critical review of the manuscript. A.Z. performed differentiation of human iPSCs and all related controls. E.B., C.H., and A.G. wrote the paper. All co-authors read and approved the manuscript.

#### DECLARATION OF INTERESTS

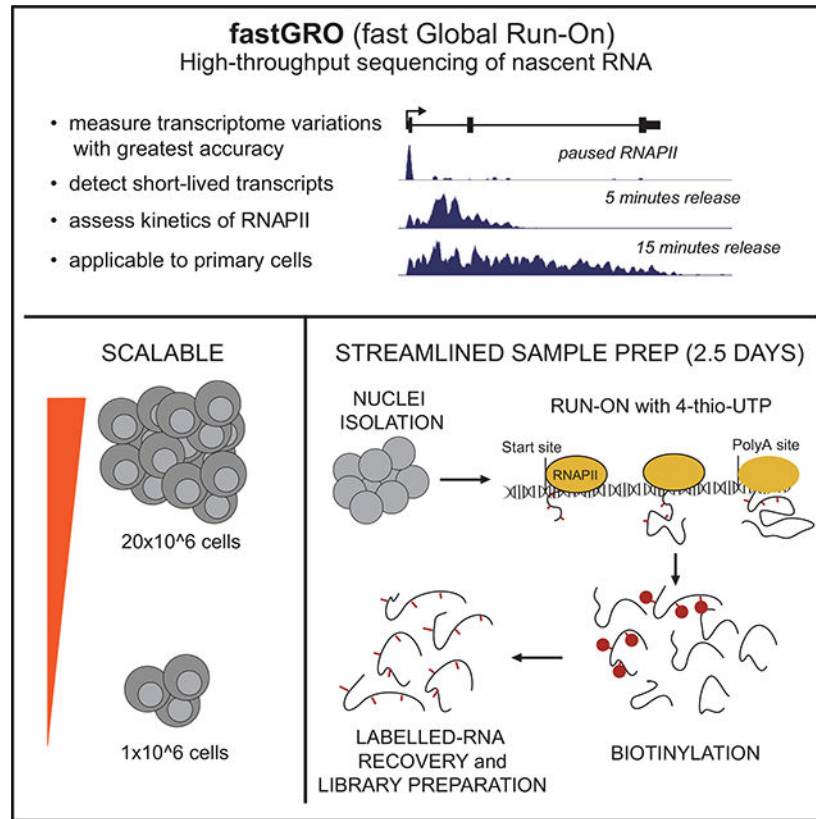
The authors declare no competing interests.

#### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.108373>.

Barbieri et al. developed fastGRO, a nascent RNA-sequencing technique based on nuclear run-on. Using a streamlined, under-3-days protocol, fastGRO tracks the activity of RNA polymerase for differential gene expression analysis, polymerase kinetic studies, and profiling of lowly expressed and unstable RNA species. A low-input fastGRO protocol profiles nascent RNA in as little as  $0.5 \times 10^6$  cells.

## Graphical Abstract



## INTRODUCTION

In slightly more than a decade, next-generation sequencing (NGS) technology has revolutionized the field of transcription by allowing precision mapping of most RNA species, from mRNAs to large intergenic noncoding RNAs (lincRNAs). Usually, RNA is extracted from crude cell extracts via acidic phenol-chloroform precipitation and either reverse transcribed with oligo(dT) or subjected to ribosomal RNA depletion, first, followed by reverse transcription using a pool of short random oligos (thus avoiding the polyadenylation bias). Adaptor ligation and PCR-based amplification convert the original pool of RNAs into a sequencing-ready library that will generate quantitative transcriptomic profiles (Stark et al., 2019). Regardless the specific protocol of choice, there are several limitations to these widely used RNA-sequencing (RNA-seq) techniques. First, traditional RNA-seq measures steady-state, mostly cytoplasmic, RNA species. Steady-state RNA levels are the ultimate result of synthesis rate, RNA processing, and RNA stability. Both RNA

processing and stability are highly regulated in every cell type (Schoenberg and Maquat, 2012; Pai and Luca, 2019; Yamada and Akimitsu, 2019). Therefore, RNA-seq alone is insufficient to infer the accurate transcriptional activity of any given gene. Additionally, transcription by RNA polymerase II (RNAPII) is not a steady and passive process of ribonucleotide chain assembly. There are multiple, critical, regulatory steps and checkpoints all along the transcription cycle that cannot be discerned with the resolution offered by RNA-seq (Adelman and Lis, 2012; Kwak and Lis, 2013; Proudfoot, 2016). Lastly, there are low-abundant and poorly stable RNA species that fall below the RNA-seq detection threshold. For instance, biologically active enhancer RNAs (eRNAs) and other lincRNA species are hardly represented in conventional transcriptomic data (Lai and Shiekhattar, 2014; Gardini and Shiekhattar, 2015). To overcome limitations of RNA-seq, several groups have developed high-throughput methods that tap into the so-called “nascent” fraction of cellular RNA (Wissink et al., 2019). Nascent transcripts represent the small RNA fraction (>0.5% of total RNA content in a cell) that is actively synthesized and still associated with RNA polymerase. Global run-on sequencing (GRO-seq) was the first genome-wide technique developed to probe nascent transcription genome-wide (Core et al., 2008). GRO-seq yields an exact footprint of already engaged RNA polymerase by building upon the strengths of a 40-year-old assay (Gariglio et al., 1974, 1981). In GRO-seq, nuclei are isolated and flash-frozen only to resume transcription *in vitro* in the presence of a labeled nucleotide (Core et al., 2008; Gardini, 2017). Precision nuclear run-on (NRO) sequencing (PRO-seq) was developed years later as a modification of GRO-seq using biotinylated nucleotides (Core et al., 2014). Both techniques are time consuming and marred by non-standardized library preparation (Mahat et al., 2016; Gardini, 2017). Another popular method for deep sequencing of nascent RNA, transient transcriptome sequencing ([TT-seq] as well as its parent technique, 4-thiouridine sequencing [4SU-seq]), relies on metabolic labeling of RNA but also recovers partly and fully processed RNA that is not associated with RNAPII (Schwanhäusser et al., 2011; Schwab et al., 2016). Additional strategies to purify RNAPII-associated transcripts, such as native elongating transcript sequencing (NET-seq) and mammalian NET-seq (mNET-seq), are biased toward identifying pausing sites of polymerase and depend on affinity purification or subcellular fractionation (Mayer et al., 2015; Nojima et al., 2015).

We have developed a run-on technique (fast Global Run-On; fastGRO) that allows robust mapping of the nascent transcriptome in under 3 days. Our technique optimizes the usage of 4-thio ribonucleotides (4-S-UTPs) for NRO assays. We take advantage of reversible biotinylation to label and enrich for newly synthesized RNA species, and we ultimately generate strand-specific libraries for Illumina sequencing using commercially available prep kits. Here, we use fastGRO to measure nascent RNA in HeLa and THP1 cells, and we compare our technique to a variety of nascent RNA assays that have been widely adopted over the past years. While reducing processing time by more than half, we show that fastGRO yields more consistent coverage across gene bodies (spanning introns and post-termination RNAs) than other benchmark techniques and can reliably gauge the kinetics of RNAPII. We also find that processed RNA contamination is significantly lower in fastGRO. We use fastGRO to measure a variety of RNA species, including low-abundant lincRNAs and short-lived eRNAs and antisense promoter transcripts. A major limitation of current

techniques is the large amount of starting material required, which restricts their applicability to inexpensive, fast-growing cell lines. Here, we show that fastGRO is down-scalable and can be performed with as few as  $0.5\text{--}1 \times 10^6$  cells, potentially extending nascent RNA studies to a variety of model systems.

## RESULTS

### fastGRO Yields Comprehensive Nascent Transcriptome Data in Human Cells

We obtained fastGRO libraries using the suspension cell line THP-1. These widely used leukemic cells are poorly differentiated myeloid progenitors that can respond to inflammatory stimuli (such as lipopolysaccharide [LPS]) similar to monocytes (Bosshart and Heinzelmann, 2016). We processed both unstimulated and LPS-treated THP-1 cells by, first, incubating whole cells in hypotonic solution to cause swelling and subsequent lysis of the plasma membrane (Figure 1A). Next, we used isolated nuclei to perform *in vitro* run-on reactions with the addition of 4-S-UTP in lieu of the brominated or biotinylated ribonucleotides used in GRO-seq and PRO-seq, respectively. Nucleosidetriphosphates (NTPs) containing a reactive thiol group are efficiently incorporated by RNA polymerase as evidenced by techniques such as 4SU-seq and TT-seq (Schwanhäusser et al., 2011; Schwalb et al., 2016) that rely on metabolic labeling starting from thionucleoside analogs. Following isolation by TRIzol, we subjected RNA to mild sonication (Figures 1B and 1C) using a Bioruptor device. This step is necessary to improve the efficiency of the downstream immunopurification and to obtain an even representation of the fully unprocessed mRNA transcript (Figure S1). In fact, we observed that undersonicated RNA resulted in significant loss of resolution at the 3' of most genes (Figure S1). The incorporated 4-S-UTPs are covalently biotinylated using a pyridyldithiol-biotin compound. The reaction forms a reversible disulfide bridge between biotin and the uracil base and allows enrichment of bona fide nascent RNA by affinity purification via streptavidin-conjugated beads. Lastly, affinity-bound molecules are eluted with harsh reducing conditions to cleave off the biotin adduct and recover nascent RNA fragments that will be incorporated into directional (stranded) Illumina-compatible libraries (Figure 1A).

We initially assessed the coverage of fastGRO sequencing by using *de novo* transcript identification with the HOMER suite (Heinz et al., 2010). In this analysis, we retrieved nearly 23,000 newly annotated, independent transcripts in both unstimulated and LPS-stimulated THP-1 cells (Figure 1D). Over 90% of transcripts were expressed at similar levels between both conditions, and 1,951 were upregulated by LPS (as compared to 1,192 that were downregulated). Importantly, we assessed that fastGRO is a highly reproducible technique, since replicated experiments that were independently performed show highly significant correlation (Figure S2). Average read profiles of the 300 most expressed genes (Figure 1E) show a robust signal with seamless coverage along the entire gene body, including the 3' post-termination region (data were normalized using spike-in of *Drosophila* RNA; see STAR Methods for details). Furthermore, fastGRO detected strong nascent transcription at LPS-induced genes as well as LPS-related enhancer and super-enhancer sites (Figures 1F–1H; Figure S3). On average, fastGRO proficiently detects nascent RNA at all RNAPII sites, including antisense promoter transcripts that are known to be rapidly

degraded by the exosome (Figure 1J; Flynn et al., 2011). Lowly abundant lincRNAs were also well represented in our dataset (Figure 1K). These transcripts are highly regulated and contribute to the expression of neighboring genes, as in the case of PVT1 (Figure 1K), which is adjacent to the MYC locus and essential for MYC-driven oncogenesis (Tseng et al., 2014). lincRNAs are conventionally defined as noncoding transcripts longer than 200 bp. Importantly, even shorter noncoding RNAs, either RNAPII or RNAPIII dependent, were robustly detected by fastGRO. For instance, we were able to profile tRNAs, small nucleolar RNAs (snoRNAs), and uridylylate-rich small nuclear RNAs (UsnRNAs) (Figure S4).

Similar to GRO-seq and PRO-seq, fastGRO profiles are a reflection of RNAPII occupancy and incorporate information on polymerase activity, such as the rate of pause-release and elongation. We used our dataset to calculate pausing indexes (the read ratio between the proximal promoter and the remaining gene body) at 300 highly expressed, constitutive genes and at a group (100) of LPS targets (Figure 1L). Our data show no significant changes (with or without LPS) in the control group, while the pausing index of LPS-responsive genes decreases dramatically upon stimulation, suggesting steady accumulation of RNAPII into the gene body.

### **fastGRO Detects Real-Time Kinetics of RNAPII**

To evaluate the sensitivity of fastGRO and its ability to time-resolve the dynamics of RNAPII in human cells, we performed a time course analysis of nascent transcription in THP-1 cells, at both constitutively active and LPS-induced genes (Figure 2A). Briefly, we blocked elongation genome-wide using the CDK9 inhibitor flavopiridol, thereby inducing widespread pausing of RNAPII at proximal promoters. After flavopiridol washout (and simultaneous stimulation with LPS), we collected nuclei at 5, 15, and 30 min and performed fastGRO to gauge the release of RNAPII into the gene body (Figure 2A). Due to much reduced RNA yield of the flavopiridol-treated samples, we increased the starting number of nuclei to 60 million (as opposed to 20 million for all other time points). Upon 2-h treatment of THP-1 with flavopiridol, we observed that nearly all nascent transcription originated from the proximal promoter, suggesting a near-complete elongation block (Figure 2B). We first examined 437 constitutively active genes (non-responsive to LPS). While few RNAPII molecules seemed capable of escaping the flavopiridol blockage, 5 min after washout, RNAPII markedly transitioned into the first third of the gene body. Subsequently, the transcriptional front traveled steadily toward the transcription end site (TES) region over the 15- and 30-min time points (Figures 2B and 2C). Notably, significant coverage after the TES was visible only after 30 min after the flavopiridol washout.

We extended our analysis to LPS genes by subjecting THP-1 cells to LPS stimulation at the washout step. We selected a subgroup of 55 larger (>10 kb) genes that were suitable for analysis of RNAPII kinetics (Figure 2D). Unlike constitutive genes, which showed robust pausing of RNAPII before stimulation, LPS genes piled up much fewer reads at their proximal promoter upon flavopiridol treatment (Figure 2E). Our data showed gradual increase of nascent RNA reads along the gene body over time, as exemplified by low-density lipoprotein receptor (LDLR) (Figure 2E). Analysis of the read distribution of 55 genes, divided by quartiles, showed significant increase only at the first and second quartiles after 5

min of LPS, while the third quartile peaked at 15 min and the fourth quartile rose higher only after 30 min post-stimulation (Figure 2D). Taken together, these data suggest that fastGRO is accurately picking up the real-time dynamics of RNAPII and is suitable to determine RNAPII kinetics under different conditions.

### Unbiased Recovery of Unprocessed Transcripts by fastGRO

Unlike steady-state RNA-seq, nascent RNA-seq captures transcripts before they have been fully processed. Since the vast majority of eukaryotic protein coding genes contain multiple introns, splicing is deemed one of the most frequent and abundant RNA processing events. Therefore, we sought to measure residual splicing events in fastGRO data to probe the actual enrichment of nascent, unprocessed transcripts. We used MAJIQ (Vaquero-Garcia et al., 2016) to determine the relative frequency of splicing junctions, normalized by transcriptome coverage. We stacked up fastGRO of THP1 cells against ribodepleted and poly(A)-selected RNA-seq data that we obtained from the same batch of cells (Figure 3A). As expected, poly(A) RNA-seq data bear the highest fraction of spliced transcripts (with a median >80%) as opposed to ~20% of fastGRO. Ribo-depleted RNA-seq also carries significantly more junctions, albeit slightly lower than poly(A) RNA-seq (as expected, ribodepletion allows minimal retention of non-polyadenylated, unprocessed RNA species). Next, we compared fastGRO to previously published techniques that capture nascent transcription by means of run-on assay (GRO-seq and PRO-seq), metabolic labeling (TT-seq), and isolation of RNAPII/RNA complexes (NET-seq). By taking advantage of commercially available library preparation kits and a single round of biotin enrichment, fastGRO samples can be prepared within 2.5 days, while most other protocols require 5 days of processing time before obtaining sequencing-ready libraries (Figure 3B). We mined public repositories for previously published nascent transcriptomic data of THP-1 cells. We retrieved datasets of GRO-seq, PRO-seq, and TT-seq. First, we compared the relative enrichment of nascent, unspliced transcripts using MAJIQ. Strikingly, fastGRO showed the least contamination of spliced RNA of all techniques (Figure 3C). Long read mNET-seq, however, displayed superior enrichment for unspliced transcripts compared to fastGRO (Figure S5; we compared available long read mNET-seq data in HeLa cells to data from a fastGRO experiment we performed in the same cells). Next, we compared the average read density profile of fastGRO, GRO-seq, PRO-seq, and TT-seq (normalized by sequencing depth). Across a group of highest expressing protein coding genes, fastGRO and TT-seq similarly displayed smooth and continuous density profiles across the entire gene body (Figures 3D and 3E). PRO-seq and GRO-seq profiles appeared more irregular and biased toward the 5' promoter proximal region (Figures 3D and 3E). Since both GRO-seq and PRO-seq protocols comprise multiple size selection steps performed on polyacrylamide gels, they are more likely to introduce a size bias toward smaller RNA fragments.

To ensure that fastGRO is applicable to other cell systems and further the comparison to similar techniques, we generated libraries from HeLa cells. fastGRO showed the most homogeneous coverage across the whole gene body of the top 300 expressed genes (Figures 3F and 3G). PRO-seq and GRO-seq profiles showed more 5' bias and scattered coverage (Figures 3F and 3G). We also analyzed an available standard mNET-seq dataset. The signal was more robust than all other techniques (as per read depth normalization) but heavily

scattered due to the nature of NET-seq technology that favors discovery of polymerase pausing sites. Furthermore, fastGRO data ensured comprehensive coverage of bidirectional enhancer RNAs (Figure 3H), comparable to other techniques.

### fastGRO Maps the Fate of RNA Polymerase Post-termination

We observed increased coverage of 3' regions by fastGRO at several protein coding genes (Figures 3D and 3E). In particular, we noticed a robust pile-up of sequencing reads surrounding the annotated TES. Upon recognition of the polyadenylation site (PAS), the cleavage and polyadenylation machinery is recruited by RNAPII, resulting in the release of a full-length, capped mRNA precursor that will be handed over to poly(A) polymerase (Shi and Manley, 2015; Proudfoot, 2016). However, RNAPII moves further downstream and elongates the post-termination uncapped 3' RNA, which is promptly degraded by the Xrn2 exonuclease. Running after polymerase for several hundreds of nucleotides, Xrn2 eventually prompts RNAPII arrest and unload from its chromatin template (Eaton and West, 2018). Post-termination RNA is rapidly degraded—and, hence, difficult to recover—but provides unique insight into the mechanisms and protein complexes that oversee termination. To fare nascent RNA protocols on their ability to recover 3' RNA, we generated a “termination index” for all highly expressed genes by calculating the ratio of normalized read density after and before the annotated PAS (Figure 4A). We found that fastGRO allowed far more significant recovery of post-termination RNA than PRO-seq and TT-seq (and similar to GRO-seq) (Figure 4A). This was also evident by plotting read-density profiles centered around the TES of the top expressed transcripts (Figure 4B) and by looking at specific genes that present exceptionally extended 3' ends such as FUT4, SFPQ, and HNRNPK (Figures 4C and 4D; Figure S6).

### A Scalable Global Run-On Assay

Nascent RNA techniques require a large amount of starting material, restraining their applicability to easily cultured, inexpensive cell types (Wissink et al., 2019). The recommended starting cell number for optimal GRO-seq and PRO-seq experiments ranges from  $1.5 \times 10^7$  to  $2 \times 10^7$  (Gardini, 2017; Wissink et al., 2019). Similarly, NET-seq is optimized for  $1 \times 10^7$  cells (Mayer and Churchman, 2016), while mNET-seq requires up to  $1.6 \times 10^8$  starting cells due to the additional RNAPII immunoprecipitation (IP) step (Nojima et al., 2016). TT-seq, which is based on metabolic labeling, necessitates 300  $\mu\text{g}$  of total RNA before streptavidin immunopurification, equaling up to  $3 \times 10^7$  starting cells (Schwalb et al., 2016). We initially obtained fastGRO datasets using 1.5 to  $2 \times 10^7$  cells (Figures 1, 2, and 3), which appeared sufficient to yield extensive coverage of nascent transcripts across most genes. In fact, as we increased starting material by 5-fold ( $1 \times 10^8$  cells, similar to mNET-seq) we did not observe increased coverage (Figures S7A and S7B). Conversely, we attempted to reduce input material and noticed that the use of fewer than  $5 \times 10^6$  HeLa cells almost invariably resulted in undetectable amounts of RNA after IP and poor-quality libraries (Figure S7C). We reasoned that boosting the efficiency of thio-UTP biotinylation could improve RNA recovery and the overall quality of sequencing libraries. Therefore, we developed a low-input variant of fastGRO (STAR Methods) by taking advantage of the recently optimized methane thiosulfonate biotin (biotin-MTS) (Duffy et al., 2015). While biotin-MTS is more efficient in forming disulfide bonds with 4-S-UTP, it may also cross-

react with non-thiolated UTP, exposing the entire procedure to contamination of unlabeled, steady-state RNA (Marzi et al., 2016). In fact, we observed increased carryover of processed RNA when using a large number of cells ( $1.5 \times 10^7$ ). We measured spliced RNA content with MAJIQ and found that biotin-MTS samples carry significantly more contamination than samples prepared with biotin-HPDP (Figure S7D). However, the relative contamination of processed RNA was much reduced in small-scale experiments, due to lower RNA concentration in the reaction (Figure S7E). Hence, we adjusted the fastGRO protocol for smaller reaction volumes (we named the low-input variant fastGRO-LI), and we generated, first, a scale-down dataset using  $5 \times 10^6$  HeLa cells. Read density profiles of the top 200 genes and the top induced epidermal growth factor (EGF) genes showed continuous coverage across the gene body, without loss of resolution as compared to the  $2 \times 10^7$  cell dataset (Figures 5A and 5B). Next, we set up an extended downscale experiment using 2.5, 1, and  $0.5 \times 10^6$  cells (8- to 40-fold less than the original experiment). We gauged the fraction of high-input annotated transcripts that were still detected in the low-input samples. Protein-coding genes that were undetectable equaled 10% in the  $2.5 \times 10^6$ - and  $1 \times 10^6$ -cell experiments and up to 14% in the  $0.5 \times 10^6$ -cell sample (Figure 5C). The vast majority of transcripts were reliably detected by fastGRO-LI, and read density profiles were remarkably consistent with those of high-input fastGRO experiments, with loss of resolution at the post-termination RNA (Figures 5D and 5E).

In an effort to extend the portability of our technique to primary-like cells or tissues, we performed another set of scale-down experiments in neural progenitor cells (NPCs), obtained from a human induced pluripotent stem cell (iPSC) line. iPSCs are slow-growing, non-transformed, human pluripotent cells (with a diploid DNA content) that can be differentiated into a variety of human tissues and cell types. First, we generated NPCs (Figure S7F) over the course of 13 days, treating cells with PSC Neural Induction Medium for 7 days and then expanding them using GIBCO Neural Expansion Medium for 6 days. Next, we performed fastGRO in NPCs using the standard protocol (20 million cells), and we identified a group of about 300 highly expressed genes, many of which were neuronal specific. Lastly, we performed fastGRO-LI using 5 million and 1 million cells and observed no loss of resolution, except for the post-termination signal (as observed and discussed in the case of HeLa cells) (Figures 5E and 5F). Elevated reproducibility (measured as Spearman correlation) between samples of different input sizes suggests that low-input fastGRO protocols can be applied to a broad range of primary samples/tissues (Figure S7G). Taken together, we demonstrate the feasibility of fastGRO with less than a million cells, potentially extending nascent transcriptome analysis to a wider pool of model systems and experimental conditions.

## DISCUSSION

We developed a fast protocol to generate genomic libraries of nascent RNA, based on a NRO assay. During the run-on reaction, newly synthesized RNA incorporates the ribonucleotide analog 4-S-UTP (Figure 6). Sulfhydryl-reactive biotin is then covalently bound to UTP analogs, allowing the affinity-based purification of fragmented nascent transcripts. After elution with a harsh reducing buffer, RNA is subjected to directional library preparation as per Illumina guidelines (Figure 6). fastGRO generates comprehensive and seamless coverage



of RNAPII (and RNAPIII) activity for the most abundant RNA species (highly expressed protein coding genes, and small noncoding RNAs) as well as those harder to detect (antisense promoter transcripts, eRNAs, and lincRNAs).

We show that fastGRO offers practical advantages over similar techniques, such as GRO-seq and PRO-seq, that are frequently used to gauge the dynamics and processivity of RNAPII: the overall sample processing time is reduced by more than half, and custom library preparation is replaced by Illumina-compatible prep kits that are streamlined and generate more reproducible libraries (cutting off the user-dependent size selection steps). Therefore, fastGRO becomes particularly suitable for conducting kinetic studies of RNAPII that carefully assess elongation dynamics, as we demonstrated by releasing transcription after pharmacologically induced pausing (Figure 2). Furthermore, fastGRO yields significantly better enrichment of nascent, genuinely unprocessed RNA. In fact, contamination of spliced RNA in fastGRO is the lowest among most nascent RNA-seq types, including run-on-based techniques as well as metabolic labeling methods (i.e., TT-seq). While mNET-seq carries lower contamination of processed RNA than fastGRO, the applicability is limited by its onerous requirement of starting material. The propensity of fastGRO to enrich for unprocessed, short-lived transcripts is also visible at the post-termination site of protein-coding genes. After encountering the first polyadenylation signal, RNAPII continues its course for hundreds to thousands of kilobases until Xrn2-dependent termination effectively dislodges the enzyme off of its DNA template. We calculated a termination index for a set of highly expressed genes to demonstrate that fastGRO provides better coverage of RNAPII activity past the PAS.

Last, we established that fastGRO can be performed with lower amounts of input material. A major limitation of all current methods of nascent RNA sequencing is the requirement of 10 to 30 million cells per single library. We performed a serial downscale of fastGRO using a modified protocol (fastGRO-LI) that uses a highly reactive biotin conjugate (biotin-MTS). Our results suggest that fastGRO can be performed with as few as  $0.5 \times 10^6$  cells (30 to 60 times fewer than GRO-seq, PRO-seq, or TT-seq), at the expense of less than 15% of the transcriptome. While HeLa cells are easy to manipulate, we further demonstrate that fastGRO-LI can be applied to iPSC-derived NPCs (with as low as  $1 \times 10^6$  cells). Our technique brings about, for the first time, portability of nascent RNA technology to a wider range of model systems, including primary human and mouse cells.

Nascent RNA-seq is a method of choice when dissecting RNAPII regulatory steps (such as pausing, elongation, and termination). Additionally, it provides the most accurate quantitative data on gene regulation since it reflects the real-time activity of polymerase. In addition to protein-coding genes, nascent RNA-seq methods are also used to measure RNA species that are undetectable or poorly represented in canonical RNA-seq datasets. There has been a recent research focus on unstable and/or low-abundant RNA categories, which has propelled RNA biology and impressed a new mark on the fields of epigenetics and transcription regulation. For instance, the low-expressed and poorly evolutionarily constrained lincRNAs were found to regulate a variety of biological processes by associating with either repressing or activating chromatin complexes (Gardini and Shiekhhattar, 2015; Ransohoff et al., 2018). Additionally, eRNAs have been used to gauge enhancer activity

throughout development and were shown to directly impact gene expression (Andersson et al., 2014; Lai et al., 2015; Bose et al., 2017). Finally, antisense promoter transcripts are rapidly degraded but offer an invaluable insight into the architecture of eukaryotic promoters and the mechanisms of productive elongation and early termination (Almada et al., 2013; Andersson et al., 2015; Jin et al., 2017). Nascent RNA methods provide existential support to these lines of research, and fastGRO represents a standardized, user-friendly, and scalable technique that can be integrated into several experimental settings.

## STAR★METHODS

### RESOURCE AVAILABILITY

**Lead contact**—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Alessandro Gardini (agardini@Wistar.org).

**Materials availability**—This study did not generate new unique reagents.

**Data and code availability**—Original high-throughput sequencing data are deposited at the Gene Expression Omnibus with the accession number GSE143844.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

**Cell lines**—Human THP-1 cells were obtained from American Type Culture Collection (ATCC) and maintained in Roswell Park Memorial Institute (RPMI)-1640 medium (Corning) supplemented with 10% (v/v) of super calf serum (GEMcell) and 2 mM of L-glutamine (Corning) at 37°C and 5% CO<sub>2</sub>. Human HeLa cells were obtained from ATCC and maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% super calf serum (GEMcell) and 2 mM L-glutamine (Corning) at 37°C and 5% CO<sub>2</sub>. Male SV20 induced pluripotent stem cells were obtained from the University of Pennsylvania Human Pluripotent Stem Cell Core and differentiated over 7 days in PSC Neural Induction Medium (GIBCO) and expanded for 6 days in Neural Expansion Medium (GIBCO) at 37°C and 5% CO<sub>2</sub>. THP1 cells were treated with 2 μM flavopiridol (Sigma) for 2 h or 5 μg/ml of LPS (Invitrogen) for 5, 15, and 30 min or 4 h in growing medium. HeLa were treated with 100 ng/ml of rEGF (Invitrogen) for 15 min in growing medium.

### METHOD DETAILS

**Experiments performed**—A detailed table of experiments with cell number, cell line, treatment, and replicate number is available in Table S1.

**fastGRO**—(A step-by-step protocol is publicly available at <https://doi.org/10.17504/protocols.io.bbmgek3w>) 20–5 million of cells were washed twice with ice-cold PBS before adding swelling buffer (10 mM Tris-HCL pH 7.5, 2mM MgCl<sub>2</sub>, 3 mM CaCl<sub>2</sub>, 2U/ml Superase-in (Invitrogen)). Cells were swelled for 5 min on ice, washed with swelling buffer + 10% glycerol and then lysed in lysis buffer (10 mM Tris-HCL pH 7.5, 2 mM MgCl<sub>2</sub>, 3 mM CaCl<sub>2</sub>, 10% glycerol, 1% Igepal (NP-40), 2 U/ml Superase-in) to isolate nuclei. Nuclei were washed twice with lysis buffer and resuspended in freezing buffer (40% glycerol, 5

mM MgCl<sub>2</sub>, 0.1 mM 0.5M EDTA, 50 mM Tris-HCL pH 8.3) to a concentration of  $2 \times 10^7$  nuclei per 100  $\mu$ L. Nuclei were then frozen in dry ice and stored at  $-80^\circ\text{C}$ . Nuclei were thawed on ice and spike in nuclei were added if used. An equal volume of pre-warmed nuclear run-on reaction buffer (10 mM Tris-HCl pH 8, 5 mM MgCl<sub>2</sub>, 300 mM KCl, 1 mM DTT, 500  $\mu$ M ATP, 500  $\mu$ M GTP, 500  $\mu$ M 4-thio-UTP, 2  $\mu$ M CTP, 200 U/ml Superase-in, 1% Sarkosyl (N-Laurylsarcosine sodium salt solution) was added and incubated for 7 min at  $30^\circ\text{C}$  for the nuclear run-on. Nuclear run-on RNA was extracted with TRIzol LS reagent (Invitrogen) following the manufacturer's instructions and ethanol precipitated. NRO-RNA was resuspended in water and concentration was determined with Qubit High Sensitivity Assay kit (Invitrogen). Up to 150  $\mu$ g of RNA was transfer to a new tube and 5%–10% of spike-in RNA was added if spike in nuclei were not added prior to the Nuclear run on. RNA was then fragmented with a Bioruptor UCD-200 for 1–5 cycles of 30 s ON / 30 s OFF, high settings. Fragmentation efficiency was analyzed by running fragmented and unfragmented RNA on Agilent 2200 TapeStation using High Sensitivity RNA ScreenTapes following manufacturer's instructions. Fragmented RNA was incubated in Biotinylation Solution (20 mM Tris pH 7.5, 2 mM EDTA pH 8.0, 40% dimethylformamide, 200  $\mu$ g/ml EZ-link HPDP Biotin (Thermo Scientific)) for 2h in the dark at  $25^\circ\text{C}$ , 800 rpm. After ethanol precipitation, the biotinylated RNA was resuspended in water and biotinylated-RNA was separated with M280 Streptavidin Dynabeads (Invitrogen). 100  $\mu$ l/sample of beads were washed twice with 2 volumes of freshly prepared wash buffer (100 mM Tris pH 7.5, 10 mM EDTA pH 8.0, 1M NaCl, 0.1% (v/v) Tween-20) and resuspended in 1 volume of wash buffer and added to the biotinylated-RNA. After 15 min in rotation at  $4^\circ\text{C}$ , beads were washed three times with wash buffer pre-warmed at  $65^\circ\text{C}$  and three times with room temperature wash buffer. 4-S-UTP containing RNA was eluted in 100 mM DTT buffer and purified with RNA Clean and Purification kit (Zymo Research) with in-column DNaseq reaction to eliminate traces of genomic DNA. The eluted RNA was quantified with Qubit High Sensitivity Assay kit (Invitrogen) and used to produce barcoded RNA sequencing libraries using the NEBNext Ultra II Directional RNA Library Prep kit (New England Biolabs). Libraries were sequenced on Illumina NextSeq 500.

**Low-input fastGRO (fastGRO-LI)**—(A step-by-step protocol is publicly available at <https://doi.org/10.17504/protocols.io.bkdtks6n>) 5–0.5 million nuclei were extracted as described for fastGRO and resuspended in freezing buffer (40% glycerol, 5 mM MgCl<sub>2</sub>, 0.1 mM 0.5M EDTA, 50 mM Tris-HCL pH 8.3) to a concentration of up to  $5 \times 10^6$  nuclei per 25  $\mu$ L. Nuclei were then frozen in dry ice and stored at  $-80^\circ\text{C}$ . Run-on reaction was performed as described for fastGRO, NRO-RNA was resuspended in water and concentration was determined with Qubit High Sensitivity Assay kit (Invitrogen). Up to 30  $\mu$ g of RNA was transfer to a new tube and 5%–10% of spike-in RNA was added, if spike in nuclei were not added prior to the Nuclear run on. RNA was then fragmented with a Bioruptor UCD-200 for 30 s, low settings. Fragmentation efficiency was analyzed by running fragmented and unfragmented RNA on Agilent 2200 TapeStation using High Sensitivity RNA ScreenTapes following manufacturer's instructions. Fragmented RNA was incubated in low-input Biotinylation Solution (25 mM HEPES pH 7.5, 1 mM EDTA, 25% dimethylformamide, 16.4  $\mu$ M MTS-Biotin (Biotium)) for 30 minutes in the dark at  $25^\circ\text{C}$ , 800 rpm. After ethanol precipitation, the biotinylated RNA was resuspended in water and

DNase treatment was performed with TURBO DNase (Invitrogen) following manufacturer instructions. Biotinylated-RNA was separated with M280 Streptavidin Dynabeads (Invitrogen): 25  $\mu$ l/sample of beads were washed twice with 2 volumes of freshly prepared wash buffer (100 mM Tris pH 7.5, 10 mM EDTA pH 8.0, 1M NaCl, 0.1% (v/v) Tween-20) and resuspended in 1 volume of wash buffer and added to the biotinylated-RNA. After 15 minutes in rotation at 4°C, beads were washed three times with wash buffer pre-warmed at 65°C and three times with room temperature wash buffer. thio-UTP containing RNA was eluted in 100 mM DTT buffer, ethanol purified used to produce barcoded RNA sequencing libraries using the NEBNext Ultra II Directional RNA Library Prep kit (New England Biolabs). Libraries were sequenced on Illumina NextSeq 500.

**Spike-in RNA preparation**—*Drosophila* S9 cells were incubated for 5 minutes with 50mM of 4-thiouridine (4sU) at room temperature. Cells were then washed twice with 1X PBS, lysed in Trizol reagent. RNA was extracted with Direct-zol Mini prep kit (Zymo research). Aliquots of 2  $\mu$ g were prepared, snap-frozen in liquid nitrogen and store at –80°C.

**Spike-in nuclei preparation**—Nuclei from *Drosophila* S2 cells were extracted as described for fastGRO. One million nuclei aliquots were prepared.

**Real-time quantitative polymerase chain reaction**—iPSCs and NPCs were lysed in Tri-reagent and RNA was extracted using the Direct-zol RNA MiniPrep kit (Zymo research). 900 ng of template RNA was retrotranscribed into cDNA using random primers and the Revertaid first strand cDNA synthesis kit (Thermo Scientific) according to manufacturer directions. 50 ng of the cDNA were used for each real-time quantitative PCR reaction with 0.4 mM of each primer, 10  $\mu$ L of iQ SYBR Green Supermix (BioRAD) in a final volume of 20  $\mu$ l, using a CFX96 real-time system (BioRAD). Thermal cycling parameters were: 3 min at 95°C, followed by 40 cycles of 10 s at 95°C, 20 s at 63°C followed by 30 s at 72°C. Each sample was run in triplicate. GAPDH was used as normalizer. Primer sequences are reported in Table S2.

**PolyA RNA-seq and ribodepleted RNA-seq**—Total RNA was extracted using Direct-zol RNA Miniprep kit (Zymo Research). For polyA RNA-seq, the polyA fraction was isolated by running RNA samples through the Oligo(dT) Dynabeads (Invitrogen). For ribodepleted RNA-seq, ribosomal RNA was removed by the KAPA RNA HyperPrep Kit (Roche). The resulting RNA was subjected to strand-specific library preparation using the SENSE mRNA-Seq Library Prep Kit V2 (Lexogen). Sequencing was performed on Nextseq500 (Illumina).

## QUANTIFICATION AND STATISTICAL ANALYSIS

**Analysis of RNA-seq data**—Reads were aligned to hg19 using STAR v2.5 (Dobin et al., 2013), in 2-pass mode with the following parameters:–quantMode TranscriptomeSAM–outFilterMultimapNmax 10–outFilterMismatchNmax 10–outFilterMismatchNoverLmax 0.3–alignIntronMin 21–alignIntronMax 0–alignMatesGapMax 0–alignSJoverhangMin 5–runThreadN 12–twopassMode Basic–twopass1readsN 60000000–sjdbOverhang 100. The latest annotations obtained from Ensembl were used to build reference indexes for the STAR

alignment. Bam files were filtered based on alignment quality ( $q = 10$ ) using Samtools v0.1.19 (Li et al., 2009). Bam files were then normalized based on the number of reads of spike-in/total read number with Samtools and bigwig files were built with deeptools 3.3.1 (Ramírez et al., 2016).

For nascent RNA analysis, bam files were transformed in bed file with bedtools (bamtoBED option) and subjected to analysis with HOMER v4.11 (Heinz et al., 2010). For the identification of new transcripts, findPeaks.pl was used to analyze fastGRO data with the following parameters: -style groseq -tssFold 6 -bodyFold 5 -pseudoCount 0.5 -minBodySize 500 -maxBodySize 100000. To analyze gene expression, FPKM (Fragments Per Kilobase of exon per Million fragments mapped) was calculated with HOMER using analyzeRepeats.pl (parameters: rna -count genes -strand - -rpkm -condenseGenes) and addGeneAnnotation.pl. FPKM were used to analyze differential gene expression levels, normalized by feature length with DESeq2 (Love et al., 2014).

**Published genome-wide data and analysis**—Published data were downloaded and re-analyzed as described for nascent RNA for GRO-seq, PRO-seq, TT-seq and mNET-seq. H3K27ac ChIPseq data were aligned to hg19, using Burrows Wheeler Alignment tool (BWA) (Li and Durbin, 2010), with the MEM algorithm. Aligned reads were filtered based on mapping quality (MAPQ > 10) to restrict our analysis to higher quality and likely uniquely mapped reads, and PCR duplicates were removed.

**Average density analysis, pausing index, and termination index analysis**—fastGRO, GRO-seq, PRO-seq, TT-seq, mNET-seq and RNA-seq data were subjected to read density analysis after spike-in (for fastGRO) and sequencing depth normalization. seqMINER 1.3.3 was used to extract read densities, and mean density profiles were then generated in R 3.5.3 using ggplot2 (Villanueva and Chen, 2019). For pausing index analysis, the ratio between read counts at the TSS (−50/+150 bp) and read counts across the rest of the gene body (+150/termination end site) was calculated. For termination index analysis, the ratio between read counts post the termination end site (TES, TES / +3 kb) and read counts pre-TES (−0.5 kb / TES) were calculated. Statistical robustness was calculated with Wilcoxon rank-sum tests.

**Splice junction analysis**—Reads were trimmed to the average length of the reads in the dataset with the shortest reads in each given comparison. Only one of two reads from paired-end data were used in cases involving comparisons between single- and paired-end datasets. Reads were aligned to human genome assembly GRCh38 using STAR (version 2.5.4b) (Dobin et al., 2013). Splice junctions were identified and quantified using MAJIQ (Vaquero-Garcia et al., 2016) requiring a minimum number of reads on average in intronic sites (−min-intronic-cov) of 0.005 and the number of intronic bins with some coverage (−irnbins) of 0.1. Only junctions common to all samples in any given comparison were used in the analyses. Statistical tests performed were Wilcoxon rank-sum tests.

**Additional Resources**—Step-by-step protocol for fastGRO is publicly available <https://doi.org/10.17504/protocols.io.bbm3k3w>. Step-by-step protocol for fastGROli is publicly available <https://doi.org/10.17504/protocols.io.bkdtk6n>

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (NIH; R01 HL141326 to A.G.), the American Cancer Society (RSG-18-157-01 to A.G.), the G. Harold and Leila Y. Mathers Foundation (MF-1809-00172 to A.G.), and the Emerson Collective. A.Z. is supported by NIH training grant T32-CA009171. All genome-wide data were generated and processed at the Wistar Institute with outstanding technical support from all members of the Genomics Facility, particularly Dr. Shashi Bala (NIH-NCI funded CCSG: P30-CA010815).

## REFERENCES

- Adelman K, and Lis JT (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet* 13, 720–731. [PubMed: 22986266]
- Almada AE, Wu X, Kriz AJ, Burge CB, and Sharp PA (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499, 360–363. [PubMed: 23792564]
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461. [PubMed: 24670763]
- Andersson R, Chen Y, Core L, Lis JT, Sandelin A, and Jensen TH (2015). Human Gene Promoters Are Intrinsically Bidirectional. *Mol. Cell* 60, 346–347. [PubMed: 26545074]
- Bose DA, Donahue G, Reinberg D, Shiekhhattar R, Bonasio R, and Berger SL (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell* 168, 135–149.e22. [PubMed: 28086087]
- Bosshart H, and Heinzelmann M (2016). THP-1 cells as a model for human monocytes. *Ann. Transl. Med* 4, 438. [PubMed: 27942529]
- Bouvy-Liivrand M, Hernández de Sande A, Pölonen P, Mehtonen J, Vuorenmaa T, Niskanen H, Sinkkonen L, Kaikkonen MU, and Heinäniemi M (2017). Analysis of primary microRNA loci from nascent transcriptomes reveals regulatory domains governed by chromatin architecture. *Nucleic Acids Res* 45, 12054. [PubMed: 29036333]
- Core LJ, Waterfall JJ, and Lis JT (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848. [PubMed: 19056941]
- Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, and Lis JT (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet* 46, 1311–1320. [PubMed: 25383968]
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. [PubMed: 23104886]
- Duffy EE, Rutenberg-Schoenberg M, Stark CD, Kitchen RR, Gerstein MB, and Simon MD (2015). Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. *Mol. Cell* 59, 858–866. [PubMed: 26340425]
- Eaton JD, and West S (2018). An end in sight? Xrn2 and transcriptional termination by RNA polymerase II. *Transcription* 9, 321–326. [PubMed: 30035655]
- Flynn RA, Almada AE, Zamudio JR, and Sharp PA (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl. Acad. Sci. USA* 108, 10460–10465. [PubMed: 21670248]
- Gardini A (2017). Global Run-On Sequencing (GRO-Seq). *Methods Mol. Biol.* 1468, 111–120. [PubMed: 27662873]
- Gardini A, and Shiekhhattar R (2015). The many faces of long noncoding RNAs. *FEBS J* 282, 1647–1657. [PubMed: 25303371]
- Gariglio P, Buss J, and Green MH (1974). Sarkosyl activation of RNA polymerase activity in mitotic mouse cells. *FEBS Lett* 44, 330–333. [PubMed: 4472294]

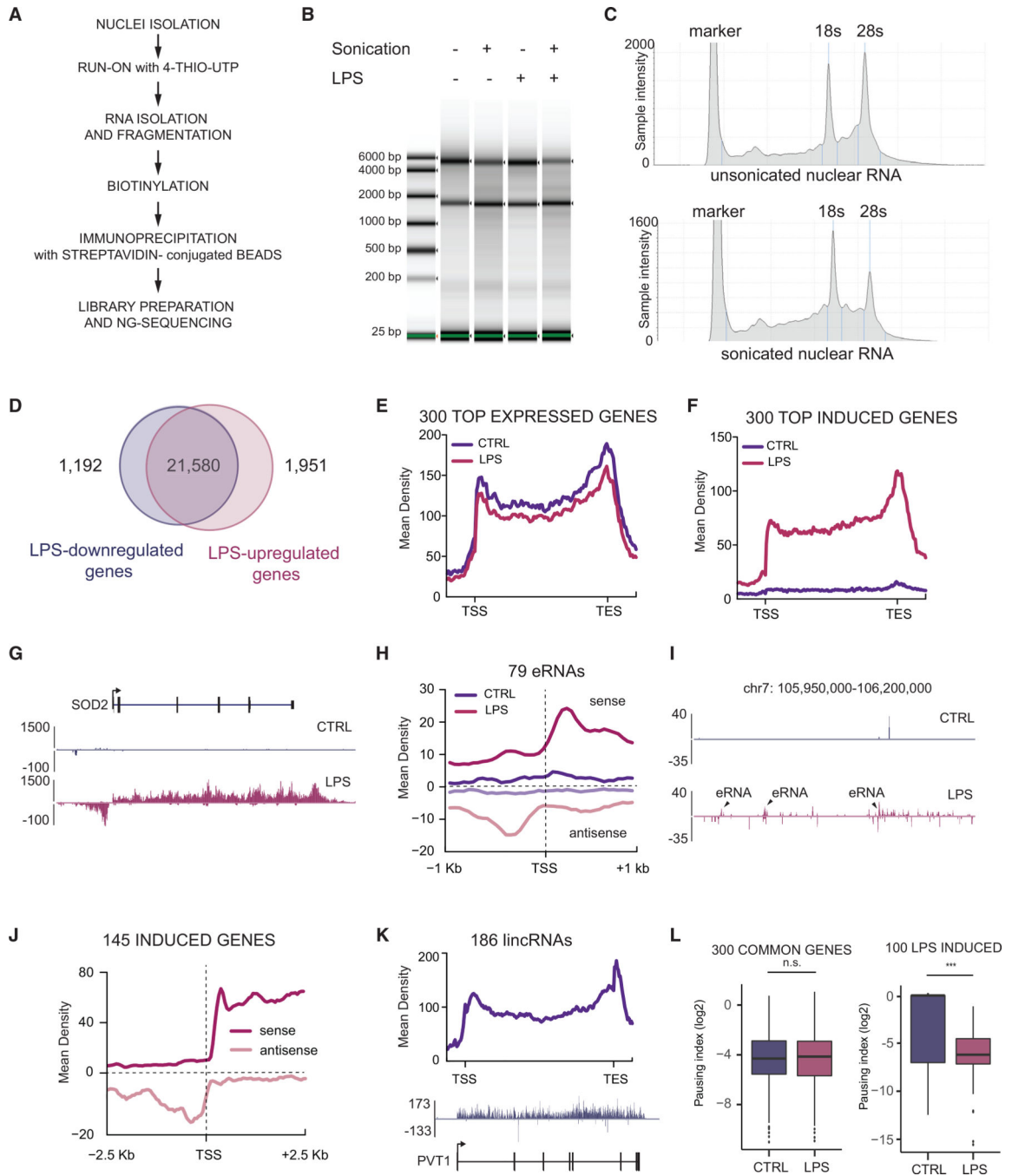
- Gariglio P, Bellard M, and Chambon P (1981). Clustering of RNA polymerase B molecules in the 5' moiety of the adult beta-globin gene of hen erythrocytes. *Nucleic Acids Res* 9, 2589–2598. [PubMed: 6269056]
- Godfrey L, Crump NT, Thorne R, Lau IJ, Repapi E, Dimou D, Smith AL, Harman JR, Telenius JM, Oudelaar AM, et al. (2019). DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nat. Commun* 10, 2803. [PubMed: 31243293]
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589. [PubMed: 20513432]
- Jin Y, Eser U, Struhl K, and Churchman LS (2017). The Ground State and Evolution of Promoter Region Directionality. *Cell* 170, 889–898.e10. [PubMed: 28803729]
- Kwak H, and Lis JT (2013). Control of transcriptional elongation. *Annu. Rev. Genet* 47, 483–508. [PubMed: 24050178]
- Lai F, and Shiekhattar R (2014). Enhancer RNAs: the new molecules of transcription. *Curr. Opin. Genet. Dev* 25, 38–42. [PubMed: 24480293]
- Lai F, Gardini A, Zhang A, and Shiekhattar R (2015). Integrator mediates the biogenesis of enhancer RNAs. *Nature* 525, 399–403. [PubMed: 26308897]
- Li H, and Durbin R (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. [PubMed: 20080505]
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. [PubMed: 19505943]
- Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: arXiv:1303.3997, <https://arxiv.org/abs/1303.3997>.
- Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. [PubMed: 25516281]
- Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, and Lis JT (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc* 11, 1455–1476. [PubMed: 27442863]
- Marzi MJ, Ghini F, Cerruti B, de Pretis S, Bonetti P, Giacomelli C, Gorski MM, Kress T, Pelizzola M, Muller H, et al. (2016). Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. *Genome Res* 26, 554–565. [PubMed: 26821571]
- Mayer A, and Churchman LS (2016). Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. *Nat. Protoc* 11, 813–833. [PubMed: 27010758]
- Mayer A, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, Sandstrom R, Stamatoyannopoulos JA, and Churchman LS (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* 161, 541–554. [PubMed: 25910208]
- Nilson KA, Lawson CK, Mullen NJ, Ball CB, Spector BM, Meier JL, and Price DH (2017). Oxidative stress rapidly stabilizes promoter-proximal paused Pol II across the human genome. *Nucleic Acids Res* 45, 11088–11105. [PubMed: 28977633]
- Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, and Proudfoot NJ (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* 161, 526–540. [PubMed: 25910207]
- Nojima T, Gomes T, Carmo-Fonseca M, and Proudfoot NJ (2016). Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat. Protoc* 11, 413–428. [PubMed: 26844429]
- Nojima T, Rebelo K, Gomes T, Grosso AR, Proudfoot NJ, and Carmo-Fonseca M (2018). RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol. Cell* 72, 369–379.e4. [PubMed: 30340024]
- Pai AA, and Luca F (2019). Environmental influences on RNA processing: Biochemical, molecular and genetic regulators of cellular response. *Wiley Interdiscip. Rev. RNA* 10, e1503. [PubMed: 30216698]

- Phanstiel DH, Van Bortle K, Spacek D, Hess GT, Shamim MS, Machol I, Love MI, Aiden EL, Bassik MC, and Snyder MP (2017). Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Mol. Cell* 67, 1037–1048.e6. [PubMed: 28890333]
- Proudfoot NJ (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* 352, aad9926. [PubMed: 27284201]
- Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, and Manke T (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44, W160–W165. [PubMed: 27079975]
- Ransohoff JD, Wei Y, and Khavari PA (2018). The functions and unique features of long intergenic non-coding RNA. *Nat. Rev. Mol. Cell Biol* 19, 143–157. [PubMed: 29138516]
- Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, and Proudfoot NJ (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol. Cell* 65, 25–38. [PubMed: 28017589]
- Schoenberg DR, and Maquat LE (2012). Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet* 13, 246–259. [PubMed: 22392217]
- Schwalb B, Michel M, Zacher B, Frühauf K, Demel C, Tresch A, Gagneur J, and Cramer P (2016). TT-seq maps the human transient transcriptome. *Science* 352, 1225–1228. [PubMed: 27257258]
- Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, and Selbach M (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. [PubMed: 21593866]
- Shi Y, and Manley JL (2015). The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev* 29, 889–897. [PubMed: 25934501]
- Stark R, Grzelak M, and Hadfield J (2019). RNA sequencing: the teenage years. *Nat. Rev. Genet* 20, 631–656. [PubMed: 31341269]
- Tseng YY, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H, Ronning P, Reuland B, Guenther K, Beadnell TC, et al. (2014). PVT1 dependence in cancer with MYC copy-number increase. *Nature* 512, 82–86. [PubMed: 25043044]
- Vaquero-Garcia J, Barrera A, Gazzara MR, González-Vallinas J, Lahens NF, Hogenesch JB, Lynch KW, and Barash Y (2016). A new view of transcriptome complexity and regulation through the lens of local splicing variations. *eLife* 5, e11752. [PubMed: 26829591]
- Villanueva RAM, and Chen ZJ (2019). ggplot2: Elegant Graphics for Data Analysis, 2nd edition. *Measurement: Interdisciplinary Research and Perspectives* 17, 160–167.
- Wissink EM, Vihervaara A, Tippens ND, and Lis JT (2019). Nascent RNA analyses: tracking transcription and its regulation. *Nat. Rev. Genet* 20, 705–723. [PubMed: 31399713]
- Yamada T, and Akimitsu N (2019). Contributions of regulated transcription and mRNA decay to the dynamics of gene expression. *Wiley Interdiscip. Rev. RNA* 10, e1508. [PubMed: 30276972]
- Zhan Xiaowei, and Liu Dajiang (2015). SEQMINER: An R-Package to Facilitate the Functional Interpretation of Sequence-Based Associations. *Genetic Epidemiology*. 10.1002/gepi.21918.



### Highlights

- fastGRO is a high-throughput nuclear run-on protocol to profile nascent transcriptomes
- Streamlined sample preparation yields sequencing-ready libraries in less than 3 days
- Enrichment of unprocessed RNA sheds light on post-termination fate of RNAPII
- Protocol scalability allows for the analysis of primary-like cells and tissues

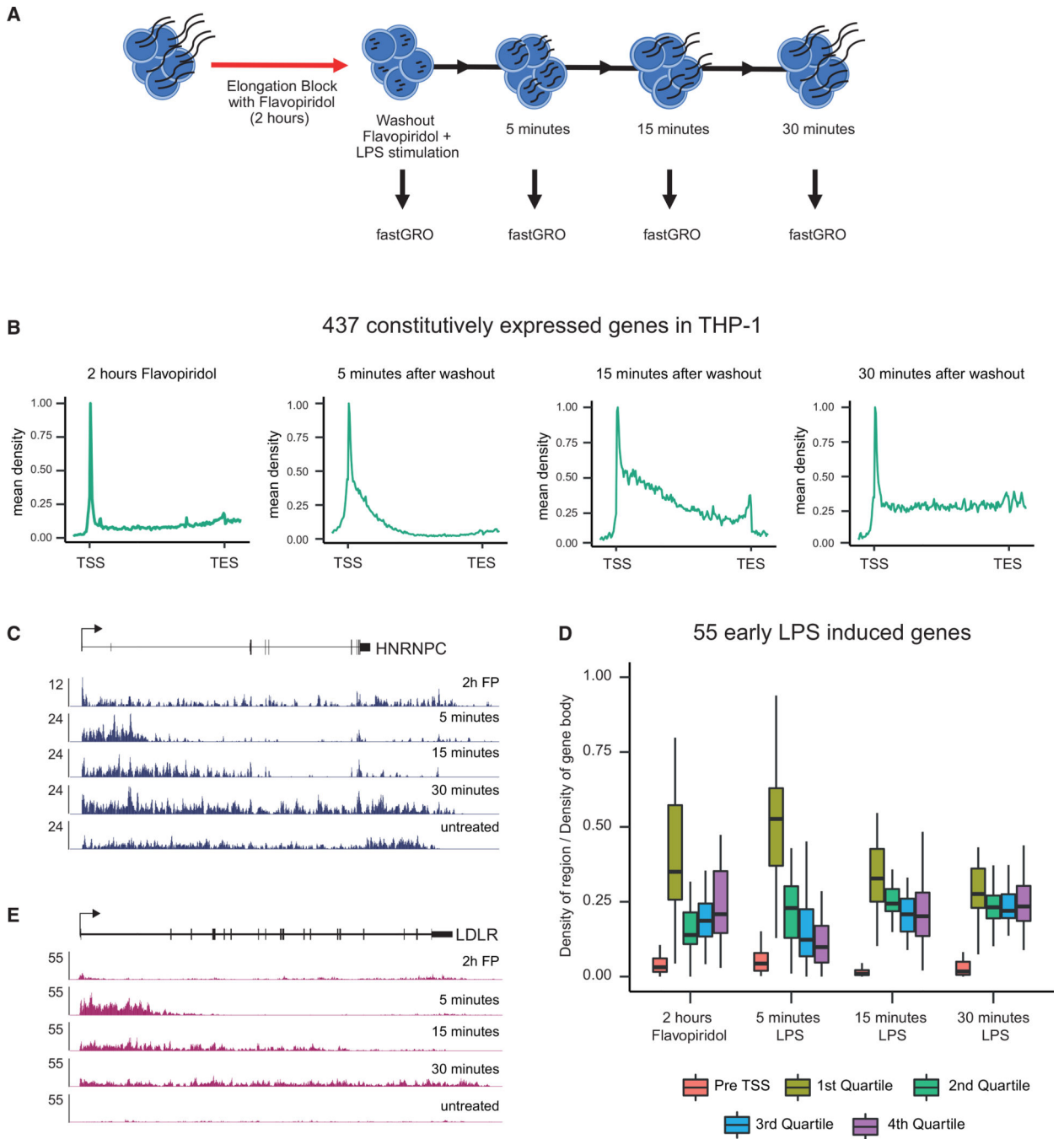


**Figure 1. fastGRO Generates Global Nascent Transcriptome Data**

(A) Schematic of fastGRO procedure. Nuclei are first isolated, and nuclear run-on (NRO) is performed *in vitro* in the presence of 4-thio-UTP. NRO RNA is isolated and fragmented, biotinylated, recovered using streptavidin-conjugated beads, and processed for library preparation. NG, next-generation.

(B and C) Examples of TapeStation run showing mild fragmentation of NRO RNA extracted from LPS-treated and untreated human THP1 cells.

- (D) Two replicates of control (CTRL; purple) and LPS-treated THP1 (dark pink) fastGRO samples were analyzed by HOMER to identify common and LPS-induced transcripts.
- (E) Average density profiles of fastGRO signals for CTRL and LPS-induced THP1 at 300 most expressed genes. TSS, Transcription Start Site, TES, Transcription End Site.
- (F) Average density profiles of fastGRO reads for CTRL and LPS-induced THP1 at 300 most LPS-induced genes.
- (G) Screenshot of region surrounding the LPS-induced gene SOD2 showing fastGRO reads along gene body, post-TES, and promoter antisense.
- (H) Average density profiles of sense and antisense fastGRO reads at 79 putative enhancer regions, identified by the level of H3K27ac (see Figure S3).
- (I) Screenshots of LPS-induced enhancer RNAs. chr7, chromosome 7.
- (J) Average density profile of sense and antisense fastGRO reads at the transcription start site (TSS) region of 145 LPS-induced genes.
- (K) Average profile of fastGRO reads from untreated THP1 at 186 long intergenic non-coding RNAs (lincRNAs) and screenshot of the PVT1 lincRNA in untreated THP1 as depicted by fastGRO.
- (L) Pausing index was calculated from fastGRO reads for 300 highly expressed genes and 300 LPS-induced genes, showing how fastGRO is a useful approach to study RNAPII elongation. Highly expressed and LPS-induced genes were identified from two replicates of CTRL and LPS using HOMER. Replicate 1 was used to generate profiles and screenshots. Correlation between the two replicates of CTRL and LPS is reported in Figure S2. \*\*\* p-value <.001; n.s., not significant.



**Figure 2. Profiling RNAPII Kinetics Using fastGRO**

(A) Diagram of the experimental design. THP1 cells were treated with 2  $\mu$ M CDK9 inhibitor flavopiridol for 2 h to block transcription elongation. To release transcription, cells were replated in fresh media to wash out flavopiridol, with the addition of LPS to further stimulate inflammatory genes. Samples for fastGRO analysis were collected at 0, 5, 15, and 30 min after washout.

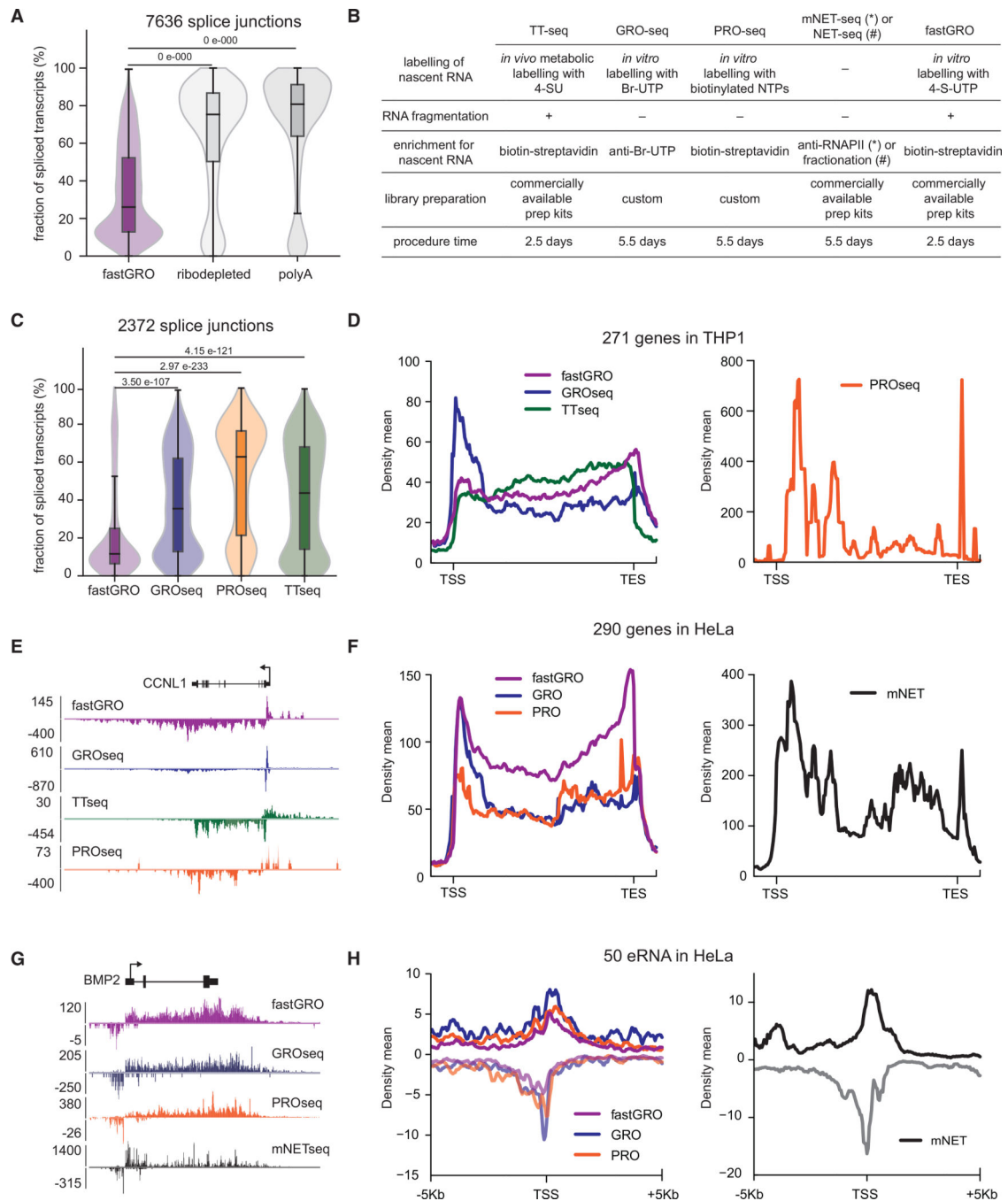
(B) Average profiles of fastGRO at 473 highly expressed genes in THP1 cells (>10 kb) reflect the transcriptional front of RNAPII moving progressively past the proximal promoter

over the course of 30 min after release of the elongation block. Profiles were normalized to their TSS.

(C) Screenshot of the constitutively expressed HNRNPC gene whose expression is fully recovered 30 min after washout of flavopiridol. As a comparison, data of asynchronously growing THP-1 cells (untreated) are provided. FP, Flavopiridol

(D) Boxplot analysis of read density at inflammatory (LPS-induced) genes. Normalized read density of fastGRO was calculated over gene quartiles (as well as 5 kb “Pre-TSS” as a control). Time-dependent increases of the 2nd, 3rd, and 4th quartiles indicate a wave of transcription migrating through the gene body after flavopiridol washout. The analysis was performed on 55 LPS-induced genes over 10 kb in length.

(E) Screenshot of the early LPS-induced LDLR gene showing the wave of transcription during the first 30 min after LPS stimulation.



**Figure 3. fastGRO Recovers Nascent, Unprocessed, and Short-Lived Transcripts**

(A) Splice junction analysis by MAJIQ shows the substantial recovery of processed RNA by rRNA-depleted (gray) and poly(A)-enriched (dark gray) RNA-seq. fastGRO (purple) is significantly enriched for nascent, unspliced RNA.

(B) Comparison of fastGRO to other nascent RNA techniques. An advantage of fastGRO is the overall short processing time (2.5 days, using commercially available library prep kits).

(C) fastGRO (purple) shows lower enrichment of spliced junctions than comparable nascent RNA-seq techniques such as GRO-seq (blue), PRO-seq (orange), and TT-seq (green) in THP1 cells.

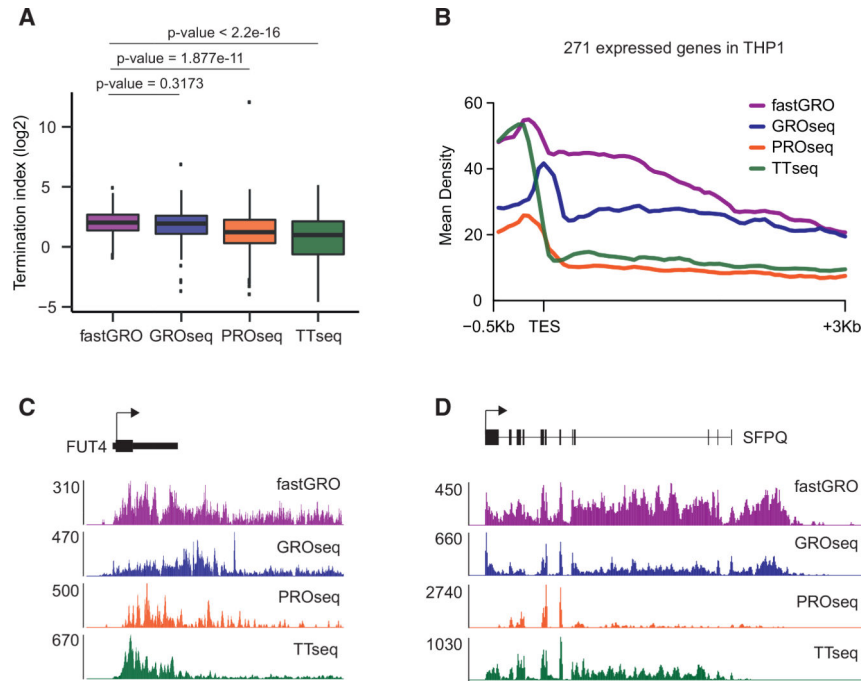
(D) Average profiles of fastGRO, GRO-seq, PRO-seq, and TT-seq at 271 highly expressed genes in THP1 cells. fastGRO shows a lower bias toward the 5' end compared to GRO-seq and recovers more post-termination RNA compared to TT-seq.

(E) Screenshot of the CCNL1 gene comparing fastGRO, GRO-seq, PRO-seq, and TT-seq in THP1 cells.

(F) Average profiles of fastGRO (purple), GRO-seq (blue), PRO-seq (orange), and mNET-seq (black) at 290 highly expressed genes in HeLa cells. fastGRO shows a homogeneous profile along the whole gene body.

(G) Screenshot of the BMP2 gene showing the comparison of fastGRO, GRO-seq, PRO-seq, and mNET-seq tracks in HeLa cells. mNET-seq data are downscaled (right y axis).

(H) Average profile of fastGRO, GRO-seq, PRO-seq, and mNET-seq reads at 50 eRNAs in HeLa cells. fastGRO recovers bidirectional short-lived eRNAs. mNET-seq data are downscaled (right y axis). Comparison between techniques were performed using replicate 1 of fastGRO in THP1 and, where possible, the best replicate of experiments deposited in GEO. In HeLa, one replicate each of CTRL and EGF fastGRO was generated and compared to published data deposited in GEO.



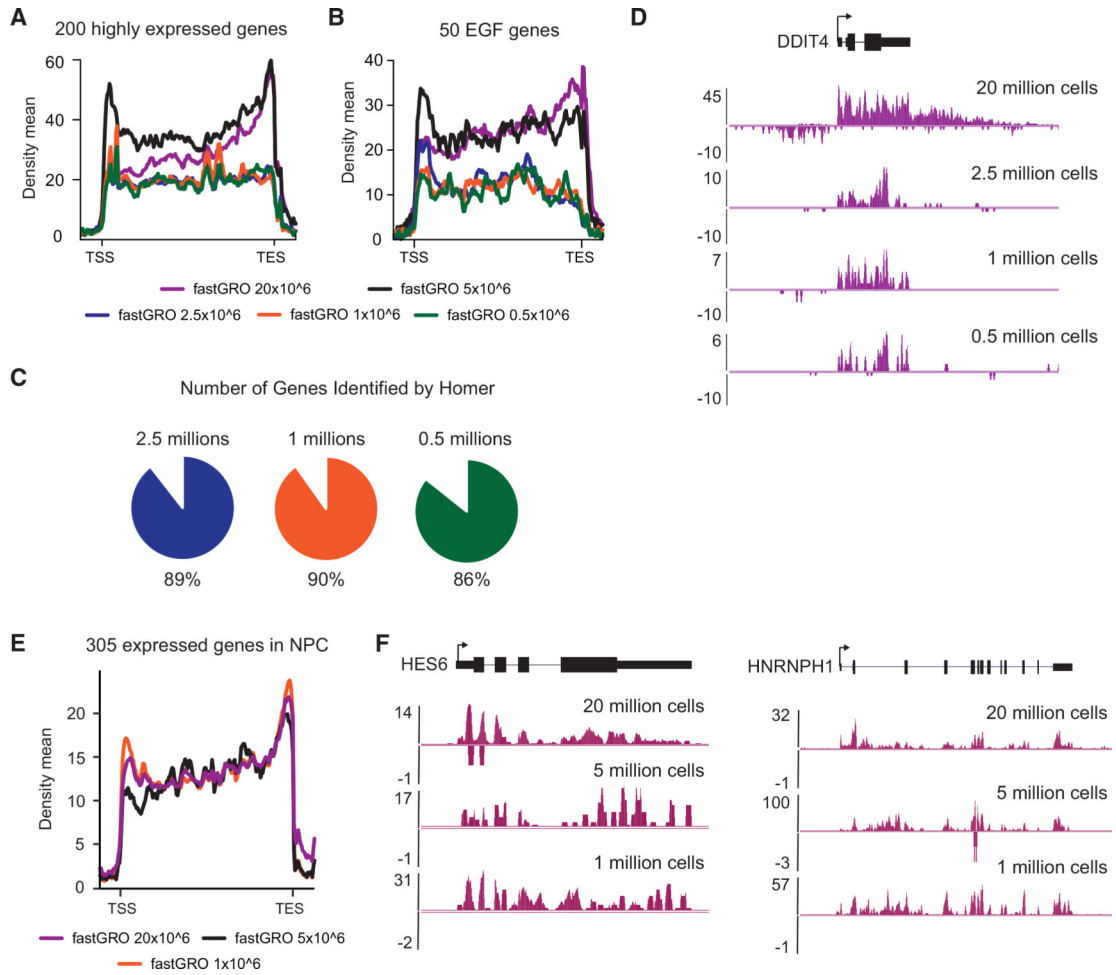
**Figure 4. fastGRO Identifies Transient RNA Downstream of the Poly(A) Signal**

(A) Boxplot of termination index (TI; calculated as ratio between number of reads post-transcription end site (TES/+3 kb) and number of reads pre-TES (-0.5 kb/TES) at 271 most expressed genes calculated from fastGRO (purple), GRO-seq (blue), PRO-seq (orange), and TT-seq (green) data. fastGRO and GRO-seq have comparable TIs, while the TIs generated from PRO-seq and TT-seq are lower (lower coverage of post-termination RNA).

(B) Average profile of reads around TES of 271 highly expressed genes calculated for fastGRO, GRO-seq, PRO-seq, and TT-seq. fastGRO shows the highest and most homogeneous profile pre- and post-TES compared to other techniques used to study nascent RNA.

(C and D) Screenshots of the monocytic gene FUT4 and the constitutively active gene SFPQ showing the high coverage of post-termination RNA retrieved by fastGRO. Comparison between techniques were performed using replicate 1 of fastGRO in THP1 and, where possible, the best replicate of experiments deposited in GEO.





**Figure 5. Low-Input fastGRO**

(A and B) Average profiles of fastGRO reads at 200 highly expressed genes (A) and 50 EGF genes (B) in EGF-treated HeLa cells obtained from 20 million (purple) and 5 million (black) cells using standard protocol and biotin-HPDP and from 2.5 million (blue), 1 million (orange), and 0.5 million (green) cells obtained with the fastGRO-LI protocol and biotin-MTS, indicating that fastGRO can be performed with low numbers of cells.

(C) Pie charts indicate the percentage of genes with FPKM (fragments per kilobase of exon per million fragments mapped) > 0 in fastGRO obtained with 20 million EGF-induced HeLa cells that have FPKM > 0 in fastGRO obtained from 2.5 million (89%), 1 million (90%), and 0.5 million (80%) EGF-induced HeLa cells.

(D) Screenshot of DDIT4 gene showing fastGRO tracks obtained using either standard protocol (20 million cells) or fastGRO-LI protocol (2.5, 1, and 0.5 million cells).

(E) Average profiles of fastGRO reads at 305 expressed genes in iPSC-derived neural progenitor cells (NPCs) were obtained from 20 million cells (purple) using standard protocol and biotin-HPDP. Profiles of 5 million cells (black) and 1 million cells (orange) were obtained with the fastGRO-LI protocol and biotin-MTS, showing that fastGRO can be performed with low numbers of primary-like cells.

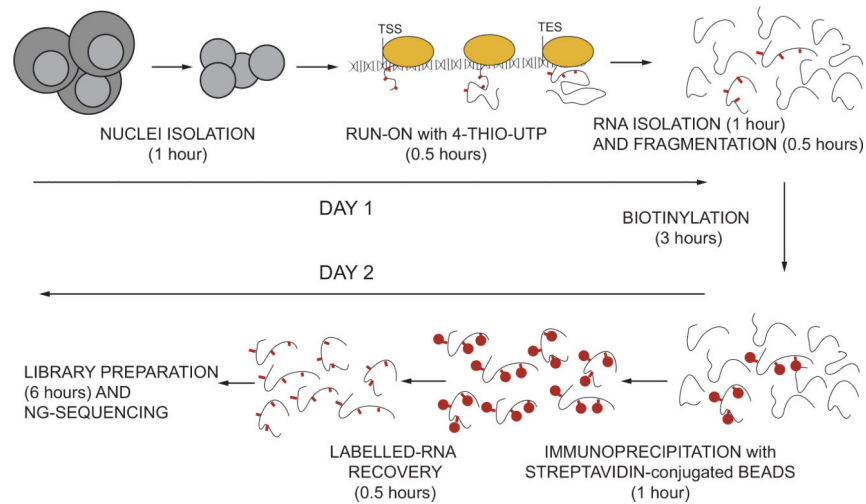
(F) Screenshots of fastGRO tracks for two genes, HES6 (neural specific) and HNRNPH1 (ubiquitously expressed), in NPCs. Scale-down experiments were obtained with 2 replicates of NPC experiments using 1 million and 5 million cells and 1 replicate for all other samples.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 6. Overview of fastGRO**

On day 1, nuclei are isolated, and *in vitro* run-on is performed in a solution containing 4-thio-UTP that is incorporated in nascent RNA. After isolation using TRIzol and ethanol precipitation, RNA is fragmented and snap-frozen. On day 2, 4-thio-UTP-containing RNA is biotinylated using either biotin-HPDP (standard protocol) or biotin-MTS (fastGRO-LI protocol for low-input sample) and recovered by IP using streptavidin-conjugated beads. Labeled RNA is recovered by elution in dithiothreitol (DTT) solution, purified, and used for NGS library preparation with commercially available kits. NG, next generation.

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
PSC Neural Induction Medium	GIBCO	Cat#A1647801
Advanced DMEM/F-12	GIBCO	Cat#12634
Neurobasal Medium	GIBCO	Cat#21103049
ROCK Inhibitor Y27632	Sigma-Aldrich	Cat#Y0503
Geltrex	GIBCO	Cat#A1413302
Flavopiridol hydrochloride hydrate	Sigma-Aldrich	Cat#F3055
Lipopolysaccharide (LPS) Solution	Invitrogen	Cat#00-4976-03
EGF Recombinant Human Protein (rEGF)	GIBCO	Cat#PHG0314
Superase-in RNase Inhibitor	Invitrogen	Cat#AM2694
4-Thiouridine-5'-Triphosphate	Trilink Bio Technologies	Cat#N-1025-1
TRIzol LS reagent	Invitrogen	Cat#10296010
EZ-link HPDP Biotin	Thermo Scientific	Cat#A35390
M280 Streptavidin Dynabeads	Invitrogen	Cat#11205D
MTSEA-biotin-XX	Biotium	Cat#90066
TURBO DNase	Invitrogen	Cat#AM1907
iQ SYBR Green Supermix	BioRAD	Cat#1708880
Oligo(dT) Dynabeads	Invitrogen	Cat#61002
Critical Commercial Assays		
Qubit RNA HS Assay Kit	Invitrogen	Cat#Q32852
Agilent 2200 TapeStation High Sensitivity RNA ScreenTapes	Agilent	Cat#5067
RNA Clean and Purification kit	Zymo	Cat# R1015
NEBNext Ultra II Directional RNA Library Prep kit	New England Biolabs	Cat# E7760S
NEBNext Multiplex Oligos for Illumina (Index Primers set 1)	New England Biolabs	Cat#E7335S
NEBNext Multiplex Oligos for Illumina (Index Primers set 2)	New England Biolabs	Cat#E7500S
Direct-zol RNA MiniPrep kit	Zymo	Cat#R2050
Revertaid first strand cDNA synthesis kit	Thermo Scientific	Cat#K1622
KAPA RNA HyperPrep Kit	Roche	Cat#08098093702
SENSE mRNA-Seq Library Prep Kit V2	Lexogen	Cat#001
Deposited Data		
fastGRO and RNaseq Data	This Paper	GEO: GSE143844
THP1 GRO-seq	Bouvy-Liivrand et al., 2017	GEO: GSM2428733
THP1 PRO-seq	Phanstiel et al., 2017	GEO: GSM2544240
THP1 TT-seq	Godfrey et al., 2019	GEO: GSM3681467
THP1 H3K27ac ChIP-seq	Godfrey et al., 2019	GEO: GSM3681459 and GSM3681461
HeLa GRO-seq	Bouvy-Liivrand et al., 2017	GEO: GSM2428725

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HeLa PRO-seq	Nilson et al., 2017	GEO: GSM2692352
HeLa mNET-seq	Schlackow et al., 2017	GEO: GSM2357382
HeLa long-read mNET-seq	Nojima et al., 2018	GEO: GSM2856679
Experimental Models: Cell Lines		
Human: THP1 Cells	ATCC	Cat# TIB-202
Human: HeLa Cells	ATCC	Cat# CCL-2
Neuronal Progenitor Cells	CHOP Human Pluripotent Stem Cell Core	<a href="https://www.research.chop.edu/human-pluripotent-stem-cell-core">https://www.research.chop.edu/human-pluripotent-stem-cell-core</a>
Drosophila: S2 cells	Capelson Lab	N/A
Drosophila: S9 cells	Capelson Lab	N/A
Oligonucleotides		
Primers (see Table S2)	This Paper	N/A
Software and Algorithms		
seqMINER v1.3.3	Zhan and Liu, 2015	<a href="https://github.com/zhanxw/seqminer">https://github.com/zhanxw/seqminer</a>
ggplot2	Villanueva and Chen, 2019	<a href="https://ggplot2.tidyverse.org/">https://ggplot2.tidyverse.org/</a>
STAR v2.5	Dobin et al., 2013	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
Samtools v0.1.19	Li et al., 2009	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>
deeptools 3.3.1	Ramírez et al., 2016	<a href="https://deeptools.readthedocs.io/en/develop/">https://deeptools.readthedocs.io/en/develop/</a>
HOMER v4.11	Heinz et al., 2010	<a href="http://homer.ucsd.edu/homer/index.html">http://homer.ucsd.edu/homer/index.html</a>
DESeq2	Love et al., 2014	<a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
MAJIQ	Vaquero-Garcia et al., 2016	<a href="https://majiq.biociphers.org/">https://majiq.biociphers.org/</a>
BWA tool	Li, 2013, Li and Durbin, 2010	<a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a>