UNIVERSITY OF CALIFORNIA SAN DIEGO

Template-Switching Enhances Primer Extension Analysis of Transcription Start Sites

A Thesis submitted in partial satisfaction of the
requirements for the degree Master of Science

in

Biology

by

Wilson Lin

Committee in charge:
James Kadonaga, Chair
Barry Grant
Shannon Lauberth

2019

The Thesis of Wilson Lin is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

I dedicate this thesis to Annie, Kendall,

Karen, and Richard for their unwavering love and support.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. James Kadonaga for graciously accepting me into his lab and continually supporting me throughout my project – both as a principle investigator and as an advisor and mentor. His optimistic and genuine advice has been irreplaceable in my endeavors, both within and beyond this project, and for that I am extremely grateful.

I would also like to acknowledge Long Vo Ngoc, who has earnestly and intensively trained me as my primary lab mentor, both in terms of technical knowledge and critical analysis of data. His guidance has been essential in my growth as a researcher, and on numerous occasions, his attention to detail has helped me to overcome multiple barriers in the development of my project.

Finally, I would like to thank the following researchers in my lab: Grisel Cruz, Jia Fei, George Kassavetis, and Cassidy Huang. Throughout my project, they have offered unique perspectives from their own experiences and taught me helpful techniques to pursue my project goals.

ABSTRACT OF THE THESIS

Template-Switching Enhances Primer Extension Analysis of Transcription Start Sites

by

Wilson Lin

Master of Science in Biology

University of California San Diego, 2019

Professor James Kadonaga, Chair

Proximal to transcription start sites, core promoters serve as key regulators of gene expression by sequence-based differential recruitment of transcriptional proteins. Thus, mapping transcription start sites (TSSs) can lead to identification and characterization of core promoter elements, thereby elucidating the relationship between DNA sequence and regulation of gene expression. Using reverse transcription of RNA from active promoters, primer extension can be employed as a quantitative single-step technique for mapping TSSs with high resolution. However, very small amounts of RNA may be challenging to detect with primer extension, and background signals commonly occur due to RNA secondary structures – both issues constrain

the utility of primer extension. One possible avenue for addressing these limitations is template-switching: a mechanism that has improved specificity and yield for numerous RNA assays. Thus, we sought to determine whether template-switching can be used to improve primer extension-based analysis of core promoters. In this paper, we outline obstacles and optimizations for implementing template-switching into primer extension. In modifying dNTP concentrations in primer extension to better suit template-switching, we observed highest template-switching efficiency in low overall (<1mM) dNTP concentrations with a relative excess of dCTP; compared to enzymatic alternatives for template-switching, this optimization may be both novel and cost-effective. Furthermore, we found that this template-switching primer extension approach provided both greater sensitivity and dramatically lower background than the traditional primer extension technique. Therefore, we conclude that template-switching effectively enhances primer extension analysis, thus broadening the usefulness of primer extension in studying core promoters.

.

I. INTRODUCTION

I-I. Traits of Eukaryotic Transcription

Transcription of genes universally involves RNA Polymerase binding to a specific region of DNA (i.e., promoter), usually with the aid of other proteins such as TATA-binding protein (TBP), and this process results in synthesis of RNA. In eukaryotes, this process is more intricate, due to several influential components such as chromatin organization, enhancer specificity, various sequence-specific DNA binding proteins, and the requirement of a large complex of transcription factors to initiate RNA synthesis (Roy & Singer 2015). Moreover, certain transcription factors are not interchangeable – for instance, TBP and TRF2 serve similar functions in setting up transcription, but they recognize different sequences and recruit distinct proteins (Vo Ngoc et al. 2017). Thus, eukaryotes possess sophisticated transcriptional systems, owing to many layers of influence in signaling, binding, and accessibility. Much of this diversity becomes apparent at the promoter – the hotspot of transcription factor recruitment and assembly.

I-II. Promoters and Core Promoters Regulate Eukaryotic Transcription

In eukaryotes, a small portion of the promoter, known as the core promoter, serves as the culminating point of transcriptional signaling. In other words, the core promoter is where enhancers and various transcription factors ultimately converge for synthesis of RNA (Vo Ngoc et al. 2017). Corresponding to different branches of transcription factors and signaling pathways, core promoter sequence motifs include: the TATA box, Initiator (Inr), DPE, TCT, and MTE (Vo Ngoc et al. 2017). Some motifs, such as the TATA box and Inr, play direct roles in recruiting transcription machinery, while others such as DPE play supporting roles (Vo Ngoc et al. 2017). Thus, since sequence-specific DNA-binding proteins will have differential interactions with these motifs, core promoters confer a wide range of transcriptional outcomes.

In addition to directing multiple transcriptional pathways, the core promoter also tunes the level of transcriptional activity. For instance, sequence variations in the TATA core promoter can lead to differential binding affinities for TBP – since TBP is necessary for certain transcriptional pathways to occur, TATA core promoters can thus influence transcriptional activity at the DNA level (Vo Ngoc et al. 2017). Therefore, consistent with its description as the gateway to transcription (Vo Ngoc et al. 2017), the core promoter plays key roles in both modulating and directing gene expression. Further study of core promoter variants and their sequence-specific traits can provide greater insight into the control and balance of eukaryotic gene expression, as well as the genetic basis of human diseases. In order to study core promoters, it is necessary to reliably locate them against the genomic background; while computational analyses have made great advancements in DNA motif searches (Leibovich & Yakhini, 2012), sequence trends alone may not accurately represent the core promoters. Major challenges would include the great diversity of human promoter sequences and the need to discern between functional core promoters and non-functional sequence matches – the latter may require an RNA-based assay (Roy and Singer 2015).

I-III. TSS Mapping can Pinpoint Functional Core Promoters

Regardless of complexity and variation among functional core promoters, all core promoters exist within the immediate vicinity of their transcription start sites (TSSs), typically within 35 nucleotides (Vo Ngoc et al. 2017). Thus, TSS mapping-based approaches can reliably locate functional core promoters (Krieg 1996). Although various TSS mapping techniques exist, they generally involve using RNA transcripts to establish a narrow range of possible TSS positions on the template DNA (typically either by resolving size or sequencing). Additionally,

TSS mapping techniques such as primer extension are compatible with RNA obtained from either *in vivo* or *in vitro* methods, thus conferring flexibility to these approaches (Krieg 1996).

In primer extension, a radioactive P32-labeled primer is annealed downstream of the 5' end on an RNA of interest (Krieg 1996, Appendix Fig A2). Then, a reverse transcriptase enzyme uses the RNA as a template to extend the primer into a cDNA strand up to the 5' end of the template RNA – this end corresponds to the TSS on the promoter sequence (Carey et al. 2013, Appendix Fig A2). Subsequently, the samples go through gel electrophoresis with a size standard or sequencing ladder, and then autoradiography reveals the size or sequence of the cDNA strand. Since the size of a fully synthesized cDNA corresponds to the distance between the primer site and the 5' end of the RNA (which also corresponds to the TSS), the observed product size on gel can be used to approximate the position of the TSS in the original DNA sequence (Krieg 1996). With fully denatured RNA, primer extension yields accurate TSS mapping results (Carey et al. 2013). Additionally, modern primer extension offers high resolution analysis – as low as a single nucleotide – of TSSs (Carey et al. 2013). Furthermore, including both sample handling and gel electrophoresis, this technique can typically yield results within 5-7 hours (Carey et al. 2013). Overall, primer extension is a simple TSS mapping technique that can quickly, accurately, and precisely represent transcription start sites from RNA (Appendix Fig A2).

On the other hand, the primer extension methodology encounters at least two major practical limitations. Firstly, RNA secondary structures (e.g. hairpins) may physically block reverse transcriptase, thus leading to synthesis of 'incomplete' cDNA strands; this can be problematic for analysis because it produces background bands on gel (Krieg 1996). Also, while some RNAs can be completely denatured for primer extension, RNA secondary structures are naturally ubiquitous and serve important biological functions, so it may be unfeasible to

completely avoid these structures (Leamy et al. 2016). Another issue arises from scarce RNA: since primer extension comprises a single annealing step and elongation step, each RNA strand is represented by one radiolabeled cDNA at most – thus, very small amounts of RNA may produce weak or undetectable signals (Carey et al. 2013). Consequently, primer extension analysis may be challenging for 'weak' promoters (i.e., promoters that synthesize low levels of RNA, or produce very unstable RNA). In summary, despite the simplicity and precision of primer extension, both background and low sensitivity are significant drawbacks which may preclude several promoter candidates from study – as such, overcoming these limitations could greatly broaden the usefulness of primer extension in identifying core promoters.

I-IV. Template-Switching may Improve Primer Extension

While primer extension has its weaknesses with sensitivity and background, these issues may be resolved by template-switching – a mechanism that has been increasingly incorporated into many RNA analysis techniques (Zhu et al. 2001, Petalidis et al. 2003, Zajac et al. 2013). For instance, template-switching has been used for assembling full-length transcriptome libraries (Zhu et al. 2001). Additionally, template-switching has excelled in quantitation for RNA microarray assays (Petalidis et al. 2003). In theory, the template-switch mechanism involves a reverse transcriptase enzyme 'switching' from an RNA template to a customized oligonucleotide template, all in one continuous cDNA synthesis process (Zajac et al. 2013). Furthermore, this phenomenon has been observed in multiple variants of both MMLV and AMV reverse transcriptases *in vitro* (Zhu et al. 2001, Ouhammouch & Brody 1992), suggesting that various templates and conditions may be compatible with template-switching approaches.

More specifically, template-switching involves using a reverse transcriptase enzyme to fully synthesize cDNA (5' to 3'), by reading RNA template (primer site to 5' end), and then

continue adding non-templated nucleotides to the cDNA (Zajac et al. 2013). In doing so, the enzyme produces a short overhanging sequence (usually cytosines) on the 3' end of the cDNA – this overhang can then anneal to terminal guanosines on a custom template-switching oligonucleotide (TSO), thus providing a new template for continued elongation of the same cDNA strand (Zajac et al. 2013). In spatial terms, the TSO is adjacently aligned with the 5' end of the RNA template, which is analogous to a longer template strand (Fig 1a). Ultimately, template-switching results in an RNA-templated cDNA sequence with an additional, known TSO-templated sequence at the 3' end (Zajac et al. 2013).

By incorporation of a known sequence to cDNA 3' ends, use of template-switching can streamline differentiation between truncated and full-length primer extension products. For this task, the conditional nature of template-switching comes into play: if reverse transcription halts prematurely due to RNA secondary structures, then template-switching does not occur, so the TSO-templated sequence would be absent in the cDNA product; conversely, products with the TSO-templated sequence would also contain the full-length RNA-templated sequence (Schramm et al. 2000). Thus, if subsequent PCR anneals one primer to the TSO-templated sequence and an accompanying primer to the RNA-templated region, then exponential amplification would occur exclusively for the full-length cDNAs (Zajac et al. 2013). Furthermore, since this TSO-templated sequence flanks one end of the RNA-templated sequence, the TSS position is faithfully preserved within the amplicon, meaning that the start site can be determined by sequencing or by subtracting the known TSO-based additions (Zhu et al. 2001). Altogether, template-switching and subsequent PCR can enrich the population of full-length cDNA products, while increasing the overall yield of products that contain the TSS – both features may substantially improve primer extension without radically changing the protocol.

Since template-switching PCR can selectively amplify full-length primer extension products, this can indirectly alleviate the background issues that typically accompany primer extension. In other words, partially synthesized cDNAs may still exist with similar abundance as in standard primer extension, but the resulting background signals would likely be dwarfed by the amplified full-length products (Zajac et al. 2013). Additionally, while primer extension involves only one round of annealing and elongation, template-switch PCR would add several more rounds of labeled DNA synthesis (Zajac et al. 2013). Thus, whereas typical primer extension techniques may fail to detect very small inputs of RNA, an integrated approach with template-switch PCR may still yield robust product signals. Overall, with potential to address both weaknesses of primer extension (i.e. background and sensitivity), template-switching could be a valuable tool for studying core promoters. As such, the aim of this study is to assess and troubleshoot whether the incorporation of template-switching enhances primer extension in reliably mapping transcription start sites.

II. RESULTS

II-I. MMLV RTase functions in both primer extension and template switching

In a broad context, known 5' ends of RNA transcripts can be mapped onto DNA to locate transcription start sites, which mark functional core promoter regions. We derived an initial template-switching protocol design on primer extension methodology, which is a simple and reliable approach for visually identifying 5' ends of RNA transcripts. By itself, primer extension starts with an RNA transcript, to which a radiolabeled DNA primer is annealed. Then, a reverse transcriptase enzyme extends the primer in a 5' to 3' direction, yielding a cDNA strand that can be visualized on gel. Ideally, the 3' end of the cDNA corresponds exactly to the 5' end of the RNA template for mapping purposes, but some strands can be slightly longer due to the terminal transferase activity of some reverse transcriptases (Zhu et. al 2001).

From here, the methodology of this project branched away from primer extension, as it deliberately utilized the non-templated incorporation to support template-switching, thereby opening downstream options for amplifying desired products. In addition to terminal transferase activity, successful template-switching requires the presence of a template-switching oligonucleotide (TSO). In total, four 30-nt TSOs were designed, each comprised of a distinct 27-nucleotide DNA sequence and three consecutive riboguanosines (rGrGrG) at their 3' ends (Fig 1a, Appendix Fig A1). As some reverse transcriptases have exhibited preferential non-templated incorporation of cytosines (Ouhammouch & Brody 1992, Zajac et al. 2013), the three riboguanosines on the TSOs are intended to reinforce base-pairing between the ends of the TSO and cDNA. According to this design, template-switching products are expected to be 30-nt longer than their simpler primer extension counterparts.

Aside from TSO, the choice of reverse transcriptase may also impact template-switching efficiency due to varying degrees of processivity and terminal transferase activity. AMV RTase is derived from the Avian Myeloblastosis Virus, and has been reported to exhibit *in vitro* template switching activity in cDNA synthesis (Ouhammouch & Brody 1992). On the other hand, MMLV RTase and Sensiscript RT are similarly derived from Moloney Murine Leukemia Virus and have exhibited template-switching (Zhu et al. 2001). We performed primer extension with each enzyme and TSO1, using purified template RNA called Target RNA and ^{32}P-labeled primer, which will also be hereby referred to as labeled reverse primer (Fig 1). Approximately 500-nt in length, Target RNA was synthesized by T7 RNA Polymerase from a plasmid containing a T7 promoter. This construct has a single known transcription start site, and the full-length RNA-templated cDNA product is expected to be 140-nt, while template-switch product should be 170-nt. All three enzymes generated strong and comparable 140-nt bands, indicating that all three are capable of effective primer extension under similar conditions (Fig 1). However, only extension with MMLV RTase produced 170-nt bands, and did so best at 37C (Fig 1). As such, these results suggest that MMLV RTase at 37C is best suited to template-switching, which is consistent with multiple other template-switching protocols (Zhu et al. 2001, Petalidis et al. 2003, Zajac et al. 2013).

II-II. Template switching efficiency depends on TSO sequence and dNTP ratios

Although template-switching activity was detectable in the initial screening, the corresponding 170-nt bands were still difficult to visualize without digitally increased exposure (Fig 1), so we also screened buffers and TSOs for template-switching compatibility (Fig 2). In the case of TSOs, the varying 27-nt regions may lead to different folding and thermostability, or other unknown interactions within the template-switching system. While all four TSO variants

produced comparable cDNA bands at 140-nt, TSO3 and TSO4 produced the strongest 170nt bands for both buffers (Fig 2a). This suggests that both TSO3 and TSO4 are both more optimal for template-switching than their counterparts, which is consistent with the low template-switching performance of TSO1 in both experiments (Fig 1, 2).

Despite the reported cytosine bias, it is possible that MMLV RTase terminal transferase activity still exhibits a high degree of randomness, which reportedly only increases with distance from the original cDNA 3' end (Zajac et al. 2013). Thus, we suspected that altering dNTP ratios may impact the selectivity of the terminal transferase activity. While H buffer followed other template-switching protocols in terms of using 1mM of each dNTP (Zhu et al. 2001, Petalidis et al. 2003, Zajac et al. 2013), L buffer was depleted to 0.2mM of all dNTPs except dCTP, which remained at 1mM (Fig 2a). These two buffers were otherwise identical and used an abundant 4ng of Target RNA, restricting any effects to absolute or relative availability of each dNTP. Compared to the H buffer, the depleted L buffer exhibited a weaker 140-nt band, indicating that the depletion of dNTPs likely limited the overall yield of primer extension (Figure 2a). Based on this effect, even with equivalent TS efficiencies, the L buffer would be expected to produce much fewer template-switch products than the H buffer. However, the absolute intensity of each 170-nt template-switch band in L buffer exceeds those from H buffer (Fig 2a), suggesting that the dNTP depletion in L buffer results in a net gain of template-switching product. Additionally, the observed decrease in 140-nt cDNA yield may be attributed to the 4ng of RNA, which may be stoichiometrically in excess of the depleted dNTPs.

To further optimize dNTP concentrations in primer extension buffer for efficient template-switching, primer extension was carried out with varying ratios and absolute concentrations of dNTPs (Fig 2c). Since template-switching requires CCC-tailing on a full-

length cDNA (Krieg 1996), dNTP ratios were tailored to favor cytosine incorporation by a molar excess of dCTP relative to other dNTPs. The RNA input was also decreased to 1ng to test whether the yield-limiting effect occurs for lower quantities of RNA template (Fig 2b). With reduced Target RNA input, the 140-nt bands appeared more comparable across all tested buffers (Fig 2b), thus confirming that the previously observed loss in yield for dNTP-depleted buffer is likely negligible for smaller inputs of RNA.

To assess overall template-switching performance, we analyzed both template-switching efficiency and absolute yields of template-switch products. Using the ratio of 170nt:140nt bands in each lane as an estimation of template-switching efficiency (Fig 2c), the highest efficiency was observed with 0.6mM dCTP and 0.2mM of dATP, dGTP, and dTTP; the second-highest efficiency was observed with 1mM dCTP and 0.2mM dATP, dGTP, and dTTP. Additionally, these two conditions exhibit the greatest absolute yield of template-switching products, based on comparison of all 170nt bands (Fig 2c). In both instances of strong template-switching product bands, 1mM or less dCTP was used, along with smaller proportions of the other dNTPs. On the other hand, the equimolar 1mM dNTP buffer, as well as the remaining buffers with >1mM dCTP, were observed with weak or negligible template-switching products in terms of autoradiography. Overall, more template-switching products and higher template-switching efficiencies were observed in conditions that used 1mM or less of dCTP alongside smaller amounts (i.e. 0.2mM) of the other dNTPs. Collectively, these results suggest that depletion of dATP, dGTP, and dTTP can create a more suitable buffer for template-switching, whereas enrichment of dCTP does not.

II-III. Reagents of *in vitro* transcription can inhibit or enhance template-switching

After consolidating optimal elements, our template-switching protocol produces efficient template-switching from pure RNA. However, the ultimate goal of this technique is to enhance detection of transcripts from weak promoters, meaning that it must be compatible with *in vitro* transcription reagents that carry over into purified RNA. Since the concept would be initially simpler to demonstrate with abundant RNA transcripts, we chose a plasmid promoter construct with high transcriptional yield, hereby referred to as pSCPX. SCPX stands for super core promoter X, which has been previously developed in our lab and contains a cluster of functional core promoter motifs, highly consistent with their respective consensus sequences (Juven-Gershon et. al 2006). More specifically, it contains the TATA box, Initiator, DPE (downstream promoter element), and MTE (motif ten element), which cumulatively drive high levels of RNA Polymerase II transcription (Juven-Gershon et. al 2006). Thus, *in vitro* transcription on this construct should produce an abundant amount of RNA, such that it is easily detectable even by standalone primer extension.

To begin with, *Drosophila melanogaster* nuclear extract was added to the plasmid to provide transcription factors, and Target RNA was simultaneously spiked in as a control. After *in vitro* transcription, products were purified by phenol-chloroform and chloroform extractions, then precipitated and washed with ethanol. The products were then resuspended in reverse transcription buffer with and without TSO to assess for detectable template-switching. The same (Fig 1-2), previously used radiolabeled primer was annealed to both Target and SCPX RNAs, and their cDNA products were run on gel for size determination and quantification.

From standalone primer extension (i.e., without template-switching), Target cDNA and SCPX cDNA are expected to be 140-nt and 116-nt, respectively. For template-switching products, Target cDNA and SCPX cDNA should be 170-nt and 146-nt, respectively, due to the

30-nt addition from template-switching. The non-template-switch products are apparent for both Target and SCPX, indicating that primer extension functions properly when used to follow up *in vitro* transcription (Fig 3a). However, no template-switching bands were observed for either product when TSO was added to the primer extension reaction; the only exception was the control primer extension, which was performed on Target RNA without the preceding *in vitro* transcription step. In short, template-switching activity - but not primer extension - disappeared upon integration of the *in vitro* transcription step. The purifications between these two steps likely removes most proteins and salts from *in vitro* transcription, but other reagents may still carry over.

To determine whether *in vitro* transcription reagents were inhibiting template-switching, we screened individual reagents from this step as direct additives into isolated primer extension/template switching reactions (Fig 3b). Aside from predicted denaturants such as chloroform, the rNTP addition most thoroughly diminished the expected 170-nt template-switching band, such that the signal was equivalent to background (Fig 3b). On the other hand, the addition of PEG in PvOH appeared to produce a much stronger 170-nt band, almost equivalent in intensity to its respective primer extension band at 140-nt (Fig 3b). Among a range of possible final PEG concentrations, there is a pronounced increase in the template-switching band intensity from 0% to 0.25% PEG, but greater concentrations show little change thereafter (Fig 4). Together, these results suggest that *in vitro* transcription reagents can lead to downstream inhibition of template-switching, except for PEG which may increase template-switching yields.

II-IV. Establishing compatibility between *in vitro* transcription and template-switching

Given that our development of template-switching is aimed towards improved detection of scarce RNA transcripts, it would be problematic if *in vitro* transcription prevents template-switching. As such, we sought out a purification method that could remove rNTPs from the *in vitro* transcription mix more effectively than phenol:chloroform extraction, which is not typically used for this purpose. One purification method of interest was Trizol extraction, since it is known to separate proteins, DNA, and RNA into distinct phases, while disrupting interactions between any of these (Rio et. al., 2010). As an initial assessment of Trizol in rNTP removal, we used Trizol extraction on mixtures of Target RNA with or without rNTPs (Fig 5). We followed this with isopropanol precipitation at room temperature and washed with 75% ethanol. The pre-extraction and post-extraction mixtures were compared both on agarose via EtBr staining and Nanodrop. For samples that contained Target RNA, 500-nt bands of comparable intensity were observed before and after extraction when analyzed on gel with EtBr staining (Fig 5a). Accordingly, the Nanodrop quantification reported minimal changes in Target RNA mass after extraction (Fig 5b). In contrast, Nanodrop values for combined Target RNA with rNTPs dropped to levels comparable to Target RNA alone, whereas no RNA was detected after extraction of rNTPs alone (Fig 5b). Collectively, these show that Target RNA transcripts are efficiently recovered by Trizol extraction whereas rNTPs are vastly removed. Next, we used the same Trizol extraction method on *in vitro* transcription mix with Target RNA spiked in, used radiolabeled primer extension or template-switching, and assessed levels of Target cDNA products on gel (Fig 6). Since the previous assay only demonstrated rNTP removal, we wanted to confirm that this was sufficient to restore detectable levels of template-switching after *in vitro* transcription.

We additionally tested a range of %PEG in the primer extension step again, in case the optimal value changed with this purification method. As a control, template-switching was performed on Target RNA at 0.25% PEG with no preceding Trizol or *in vitro* transcription step (Fig 6). We used 0.25% PEG as opposed to a higher value because 0.25% and 0.50% previously exhibited negligible differences in template-switching efficiency (Fig 4). Notably, the template-switching band intensity of Trizol-extracted Target RNA with 0.50% PEG was comparable to the template-switching controls (Fig 6). Thus, these results suggest that Trizol extraction is more suitable than phenol:chloroform extraction in supporting compatibility between *in vitro* transcription and template-switching.

II-V. Primer extension and template-switching reagents can inhibit subsequent PCR

With *in vitro* transcription and template-switching working together sufficiently well, the last step to integrate was PCR amplification. To reiterate, the primary motivation behind utilizing template-switching with primer extension was to add a PCR handle that flanks the 3' end of the cDNA. This allows for subsequent amplification of the resulting cDNA that faithfully represents the 5' end of the original RNA transcript, since the added length is known. In order for this entire method to enhance primer extension, the PCR step must dramatically increase the number of target products while minimizing off-targets and noise.

Due to the critical importance of the PCR step, we assessed inhibition of PCR by primer extension/template-switching reagents. One PCR primer was identical to the TSO except without the riboG stretch, and the other primer was identical to the radiolabeled primer except also available as non-radioactive. Since this test required a PCR template that was independent of primer extension, we designed a plasmid construct almost identical to the one used to synthesize

the Target RNA stock, with one key difference: the TSO sequence was mutated into it, immediately upstream of the TSS. This means that the amplicon is nearly identical in both sequence and size to the actual Target template-switch cDNA, but exists without any use of primer extension. In short, we assembled the typical template-switching mix (Fig 1-4), but without incubation, and then loaded varying volumes into a GoTaq PCR mix containing the PCR control plasmid (Fig 7). The unlabeled PCR products were then analyzed by PAGE and EtBr staining. Remarkably, the addition of primer extension mix quantitatively lowered PCR yields starting at 20% primer extension mix by volume, and then completely diminished PCR bands at 40% volume (Fig 7a). In terms of balancing input volume against inhibition, the band intensity 10% mix by volume was comparable to 0%, suggesting that this proportion is safe to use with negligible inhibition (Fig 7a).

We then wanted to determine whether this effect was reproducible in using actual primer extension/template-switching products for PCR rather than the control plasmid. In this case, an interesting premise arose as to whether the disinhibition of PCR by using reduced volumes of template-switch product would outweigh the loss of initial PCR template in doing so. Similar to previous template-switching experiments (Fig 1-4), we generated template-switched Target cDNA via primer extension with TSO, except with unlabeled primer. We then directly compared PCR yields after using 10% and 40% (relative to PCR) input volumes of these products, equivalent to loading 25% and 100% of the template-switch product, respectively (Fig 7b). Indeed, using 25% of the template-switch product produced substantially higher PCR yield than using 100%, suggesting that PCR inhibition was profound with increased inputs of the primer extension mixture (Fig 7b). Overall, these suggest that primer extension or template-switching

reagents can severely inhibit PCR efficiency, requiring that limited amounts be used to optimize yields.

II-VI. Optimized template-switch PCR with radiolabeled primer substantially exceeds the RNA detection threshold of standalone primer extension

From *in vitro* transcription to primer extension and PCR, the collective troubleshooting up to this point now meets practical requirements for detecting RNA transcripts. However, the ultimate goal of this project is to surpass the detection threshold of standalone primer extension, such that transcriptionally weak promoters can still easily be studied with this quick benchtop method. This also means that the PCR step must be able to efficiently amplify the desired products with radiolabeled primer, rather than unlabeled primer, due to the need to detect quantities of RNA that are likely below the detection threshold of ethidium bromide staining.

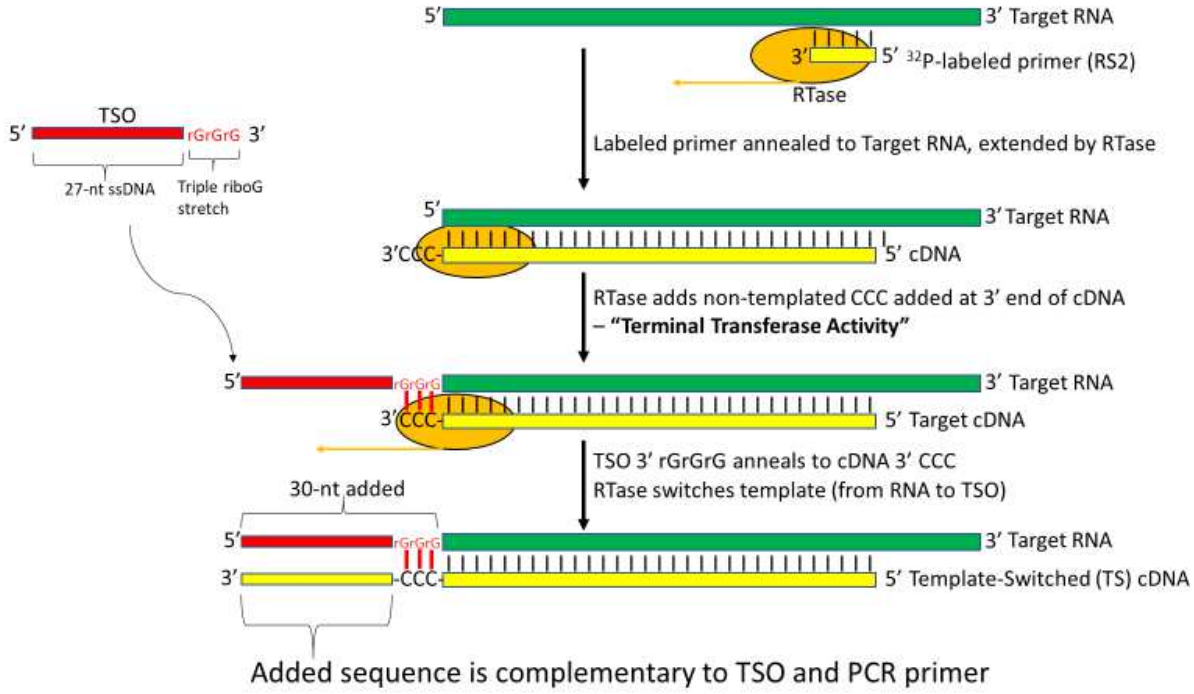
One immediate problem we noticed is that our radiolabeled primer exists in concentrations that are sufficient for primer extension but far too dilute for PCR (<1nM). This dilemma is rooted in the primer labeling method, which labels very small quantities of unlabeled primer with isotope to maximize the proportion of labeled primer produced. Assuming that this aspect stays the same, we sought instead to supplement the scarce radiolabeled primer with its otherwise identical unlabeled counterpart. For distinction, both of these primers will be referred to as reverse PCR primers, whereas the always unlabeled forward primer anneals to the TSO-based sequence. As an initial PCR template, we again used the PCR control plasmid with built-in primer sequences and nearly identical amplicon as template-switch Target cDNA (Fig 7a). For amplification, varying amounts of unlabeled reverse primer were combined with fixed levels of labeled reverse primer and unlabeled forward primer (Fig 7a). More specifically, labeled reverse primer was diluted 20-fold from stock into the PCR reaction, whereas forward primer was kept at

1μM. Like with radiolabeled primer extension, the final PCR products were loaded into PAGE and analyzed via autoradiography. The higher bands at 170-nt represent the amplified and labeled product. Additionally, the bands below these are likely part of the same product but migrating unusually due to large amounts of DNA (labeled and unlabeled) present in the lane (Fig 7a). This is supported by running the same volumes of product with larger volumes of loading buffer, which results in the lower bands diminishing while the higher bands stay constant in intensity (Fig 7a). The combined intensity of both bands suggests that 0.25μM of unlabeled reverse primer produces the greatest yield of the desired PCR product.

Lastly, we tested the final capabilities of the template-switch-PCR system by comparing it to standalone primer extension with very small inputs of Target RNA (Fig 8b). This system combines all the adjustments detailed throughout Figures 1-5, as well as using 0.25μM of unlabeled reverse primer (Fig 8a). Notably, the standalone primer extension failed to produce a detectable 140-nt band for RNA inputs below 40pg, whereas template-switch-PCR produced its expected 170-nt band for even the lowest input (0.4pg) within 20 cycles (Fig 8b). Furthermore, the 170-nt bands are all strongly detectable by 30 cycles of PCR. Both above and below the 170-nt bands, weaker bands were observed but were diminished nearly to background intensity in parameters where the 170-nt band is only moderately detectable (Fig 8b). To reiterate, only 25% of template-switch product is used for PCR and then 5% of the PCR product is loaded into gel. Despite such large dilutions, these results indicate that the finalized template-switch PCR protocol at least 200-fold exceeds the minimal RNA detection threshold of standalone primer extension.

Figure 1. MMLV reverse transcriptase (RTase) exhibits template-switching activity following primer extension in the presence of template-switching oligonucleotide (TSO). (a) Diagram of primer extension and template-switching processes *in vitro*, showing interactions between TSO, RTase, and RNA template. All TSO designs share a stretch of 3 riboguanosines at their 3' ends, but otherwise have variable 27-nt DNA sequences. (b) Screening of MMLV RTase, AMV RTase, and Sensiscript RT for detectable template-switching activity, by using radiolabeled primer extension on Target RNA, followed by PAGE and autoradiography. Different temperatures were also tested for MMLV RTase. Without template-switching, Target cDNA is expected to be 140-nt, whereas the template-switching product should be 170-nt.

1a


5′
3′ Target RNA
3′
5′ 32P-labeled primer (RS2)
RTase
TSO
5′
rGrGrG 3′
27-nt ssDNA
Triple riboG stretch
Labeled primer annealed to Target RNA, extended by RTase
5′
3′ Target RNA
3′CCC-
5′ cDNA
RTase adds non-templated CCC added at 3′ end of cDNA
– "Terminal Transferase Activity"
5′
rGrGrG
3′ Target RNA
3′CCC-
5′ Target cDNA
TSO 3′ rGrGrG anneals to cDNA 3′ CCC
RTase switches template (from RNA to TSO)
30-nt added
5′
rGrGrG
3′ Target RNA
3′
-CCC-
5′ Template-Switched (TS) cDNA
Added sequence is complementary to TSO and PCR primer

1b

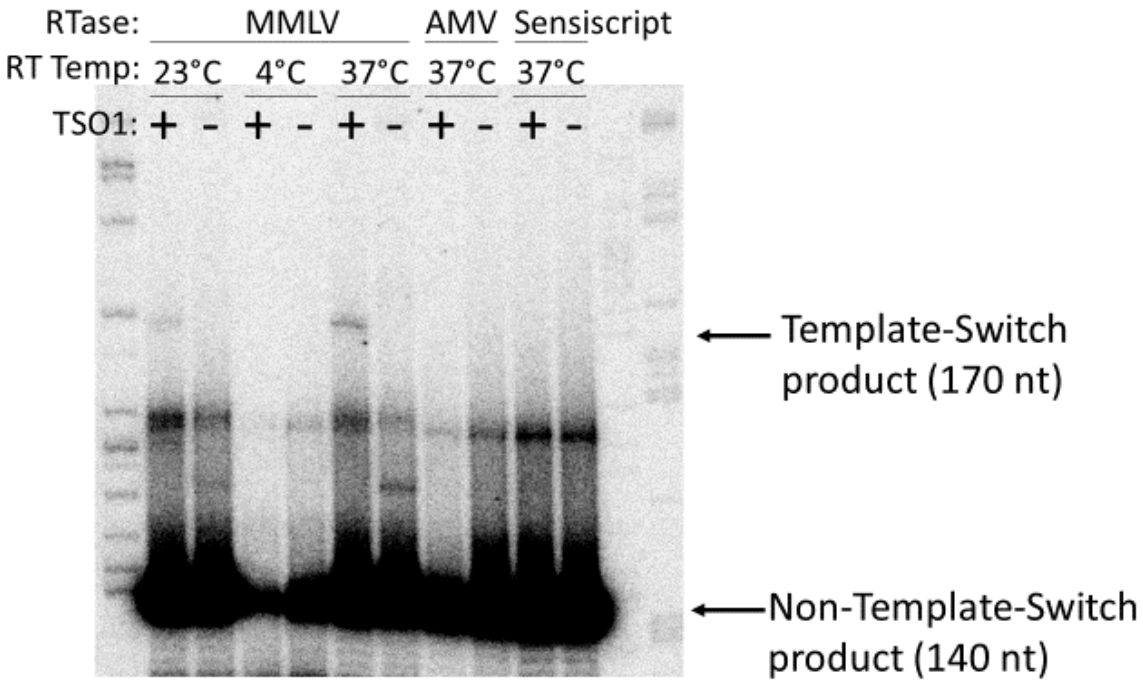
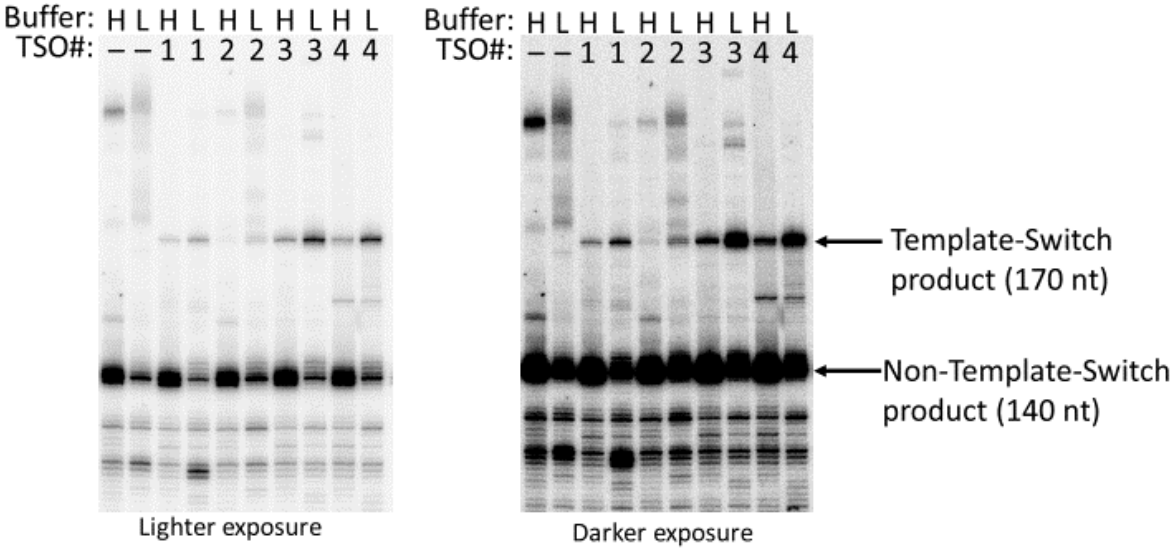
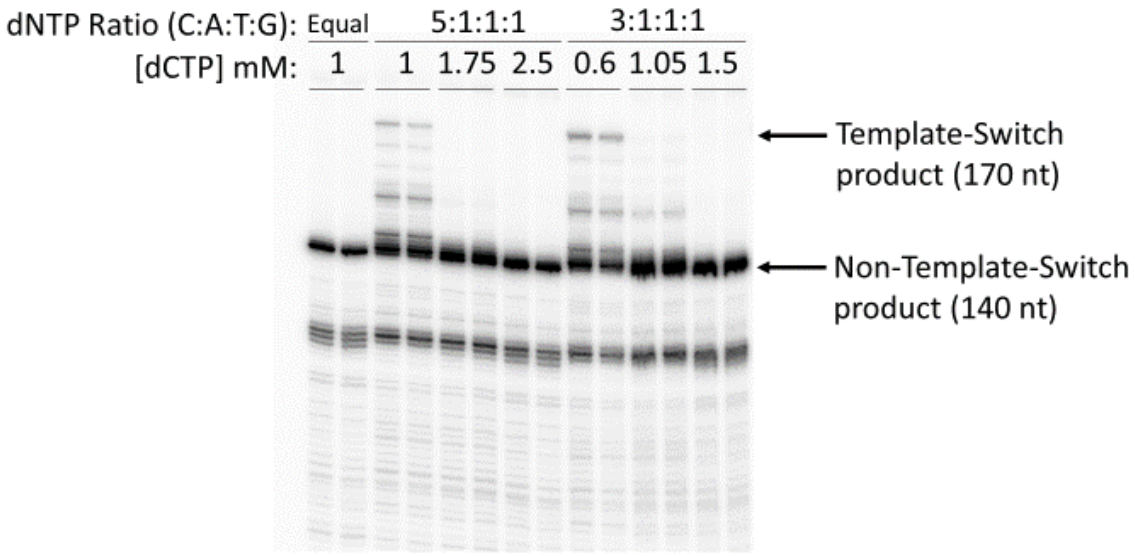

RTase: MMLV AMV Sensiscript
RT Temp: 23°C 4°C 37°C 37°C 37°C
TSO1: + - + - + - + - + -
Template-Switch product (170 nt)
Non-Template-Switch product (140 nt)

Figure 2. TSO design and dNTP ratios are critical factors in template-switching efficiency. (**a**) Screening of TSO designs 1-4 (listed in Appendix Fig. A1) with H and L buffers, where H contains 1mM of each dNTP and L contains 1mM of dCTP and 0.2mM of other dNTPs. **(b)** Comparison of template-switching efficiency across varying dNTP concentrations. Bands in **(a)** and **(b)** represent radiolabeled cDNA, synthesized by reverse transcription on 4ng (**a**) or 1ng (**b**) of T7P RNA; P^{32}-labeled primer was annealed to the RNA, 140 nucleotides downstream of the TSS. Without template-switching, a full-length cDNA is expected to be 140nt, whereas a template-switching product is expected to be 170nt. Other pronounced bands are not well-defined, such as those below the 140nt region; these may be background signal or incomplete synthesis. **(c)** Calculated values for averaged template-switching efficiency in each condition, along with summed intensity of 140nt and 170nt bands to obtain total yield. Template-switching is calculated as the percentage of 170nt band intensity over the total yield in each lane. N=2 for each dNTP composition.

2a



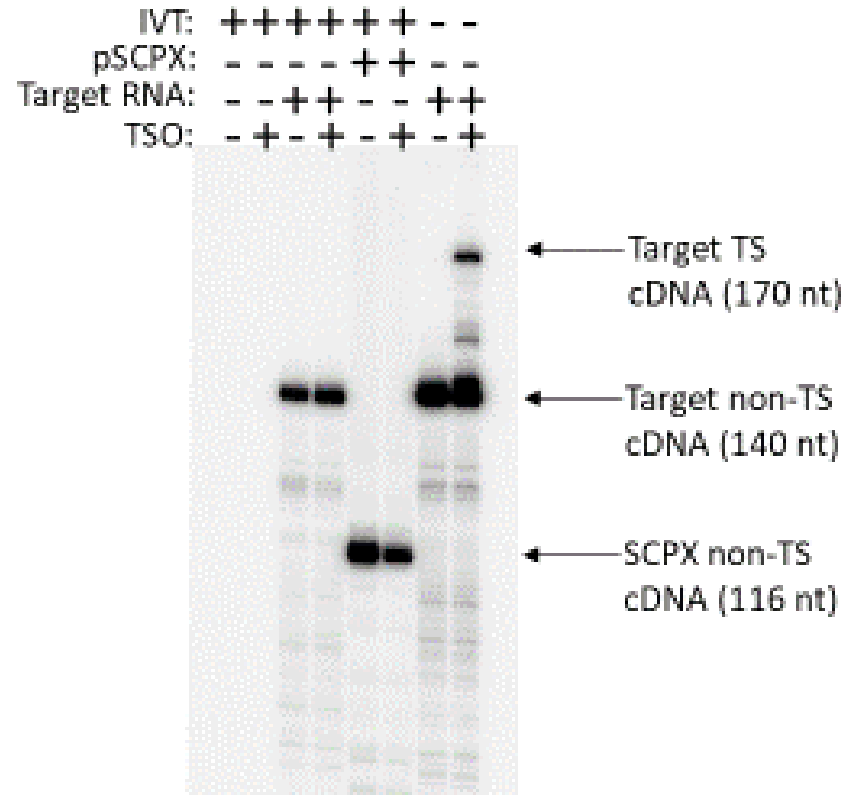
2b



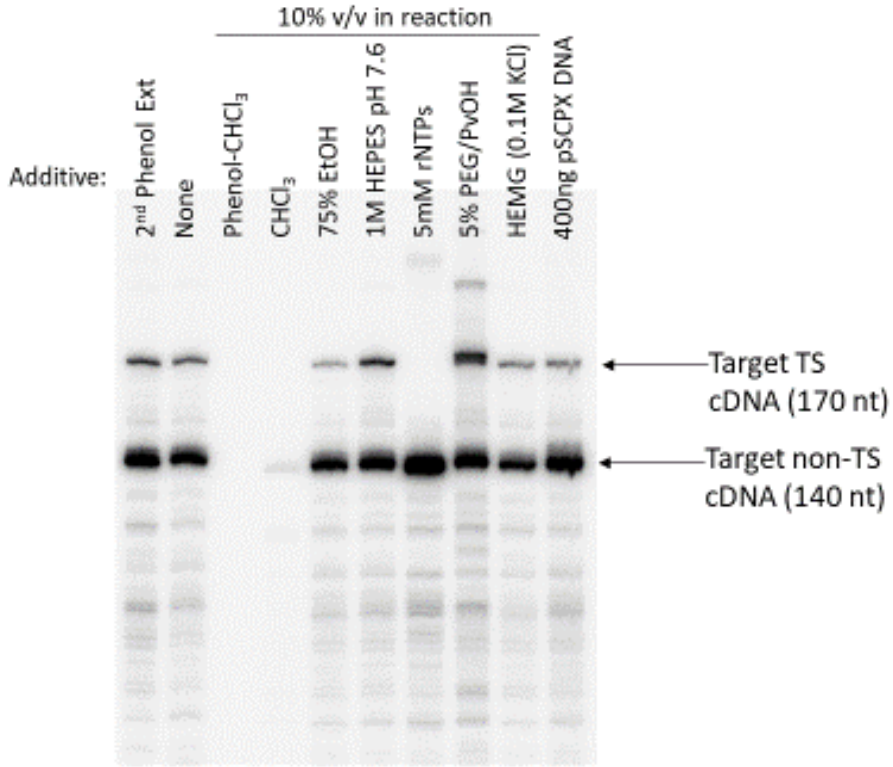
dNTP ratio (C:A:T:G)	[dCTP] in mM	Average % Template-Switching Efficiency	Total Yield (arbitrary units)
1:1:1:1	1.00	0.095	1.00
5:1:1:1	1.00	2.228	1.30
5:1:1:1	1.75	0.070	1.63
5:1:1:1	2.50	0.101	1.39
3:1:1:1	0.60	6.752	1.16
3:1:1:1	1.05	0.161	1.80
3:1:1:1	1.50	0.111	1.66

Figure 3. **Template-switching is inhibited by *in vitro* transcription reagents that persist through purification.** (**a-b**) Gel analysis of radiolabeled primer extension and template-switching products upon addition of *in vitro* transcription reagents. (**a**) Purified products of *in vitro* transcription on pSCPX, with Target RNA spike-in, were extracted with phenol:chloroform prior to primer extension. Target non-TS and TS products appear as 140-nt and 170-nt bands, respectively. SCPX non-TS and TS products are predicted to appear as 116-nt and 146-nt bands, respectively. (**b**) Screening of multiple *in vitro* transcription reagents for potential inhibition of template-switching activity, by adding individual reagents to primer extension reactions on Target RNA. As a control (**a** and **b**), TS was performed on Target RNA with no prior *in vitro* transcription or phenol:chloroform purification step. TS stands for template-switching.

3a


IVT: + + + + + + - -
pSCPX: - - - - + + - -
Target RNA: - - + + - - + +
TSO: - + - + - + - +
Target TS cDNA (170 nt)
Target non-TS cDNA (140 nt)
SCPX non-TS cDNA (116 nt)

3b


10% v/v in reaction
Additive:
2nd Phenol Ext
None
Phenol-CHCl3
CHCl3
75% EtOH
1M HEPES pH 7.6
5mM rNTPs
5% PEG/PvOH
HEMG (0.1M KCl)
400ng pSCPX DNA
Target TS cDNA (170 nt)
Target non-TS cDNA (140 nt)

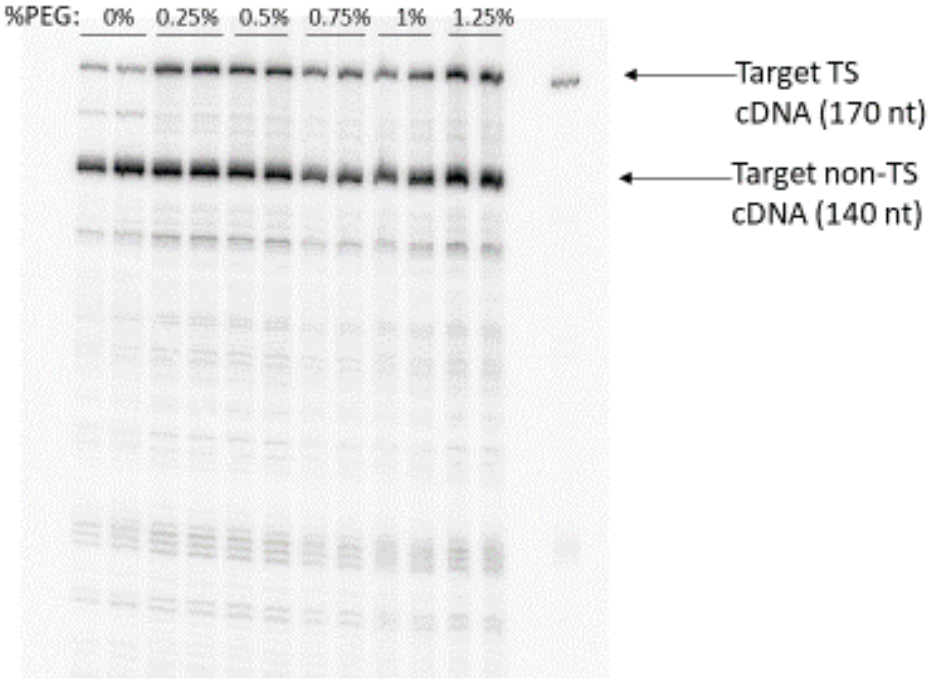
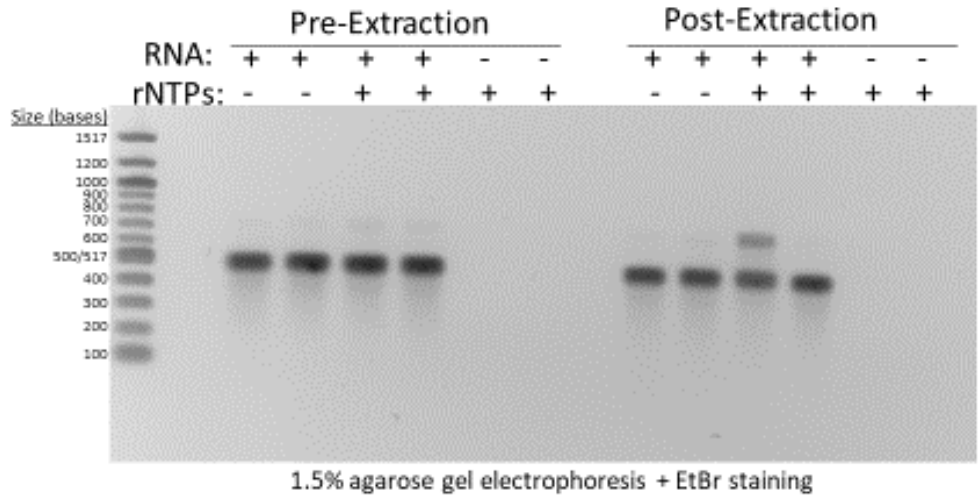


Figure 4. Addition of PEG substantially increases template-switching efficiency. Ranging PEG concentrations were tested for quantitative, positive effects on template-switching. Target RNA was used as the starting material for template-switching, with TSO3. Zero PEG was used as a control condition for the default protocol. Radiolabeled products were ran on denaturing 8% acrylamide gel and detected via autoradiography.

5a



5b

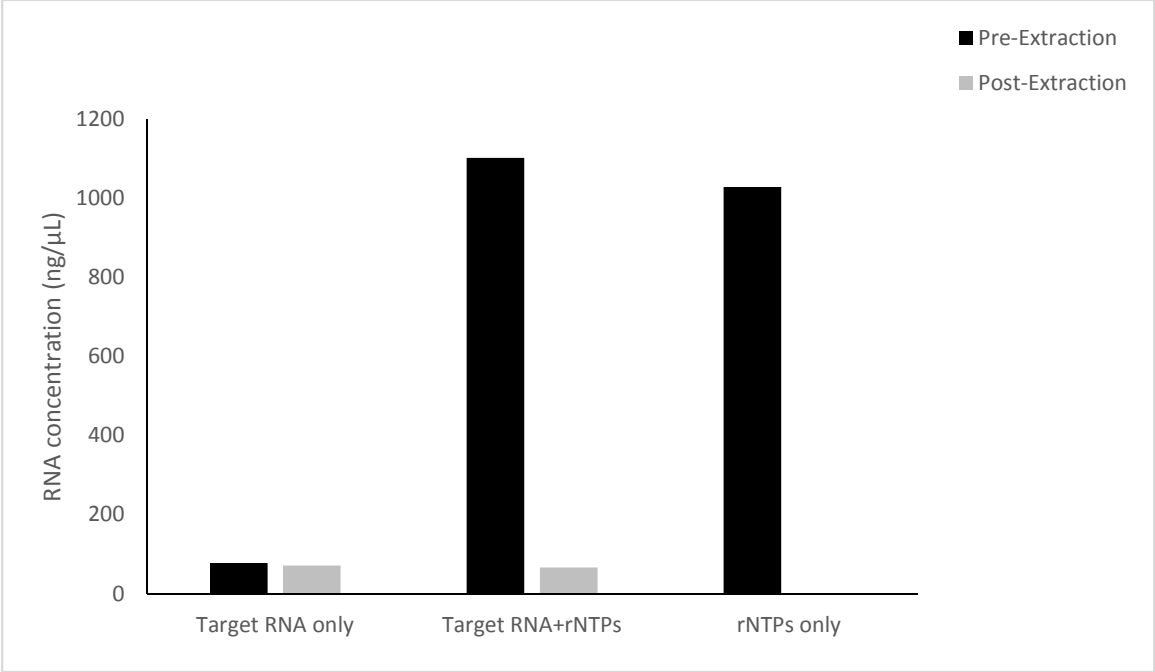


Figure 5. Trizol extraction efficiently removes free ribonucleotides from solution while recovering most RNA transcripts. (**a-b**) Duplicate 25µL mixtures of 2.5µg Target RNA and/or 20nmol of each rNTP were left alone or Trizol-extracted, then resuspended in dilution-equivalent volumes. Then, (**a**) Target RNA recovery was visualized as 500-nt bands on agarose with EtBr staining, while (**b**) overall RNA was quantified via Nanodrop to measure both Target RNA and rNTP recovery. TS is template-switching.

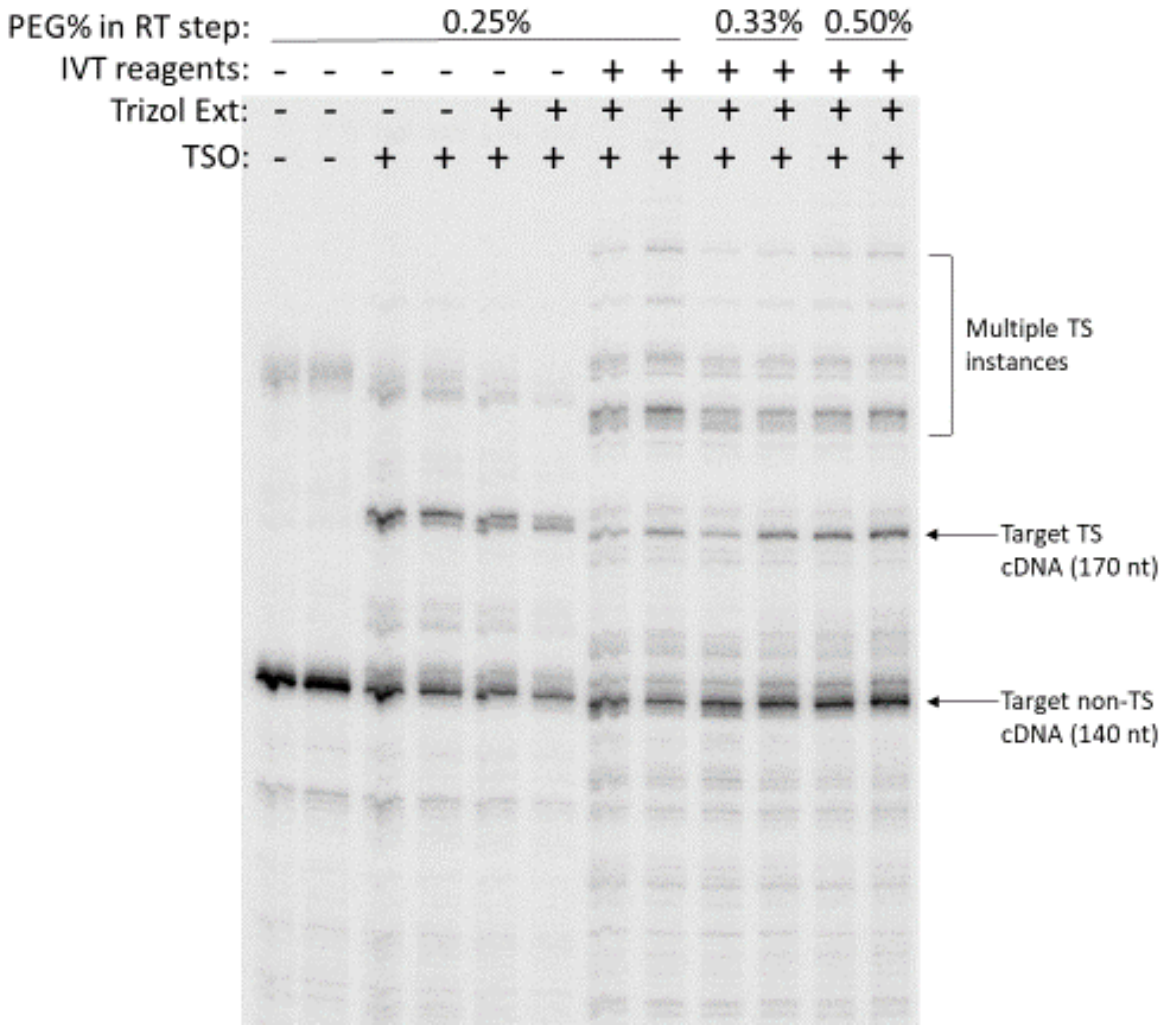
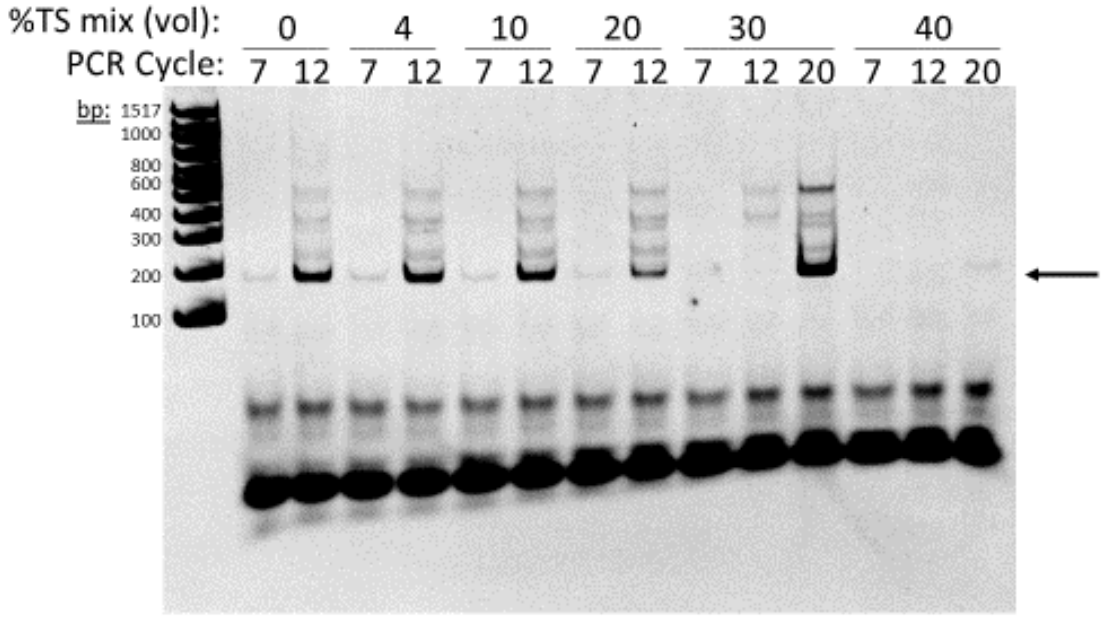


Figure 6. Trizol extraction restores template-switching activity downstream of *in vitro* transcription. Trizol extraction was used to purify *in vitro* transcription mix with Target RNA spiked in, and the resulting pellet was used for radiolabeled primer extension and template-switching under various %PEG concentrations. PAGE of resulting cDNA shows 170-nt (TS) bands and 140-nt non-TS bands, while ladderlike bands above 170-nt likely represent multiple successive TS events. As controls (leftmost four lanes), primer extension and template-switching were performed on Target RNA without preceding *in vitro* transcription or Trizol extraction steps. Each condition was tested in duplicates. TS is template-switching.

7a



7b

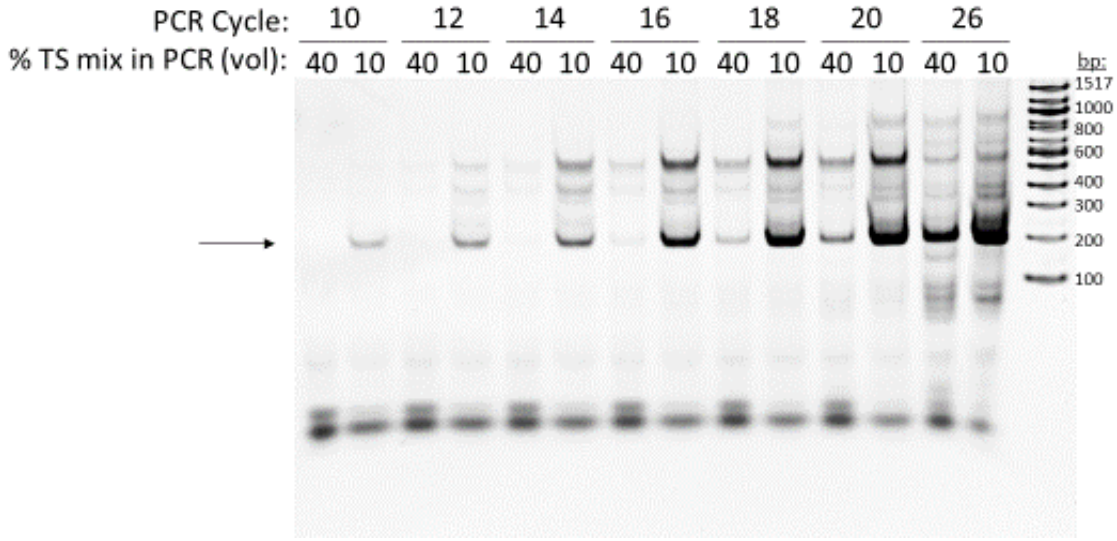
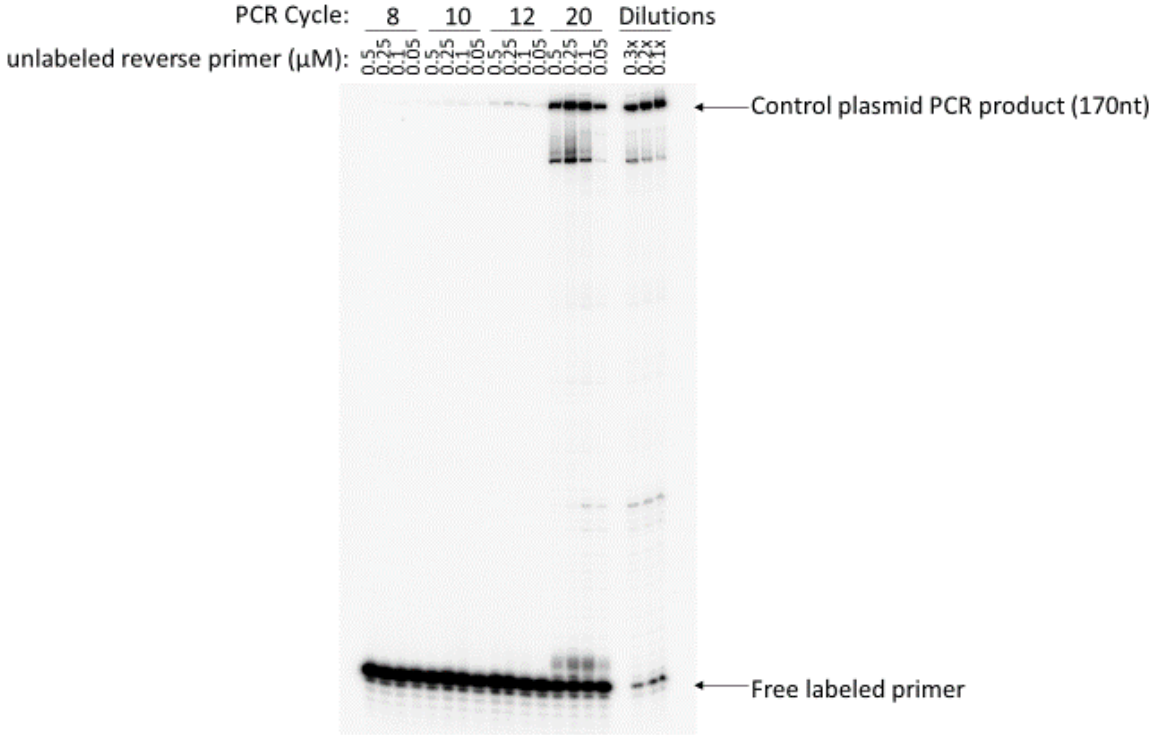


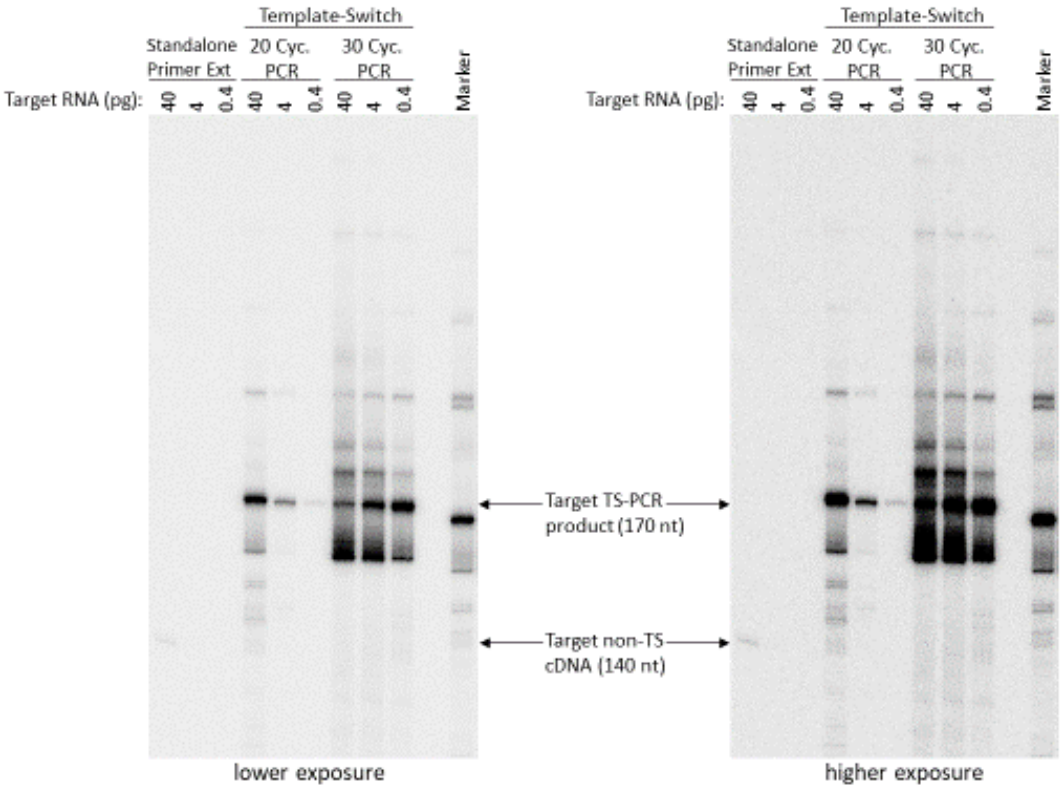
Figure 7. PCR efficiency is quantitatively reduced by reagents used for template-switching. Amounts of primer extension mix are defined as percent volume of the PCR reaction, with a maximum of 40%. (**a**) PCR was performed with a plasmid construct imitating template-switch Target cDNA in terms of sequence and size. A wide range of PE mix volumes were used to examine quantitative inhibition. (**b**) PCR was performed on template-switch Target cDNA derived from primer extension, using 10% and 40% input volumes only but measuring at much later cycles than in (**a**). Products of both (**a**) and (**b**) were run on PAGE and stained with EtBr. Arrows indicate location of 170-nt PCR product.

Figure 8. Efficient Template-Switch PCR demonstrates increased detection of RNA over standalone primer extension. (**a**) Varying amounts of unlabeled reverse primer were used to supplement its labeled equivalent, along with unlabeled forward primer, in order to amplify a plasmid with an amplicon that mimics Target template-switch cDNA. This produces a 170-nt PCR product as well as lower bands that diminish when equivalent amounts of product are diluted and loaded. Dilutions were made from sample with 0.25μM unlabeled rev. primer after 20 cycles of PCR, and listed values represent fold dilution into loading buffer. (**b**) Comparison of standalone primer extension and template-switching with PCR, both using small amounts of Target RNA as starting material. In both (**a**) and (**b**), PCR reactions use fixed 1μM unlabeled forward primer and 20-fold diluted labeled reverse primer. After PCR, 2.5% of product was loaded into PAGE and detected via autoradiography. TS is template-switch, and TS-PCR product is expected to be 170-nt, whereas non-TS primer extension product is 140-nt.

8a


PCR Cycle: 8 10 12 20 Dilutions
unlabeled reverse primer (µM): 0.5 0.25 0.1 0.05 0.5 0.25 0.1 0.05 0.5 0.25 0.1 0.05 0.5 0.25 0.1 0.05 0.3x 0.1x
Control plasmid PCR product (170nt)
Free labeled primer

8b


Template-Switch
Standalone Primer Ext
20 Cyc. PCR
30 Cyc. PCR
Marker
Target RNA (pg): 40 4 0.4 40 4 0.4 40 4 0.4
Target TS-PCR product (170 nt)
Target non-TS cDNA (140 nt)
lower exposure
higher exposure

III. DISCUSSION

Primer extension is a simple and useful method for determining 5' ends of RNA transcripts and mapping transcription start sites to locate functional core promoters (Krieg 1996). However, especially compared to modern sequencing techniques, this method suffers from lower sensitivity, which may preclude its use for promoters with low transcriptional activity. Thus, I adapted a radiolabeling primer extension technique to utilize template-switching and PCR for enhanced detection of RNA. Through initial setup, I observed that MMLV RTase and only some TSO designs work with template-switching (Fig 1), both of which are consistent with other studies (Zhu et al. 2001, Petalidis et al. 2003, Zajac et al. 2013). On the other hand, I have identified various optimizations for both template-switching and PCR that are not accounted for in these other reports. This also revealed inhibition of template-switching by preceding *in vitro* transcription and inhibition of PCR by preceding primer extension, both of which were resolvable. Lastly, this template-switching adaptation of primer extension accomplishes the initial goal by substantially surpassing the RNA detection ability of standalone primer extension. Altogether, these results provide new insight both on the characteristics of template-switching, and on methodological improvements for radiolabeled primer extension and other template-switching techniques.

III-I. Concentrations of dNTPs and PEG both exhibit dual effects on template-switching

Since template-switching methods depend on consistent CCC-tailing of full-length cDNAs (Zajac et al, 2013), we were curious about the relationship between dNTP availability and the innate bias of MMLV RTase towards non-templated cytosine incorporation. As a baseline, we used 1mM of all dNTPs, as seen in other protocols (Zajac et al. 2013), and compared this to settings with a relative excess of dCTP over other dNTPs (Fig 2a). Among these, most prevalent template-switching was observed in buffers with 1mM dCTP with 0.2mM

other dNTPs, and 0.6mM dCTP with 0.2mM other dNTPs (Fig 2). Compared to the default equimolar dNTP setup, these two dNTP-depleted buffers exhibited at least 20-fold higher template-switching efficiency (Fig 2c). While typical template-switch buffers use 1mM dNTPs successfully, the dNTP-depletion approach appears to be more optimal for maximizing template-switching yields, which is relevant to the purpose of our application.

The primary use of *in vitro*-synthesized RNA transcripts in this study may contribute to this apparent discrepancy, since other template-switch methods commonly focus on mRNA from cell extracts (Petalidis et. al., 2003; Schramm et. al., 2000; Zajac et. al., 2013; Zhu et. al., 2001). More specifically, the mRNAs in other studies likely contain a 5' cap structure (Schramm et. al., 2001); this feature may be critical for MMLV RTase to preferentially tail cDNAs with cytosines (Liu et al., 2018). By contrast, for our *in vitro* synthesized RNA samples, which typically lack the 5' cap (Liu et al., 2018), MMLV RTase may have a significantly lower preference for C-tailing compared to the typically reported bias. To confer stronger cytosine-incorporation bias to RTase on normally uncapped transcripts, one approach utilizes an additional enzymatic step that adds a 5' cap structure; thus, initially uncapped RNA species may still be used effectively in template-switching with one extra step (Liu et al., 2018). However, rather than adding a new enzyme and buffer to the system, our approach shows demonstrates a comparable template-switching efficiency (Fig 4). With our disproportionate dNTP ratio approach, we simply changed quantities of existing reagents, without decreasing overall product yield (Fig 2a) – therefore, with our approach, efficient template-switching may be achieved without managing additional proteins or buffers.

Surprisingly, our results further suggest that simply increasing dCTP concentration, especially past 1mM, may inhibit template-switching rather than enhancing it (Fig 2b). Even

with 4mM dCTP by itself on a similar model, C-tailing is still poor after an hour-long incubation (Ohtsubo et al., 2017). Paradoxically, when comparing the two dNTP-depletion buffers (Fig 2b), the one with 0.6mM dCTP produced more template-switching than 1mM dCTP as well. Thus, despite the obvious need to bias RTase toward cytosine incorporation, excess dCTP appears to have not only diminishing returns but possibly an inhibitory effect on template-switching. In other words, these results suggest that dNTPs availability influences template-switching with dual effects: one where depletion of dATP/dGTP/dTTP likely improves C-tailing and therefore template-switching, and another where lower overall dNTP concentrations promote more efficient template-switching through an unknown mechanism.

Similarly, we observed with negative and positive outcomes from adding PEG to the template-switching reaction. We investigated the role of PEG because one screening suggested that it increases template-switch efficiency (Fig 3b). While it indeed increased template-switch efficiency, even using 0.25% PEG (w/v) produced a visibly darker background, with intensified off-target bands (Fig 4). Even though this reagent is not typically used for reverse transcription, it has reported benefits for *in vitro* transcription due to its macromolecular crowding effects (Ge et. al., 2011). This may explain both observed effects in this study: PEG likely promotes association of the TSO with cDNA, which is otherwise an extremely weak interaction from only 3 base pairs; however, it may also promote non-specific annealing of TSO or labeled primer to the RNA template, thus increasing noise and off-target signals in gel.

III-II. Template-switching is compatible with *in vitro* transcription and PCR

For this project, one required feature of template-switching was compatibility with two other reactions: that it could be performed on *in vitro* transcription products, and that the resulting cDNA could then be efficiently amplified via PCR. By default, using

phenol:chloroform extraction to clean up *in vitro* transcription product was insufficient for template-switching (Fig 3a), as no template-switch product was observed. However, upon replacing this with Trizol extraction, template-switching activity was effectively restored to full. Additionally, the primer extension/template-switch mix quantitatively inhibited PCR, but this inhibition was negligible when the mix comprised 10% or less of the final PCR volume (Fig 7). Compared to loading the full template-switch product volume, this amounts to using only 25% for PCR, yet this exhibited a substantial net benefit to PCR yield (Fig 7b). From *in vitro* transcription, our assays suggest that rNTPs are the primary inhibitor of template-switching (Fig 3a), so the effectiveness of Trizol here is consistent with our observation that it removes rNTPs from solution. While we did not screen for a specific inhibitor of PCR, this was logistically unnecessary as the inhibition diminished with smaller input volumes. However, MMLV RTase is known to inhibit PCR through various interactions (Suslov & Steindler, 2005). As such, one alternative to loading smaller volumes is to instead heat-inactivate the RTase before adding any PCR reagents, for instance at 70°C for 10 minutes. Either way, these data indicate that template-switching can be practically integrated with both preceding *in vitro* transcription and subsequent PCR amplification.

III-III. Template-switch-PCR demonstrates much higher sensitivity towards RNA than radiolabeled primer extension by itself

Altogether, I have shown that the concept of template-switching and PCR can be applied to radiolabeled primer extension methodology in a way that faithfully represents 5' ends of mRNAs while selectively enriching the desired products. Furthermore, I show that this approach is substantially more sensitive than the standalone primer extension method, by comparing the two for small inputs of RNA. When scaling these inputs in gel analysis, the 20 cycles of TS-PCR

exhibits a 200-fold stronger signal for the desired product than standalone primer extension (Fig 8). More specifically, 0.4pg RNA with 20 cycles of TS-PCR produced a band twice as strong as that from 40pg RNA with primer extension alone (Fig 8b). Notably, the standalone primer extension does not actually produce a detectable signal here for 4pg and 0.4pg of RNA, whereas the TS-PCR does after 20 or 30 cycles. Together, this suggests that the TS-PCR can be used with effectively 200 times more sensitivity with only 20 cycles. Upon further comparison between the 30-cycle TS-PCR and standalone primer extension, the TS-PCR exhibits a 100-fold greater signal with 0.4pg than primer extension with 40pg – therefore, there is a 10,000-fold increase in sensitivity with 30 cycles (Fig 8b).

Lastly, it is important to consider the relative fraction of total products that were loaded into the gel: 25% of the TS product was used for PCR, and then approximately 2.5% of the PCR product was loaded into the gel. In contrast, 50% of the total primer extension product was loaded into the gel. Cumulatively, this results in an 80-fold difference in loading capacity, due either to dilutions or limited use of material in TS-PCR. If these losses can be reconciled with cleanup or treatment steps, the sensitivity of TS-PCR may effectively be 16,000-800,000 times greater than the standalone primer extension protocol. Therefore, these results suggest that template-switch-PCR has orders of magnitude more sensitive detection than our typical primer extension protocol, which should make this alternative method more applicable for studying 5' RNA transcript ends from very weak promoters.

III-IV. Future Directions

Although I have shown the potential of this template-switch-PCR approach through piecewise assessments, its practical usefulness would be more completely demonstrated with positive results from testing dilutions of plasmid promoter constructs or different constructs with

varying levels of transcriptional activity. Given its strong transcription, SCPX can serve as a control for easily detectable transcripts, whereas TATA-less or weak-consensus sequences can be used to demonstrate the advantage of template-switch PCR. Additionally, while this method may easily be used to map TSSs or qualitatively compare promoter strengths, it is also important to identify the dynamic PCR range over which two or more promoters can be quantitatively compared as well. For this to be quantitative, the signal amplification across this linear dynamic range must be consistent and predictable, so further tests can be done to characterize this range of PCR cycles.

Finally, while I screened multiple factors that affect template-switching efficiency, it may also be helpful to understand the mechanisms of these effects in order to further improve the method, or to contribute to knowledge of template-switching in general. For example, it may be useful to measure not only the number of template-switch products, but also the sequences and lengths of the N-tailing on 3' cDNA ends by MMLV RTase. To do this, we can use TSOs with degenerate anchoring nucleotides – in other words, as opposed to the typical 3'–rGrGrG (designed to bind to predicted CCC-tails on cDNA ends), the oligos would have randomized nucleotides in this region (Zajac et al, 2013). These oligos, combined with qPCR and quantitative sequencing, have already been used to study the preferences of N-tailing by MMLV RTase in equimolar 1mM dNTP conditions (Zajac et al, 2013); therefore, this can easily be adapted to explicitly assay tailing preferences for our dNTP conditions as well (e.g. 0.6mM dCTP with 0.2mM other dNTPs).

IV. Materials and Methods

Preparation of labeled reverse primer

Primer labeling equipment setup is as follows: prepare 40mL of column wash buffer: TE' with 50mM NaCl. Also thaw γ^{32}P-ATP in a water-filled beaker behind a shield. Assemble small empty 1mL syringe tube (no needle or plunger) or column, add 0.5x0.5cm filter membrane into bottom tip, then fill to 0.9mL with G50 Fine Sephadex. Rinse beads with 3-5 column volumes of wash buffer, then drain to bead surface level. Primer labeling (kinase) reaction was comprised of: 5µL of 2.5µM unlabeled reverse primer in TE' plus 13.5µL H_2O in a 1.5mL Eppendorf tube. Then incubate for 2 mins. at 58°C, and spin down contents. Add 2.5µL 10X PNK buffer, 2µL T4 PNK enzyme, and 2µL γ^{32}P-ATP. Incubate at 37°C for 1h. Stop reaction with 2µL 200mM EDTA. Column purification was as follows: prepare 11 collection tubes. For fraction #1, load 25µL reaction mix into column, rinse tube with 50µL wash buffer and load this as well. Add 150µL more wash buffer and continue collecting in fraction #1. For fractions #2-11, move column into a new fraction tube and collect after adding 50µL wash buffer. Fraction selection was as follows: Measure and record radioactivity of all fractions with 5K-Geiger counter and take 2 samples representing the first observed radioactivity peak, approximately 100 µL total. Store at -20°C.

Pre-mixed reagents for in vitro transcription

0.1M HEMG mix contained final concentrations of 25mM HEPES, 100mM KCl, 12.5mM $MgCl_2$, 0.1mM EDTA, 10% glycerol, 1mM DTT, and 0.01% NP-40. 5% PEG/PvOH was prepared from 1g PvOH with 1g PEG and 10mL water to achieve a 20mL final volume. Stop mix for *in vitro* transcription was made with final concentrations of 20mM EDTA, 200mM NaCl, 1% SDS, and 0.3mg/mL glycogen.

Pre-mixed reagents for Primer Extension and Template-Switching

5X Annealing Buffer contained final concentrations of 10mM Tris pH 7.8, 1mM EDTA, and 1.25M KCl. Labeled annealing mix was made with 20µL of 5X Annealing Buffer with 57µL H_2O and 3µL of stock ^{32}P-labeled primer. Unlabeled annealing mix was made with 20µL of 5X Annealing Buffer with 58µL H_2O and 2µL 5µM unlabeled primer. FLB (formamide loading buffer) was contained 80% formamide, 10mM EDTA, 1mg/mL xylene cyanol, and 1mg/mL bromophenol blue. 4X Extension Buffer (4X EB) was prepared in a 100µL volume with final concentrations of: 400mM betaine, 5mM $MnCl_2$, 250mM Tris pH 8.3, 50mM DTT, various dNTP concentrations, and 6U/µL reverse transcriptase (typically MMLV RTase). Equimolar dNTP variants contained 4mM of each dNTP in this 4X buffer. The finalized 3:1 dNTP variant (Fig 3-8) of the 4X Extension Buffer contained 2.4mM dCTP and 0.8mM each of dATP, dGTP, and dTTP. 2X Primer Extension mix was made by combining 40µL 4X EB with 40µL H_2O (Fig 1-2) or with 32µL H_2O and 8µL 5% PEG/PvOH (Fig 3-8) for a final volume of 80µL. 2X Template-Switch mix was made by combining 30µL 4X EB with 30µL H_2O (Fig 1-2) or with 12µL H_2O, 12µL 5% PEG/PvOH, and 6µL 20µM TSO (Fig 3-8) for a final volume of 60µL; note that TSO1 was used for Fig 1, TSO1/2/3/4 for Fig 2, and then strictly TSO3 for Fig 3-8. FLB (formamide loading buffer) was contained 80% formamide, 10mM EDTA, 1mg/mL xylene cyanol, and 1mg/mL bromophenol blue. Final FLB + NaOH mix used for loading was prepared by combining 120µL FLB with 60µL 0.1M NaOH.

Pre-mixed reagents for PCR

Unlabeled PCR Master Mix was comprised of 5.5µL each of 100µM unlabeled forward and reverse primers, 44µL H_2O, and 275µL of 2X GoTaq Mix; forward primer corresponds to TSO-templated sequence and reverse primer is identical to radiolabeled primer used in radiolabeled primer extension. Labeled PCR Master Mix contained 60µL H_2O, 30µL 10µM

unlabeled forward primer, 15µL 5µM unlabeled reverse primer, 15µL labeled reverse primer from stock, and 150µL of 2X GoTaq Mix.

In vitro transcription

All in vitro transcription reactions were assembled from a master mix, where each reaction is comprised of 0.5µL 1M HEPES pH 7.6, 2.5µL mix with 5mM of each rNTP, 5µL 5% PEG/PvOH (w/v), 3.5µL H_2O, 6.5µL HEMG in 0.1M KCl, 6µL *Drosophila* SK nuclear extract, and 1µL of 200ng/uL of DNA plasmid construct. This totals to 25µL. After 45 min. at 30°C, the reaction was stopped by adding 100µL Stop Mix and 5µL Proteinase K and incubating for 15min. at 30°C. This was followed by phenol-chloroform extraction and ethanol precipitation to pellet the products for primer extension.

Trizol Extraction

Trizol extraction was performed on some *in vitro* transcription products or reagents by adding 1mL Trizol reagent to the 100-150µL *in vitro* transcription mix. Vortex for 5 mins. Add 200µL chloroform, vortex briefly, and leave at room temperature for 5 mins. Centrifuge at 12,000xg for 15 minutes at 4°C. Transfer only the top clear layer, approximately 600µL, to a new clean 1.5mL tube. Add 60µL 3M sodium acetate, 1µL 5mg/mL glycogen, 700µL isopropanol, and invert several times. After 20 more minutes at room temp., centrifuge again at 12,000xg for 15 minutes at 4°C. Remove all supernatant, then 75% EtOH wash twice and air-dry pellet.

Primer Extension and Template-Switching

For primer extension or template-switching experiments that did not utilize PCR, 2ng or less of Target RNA in 2µL was combined with 8µL labeled annealing mix, or *in vitro* transcription product pellet was resuspended in 2µL H_2O plus 8µL labeled annealing mix. For

experiments that also would follow up with labeled PCR, the same volumes were used with unlabeled annealing mix instead. Incubate at 75°C for 90 sec., then 58°C for 40 minutes.
For standalone primer extension without template-switching, 10µL 2X Primer Extension Mix was then added to each sample. For template-switching reactions, 10µL 2X Template-Switch Mix was then added to each sample. Incubate at 37°C for 1 hour, then at 75°C for 10 mins. For samples that had no subsequent PCR step, ethanol precipitation was instead used.

Ethanol Precipitation after Primer Extension or Template-Switching

Note that liquid waste is radioactive if labeled annealing mix was used. Ethanol precipitation was performed after Primer Extension or Template-Switching by adding 300µL 100% EtOH, 5µL 3M NaOAc, and 1.5µL 5mg/mL glycogen; samples were then mixed by inversion and kept at -20°C for at least 30 mins. The samples were then centrifuged at 4°C for 15 mins. at 12,000xg, and all supernatant was removed. The samples were then washed with 300µL 75% EtOH and dried in a lyophilized at room temperature for 5 mins. Finally, each pellet was resuspended with 9uL of FLB+NaOH mix, then tubes were placed on 95°C dry heating plate for 3mins and transferred to an ice block before loading into gel.

Unlabeled and Labeled PCR

PCR cycling protocol was as programmed as follows: 2 mins. 95°C, then 30 rounds of (30s 95°C, 30s 63°C, 20s 72°C), ending with 4°C unlimited holding. Unlabeled PCR reactions used variable volumes of unlabeled template-switch product were combined with PCR Master Mix and partitioned into final volumes of 50µL. Alternatively, 1-2ng of PCR control plasmid (FTS) was occasionally used in place of template-switch product. To use 40% of PCR volume with template-switch input, 20µL of template-switch product was combined with. To achieve the established 10% v/v input of template-switch product, we used 5µL (out of 20µL total) of

unprecipitated template with 45µL of labeled or unlabeled PCR mix. If unlabeled, the 5µL of each product was loaded into a 6% acrylamide gel and ran at 120V for 45 minutes, then was stained with EtBr. If labeled, then 2.5µL PCR product was added to 7.5µL FLB+NaOH mix.

PAGE for labeled samples

For labeled samples, from primer extension, template-switching, or PCR, 4uL of resuspended product with FLB was loaded into 8% acrylamide gel on large sequencing plates and ran for 2h at 35W. Plates were then separated, and gel was transferred to 3M paper and vacuum-dried with heat. After 24 hours of screen exposure, gels were visualized via autoradiography.

Appendix

TSO #	Sequence (5' to 3')
1	AGTCGATCATGACGATGCTAGTCAGTCrGrGrG
2	GCAGAAACCACTGAGAATATCATACGArGrGrG
3	GCATCGATCATACGATGCTAGTCAGTCrGrGrG
4	AAGCAGTGGTATCAACGCAGAGTACATrGrGrG

Figure A1. TSO designs by number and sequence. Each TSO consists of a variable 27-nt ssDNA sequence from the 5' end and all possess a triple riboguanosine stretch (rGrGrG) at the 3' end. Corresponding forward PCR primers are identical to their respective TSOs except with only the 27-nt ssDNA sequence as described here. TSO1 is featured in Fig 1, all four are featured in Fig 2, and only TSO3 is featured in Fig 3-8.

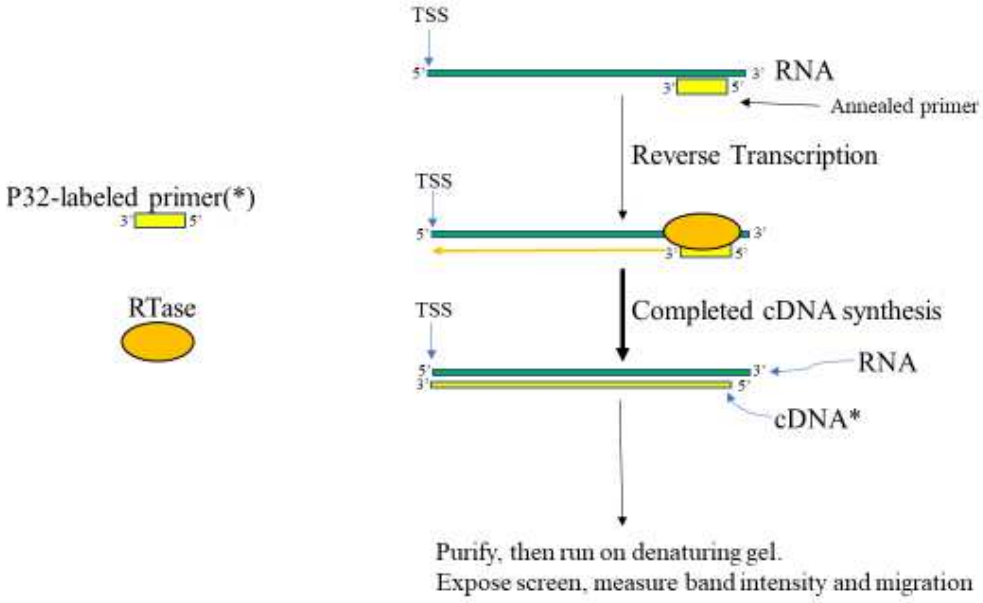


Figure A2. In primer extension, P32-labeled primer is annealed to target RNA, then elongated by reverse transcriptase (RTase). For complete synthesis of cDNA, RTase must transcribe up to the 5' end of RNA template. Both 5' end of RNA template and 3' end of cDNA correspond to transcription start site (TSS) of original promoter sequence. Purified cDNA is analyzed by gel electrophoresis and autoradiography; observed size or sequence can be used to determine TSS within ± 1 nucleotide. Asterisks denote radioactive strands.

REFERENCES

Carey, M.F., Peterson, C.L., & Smale S.T. (2013). *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*. Cold Spring Harbor Laboratory Press.

Ge, X., Luo, D., & Xu, J. (2011). Cell-Free Protein Expression under Macromolecular Crowding Conditions. *PLoS ONE*, *6*(12), e28707. https://doi.org/10.1371/journal.pone.0028707

Juven-Gershon, T., Cheng, S., & Kadonaga, J. T. (2006). Rational design of a super core promoter that enhances gene expression. *Nature Methods*, *3*(11), 917–922.

Krieg, P. A. (1996). *A Laboratory Guide to RNA: Isolation, Analysis, and Synthesis.* Wiley-Liss, Inc., New York, pp. 133-139.

Leamy, K. A., Assmann, S. M., Mathews, D. H., & Bevilacqua, P. C. (2016). Bridging the gap between in vitro and in vivo RNA folding. *Quarterly reviews of biophysics* **49**: e10.

Leibovich, L., & Yakhini, Z. (2012). Efficient motif search in ranked lists and applications to variable gap motifs. *Nucleic acids research*, *40*(13), 5832-47.

Liu, F., Zheng, K., Chen, H. C., & Liu, Z. F. (2018). Capping-RACE: a simple, accurate, and sensitive 5' RACE method for use in prokaryotes. *Nucleic acids research* **46:** e129.

Ohtsubo, Y., Nagata, Y., & Tsuda, M. (2017). Efficient N-tailing of blunt DNA ends by Moloney murine leukemia virus reverse transcriptase. *Scientific Reports*, *7*(1), 41769. https://doi.org/10.1038/srep41769

Ouhammouch, M., & Brody, E. N. (1992). Temperature-dependent template switching during in vitro cDNA synthesis by the AMV-reverse transcriptase. *Nucleic Acids Research*, *20*(20), 5443–5450. https://doi.org/10.1093/nar/20.20.5443

Petalidis, L., Bhattacharyya, S., Morris, G. A., Collins, V. P., Freeman, T. C., & Lyons, P. A. (2003). Global amplification of mRNA by template-switching PCR: linearity and application to microarray analysis. *Nucleic acids research* **31**: e142.

Rio D.C., Ares, M., Hannon, G.J., Nilsen, T.W. (2010). *Purification of RNA Using TRIzol (TRI Reagent).* Cold Spring Harbor Laboratory Press.

Roy, A. L., & Singer, D. S. (2015). Core promoters in transcription: old problem, new insights. *Trends in biochemical sciences* **40**: 165-71.

Schramm, G., Bruchhaus, I., & Roeder, T. (2000). A simple and reliable 5'-RACE approach. *Nucleic acids research* **28**: E96.

Suslov, O., & Steindler, D. A. (2005). PCR inhibition by reverse transcriptase leads to an overestimation of amplification efficiency. *Nucleic Acids Research*, *33*(20), e181. https://doi.org/10.1093/nar/gni176

Vo Ngoc, L., Wang, Y. L., Kassavetis, G. A., & Kadonaga, J. T. (2017). The punctilious RNA polymerase II core promoter. *Genes & development* **31**: 1289-1301.

Yang, M. Q., & Elnitski, L. L. (2008). Diversity of core promoter elements comprising human bidirectional promoters. *BMC genomics*, *9 Suppl 2*(Suppl 2), S3. doi:10.1186/1471-2164-9-S2-S3

Yang, M. Q., & Elnitski, L. L. (2008). Diversity of core promoter elements comprising human bidirectional promoters. *BMC genomics*, *9 Suppl 2*(Suppl 2), S3. doi:10.1186/1471-2164-9-S2-S3

Zajac P., Islam S., Hochgerner H., Lönnerberg P., Linnarsson S. (2013). Base Preferences in Non-Templated Nucleotide Incorporation by MMLV-Derived Reverse Transcriptases. *PLoS ONE* **8**: e85270. https://doi.org/10.1371/journal.pone.0085270

Zhu Y.Y., Machleder E.M., Chenchik A., Li R., Siebert P.D. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *BioTechniques* **30**: 892-897. PubMed: 11314272.