

METHODOLOGY ARTICLE

Open Access

Incorporation of non-natural nucleotides into template-switching oligonucleotides reduces background and improves cDNA synthesis from very small RNA samples

Jeremy Kapteyn¹, Ruifeng He^{1,2}, Eric T McDowell¹, David R Gang^{1,2*}

Abstract

Background: The template switching PCR (TS-PCR) method of cDNA synthesis represents one of the most straightforward approaches to generating full length cDNA for sequencing efforts. However, when applied to very small RNA samples, such as those obtained from tens or hundreds of cells, this approach leads to high background and low cDNA yield due to concatamerization of the TS oligo.

Results: In this study, we describe the application of nucleotide isomers that form non-standard base pairs in the template switching oligo to prevent background cDNA synthesis. When such bases are added to the 5' end of the template switching (TS) oligo, they inhibit MMLV-RT from extending the cDNA beyond the TS oligo, thus increasing cDNA yield by reducing formation of concatamers of the TS oligo that are the source of significant background.

Conclusions: Our results demonstrate that this novel approach for cDNA synthesis has valuable utility for application of ultra-high throughput technologies, such as whole transcriptome sequencing using 454 technology, to very small biological samples comprised of tens of cells as might be obtained via approaches like laser microdissection.

Background

Methods such as laser (capture) microdissection [1] and cell sorting are becoming increasingly accessible, enabling more researchers to address biological questions at the level of individual cell types. As the sensitivity of analytical tools such as mass spectrometry increases, very small samples become increasingly amenable to systems-level global analyses including proteomics and metabolomics investigations. At the same time, continually decreasing costs of next-generation sequencing technologies, combined with the added qualitative information value of sequence data [2-5] as compared to microarrays, could make transcriptome sequencing the preferred means of evaluating questions regarding gene expression and alternative splicing for

model organisms, and *de novo* sequencing continues to represent the best approach for non-model organisms where genomic sequence data is lacking. The ability to conduct omics level experiments using only a few cells will allow researchers to answer questions that were previously intractable due to the heterogeneous and complex nature of most tissue samples.

Despite the growing availability of next-generation sequencing technologies, significant effort is still often needed to collect the required biological starting material. For transcriptome analysis, this generally entails the synthesis and subsequent amplification of cDNA. Although exponential amplification raises concerns about representation bias, exponentially amplified cDNA is representative of the transcript population, even when starting with the equivalent of a single cell's RNA [6,7]. Moreover, full-length transcripts are highly desirable in transcriptome sequencing experiments. Several methods have been developed to preferentially enrich cDNA for

* Correspondence: gangd@wsu.edu

¹School of Plant Sciences and BIO5 Institute, The University of Arizona, 1657 E Helen Street, Tucson, AZ 85721, USA

full length transcripts using the 5' cap structure of mature mRNA transcripts [8-12]. Among these is the template switching PCR (TS-PCR) method [12], which takes advantage of the terminal transferase activity and template-switching ability of Moloney murine leukemia virus reverse transcriptase (MMLV-RT) to add an arbitrary sequence to the 5' end of a transcript in a manner that occurs preferentially for capped, full-length transcripts [12-16]. This arbitrary sequence, encoded in a TS oligo, along with a second arbitrary sequence added at the 3' end of the cDNA by the oligo-dT-containing primer used to initiate reverse transcription, is used to subsequently amplify the total pool of putatively full-length transcripts in a completely sequence-naïve manner. For investigators interested in transcriptome analysis and gene discovery from non-model organisms or those interested in transcriptional processes at a cell-specific level, TS-PCR cDNA synthesis represents the most straightforward approach to generating cDNA for sequencing efforts. However, this approach does suffer from some drawbacks, including the potential for high background that interferes with downstream sequencing efficiencies when starting with very small quantities of RNA.

In this study, we describe a novel approach to eliminate background and increase cDNA yield in TS-PCR. The key feature of our modification that enables this improvement of the TS-PCR method is the inclusion of isomeric nucleotide bases at the 5' end of the TS oligo to inhibit MMLV-RT template switching activity after incorporation of the first TS oligo and thereby reduce or prevent TS oligo concatamerization (see Figure 1). We applied the modified method to construct cDNA libraries from glandular trichome secretory cells of tomato (*Solanum*) species and petunia, as well as from rhizome tips and elongation zones of scouring rush (*Equisetum hyemale*) and red rice (*Oryza longistaminata*). Our results demonstrate that application of the improved TS-PCR method dramatically reduced the proportion of oligo concatamers in the resulting sequence data, and the cDNAs produced were of high quality and yield, useful for both conventional Sanger sequencing as well as next-generation sequencing.

Results

Based on results presented below, we developed the following model that outlines how cDNA synthesis occurs in the absence and presence of RNA using standard TS oligo vs. iso₃TS oligo (Figure 1). When a total RNA sample including mRNA transcripts is combined with oligo-dT primer sequence to prime first strand synthesis in the presence of TS oligo, the following can occur. **Step 1** - Oligo-dT primer anneals to the polyA tail of an mRNA transcript (Figure 1B) and reverse

transcriptase (RT) synthesizes first strand cDNA, or to some other polynucleotide entity in solution (Figure 1A), such as a standard TS oligo. Upon reaching the 5' end of the transcript (or other nucleotide entity), the terminal transferase activity of MMLV RT adds several additional bases, preferentially dC, to the 3' end of the cDNA strand. The terminal transferase activity of MMLV-RT is greatly enhanced in the presence of Mn²⁺ [17]. **Step 2** - The rG₃ bases at the 3' end of a standard TS oligo anneal to the polyC bases left by the RT at the 3' end of the cDNA strand and RT switches templates, transcribing through the TS oligo DNA template (Figure 1A and 1B). The TS oligo sequence and the extended 5' portion of the oligo-dT primer are used as primer sites for cDNA amplification in the LD PCR step which follows the RT reaction. **Step 3** - The terminal transferase activity of RT can add polyC bases to the end of a cDNA strand when using the standard TS oligo DNA as a template (Figure 1A and 1B). **4** - Another TS oligo can then anneal and the process of TS oligo transcription and addition can continue indefinitely, adding a series of direct repeats of TS oligo sequence to the growing cDNA strand. **Alternative Step 3** (iso₃TS oligo is used) - Upon reaching the poly iso-nucleotide bases at the 5' end of the iso₃TS oligo, RT is inhibited from incorporating the corresponding nucleotides and reverse transcription terminates without addition of a polyC tail (Figure 1C). **Alternative Step 4** - No additional iso₃TS oligo can anneal and concatamerization of TS oligo does not occur (Figure 1C).

We developed this model as the result of performing experiments where we used TS-PCR to construct cell-type specific cDNA libraries from small numbers of isolated glandular trichome secretory cells from tomato and related *Solanum* species. These species contain several different types of glandular trichomes, some of which can only be isolated from other types by hand-selecting them from the surfaces of fresh and intact plant leaves using microcapillary probes under the guidance of a stereoscope, a time consuming and tedious procedure that yields only small numbers of isolated trichomes (100 and 500 gland samples were evaluated, see Figure 2). While constructing libraries from these very small samples, we encountered a significant flaw in the TS-PCR method as originally described [12]: a substantial quantity of cDNA spanning a size range that might be expected to represent a typical biological sample is synthesized in the absence of any starting RNA (Figure 2A, lane 7; 2B, lane 1; 2C, lane 1; see Figure 1A). Differences in level of cDNA synthesis for different quantities of starting material are difficult to distinguish after TS-PCR (Figure 2A, lanes 1, 4 and 7). Substantial cDNA is generated in TS-PCR step even for negative control (Figure 2A, lane 7). However, after *Sfi*I digestion, the

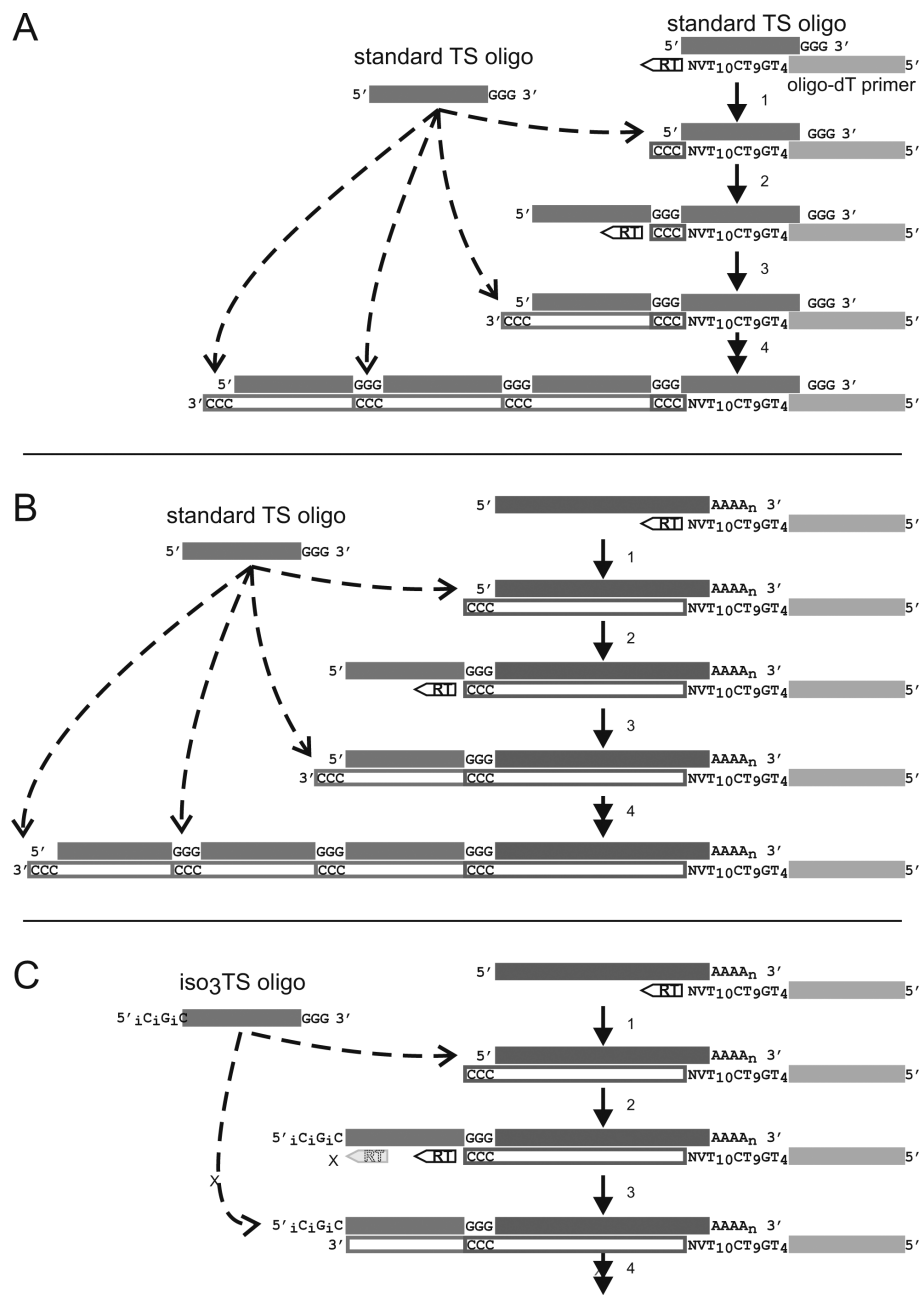
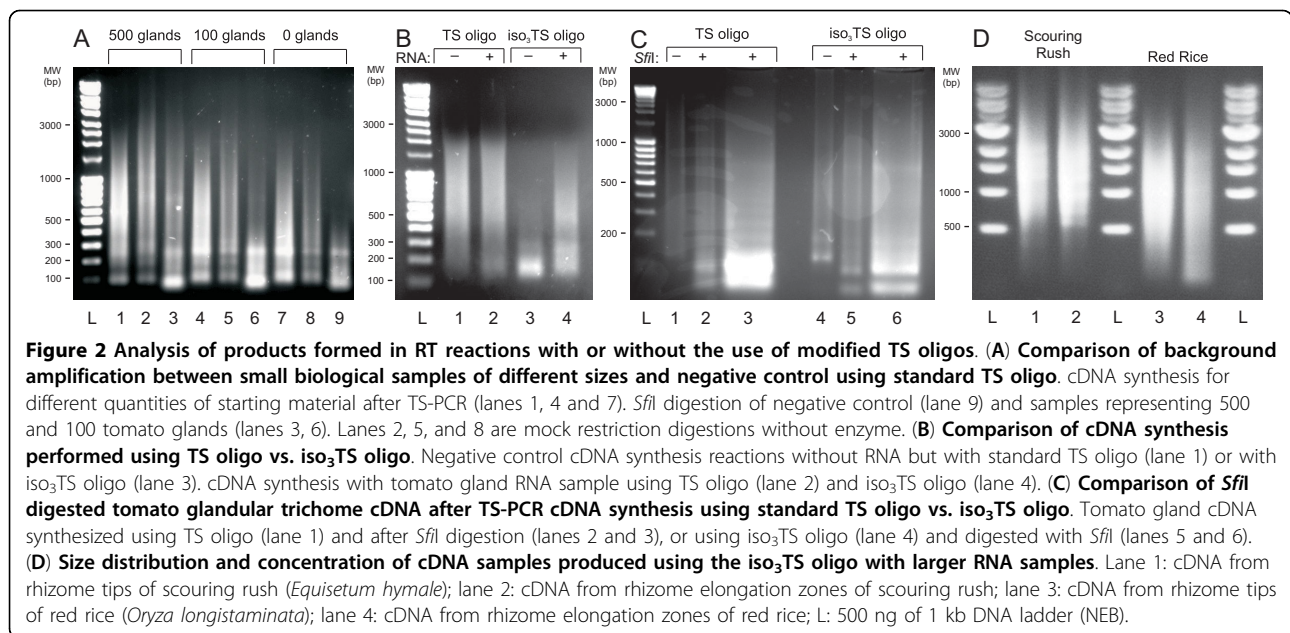


Figure 1 Model of cDNA synthesis that occurs in the absence (A) and presence (B and C) of RNA using standard TS oligo (A and B) vs. iso₃TS oligo (C).

cDNA in the negative control is digested into low MW fragments (Figure 2A, lane 9), while a smear of higher MW DNA is present for samples representing 500 and 100 tomato glandular trichomes ("glands") (Figure 2A, lanes 3, 6). Lanes 2, 5, and 8 are mock restriction digestions without enzyme added.

Negative control cDNA synthesis reactions performed without RNA show high levels of cDNA synthesis for TS oligo as large as 2000 bp (Figure 2B, lane 1), while

the negative control cDNA synthesis performed using iso₃TS oligo demonstrates almost complete suppression of artefactual synthesis of cDNA above 200 bp (Figure 2B, lane 3). It is difficult to distinguish the results of cDNA synthesis for the negative control vs. tomato gland RNA sample using TS oligo (Figure 2B, lane 1 vs. 2), while the results of cDNA synthesis using iso₃TS oligo for the negative control (Figure 2B, lane 3) and the tomato gland RNA sample (Figure 2B, lane 4) show



strong dependence on the presence of sample RNA to generate cDNA.

After preparative DNA electrophoresis, tomato gland cDNA synthesized using TS oligo (Figure 2C, lane 1) shows a stronger shift toward lower molecular weight fragments after *SfiI* digestion (Figure 2C, lanes 2 and 3) than the cDNA synthesized from the same RNA sample amplified using iso₃TS oligo (Figure 2C, lane 4) and digested in parallel with *SfiI* (Figure 2C, lanes 5 and 6). Lanes 3 and 6 were large preparative lanes used for subsequent cDNA library production as described in the Methods. Equal quantities of DNA were loaded in lanes 1, 2, 4, and 5 and in 3 and 6, but the proportion of the DNA sample comprising lower molecular weight fragments as evidenced by ethidium bromide staining is much higher after *SfiI* digestion for samples synthesized using the unmodified TS oligo (Figure 2C, lanes 2 and 3) as compared to that for cDNA synthesized using the modified, iso₃TS oligo (Figure 2C, lanes 5 and 6). This suggests that a higher proportion of cDNA synthesized with the unmodified TS oligo consists of concatamers of the TS oligo.

To characterize the sequences of the apparently artefactual cDNA generated using TS oligo in first strand cDNA synthesis reactions, performed in the absence of RNA or using very small quantities of sample RNA, we cloned amplified and undigested cDNA into pCR2.1-TOPO and sequenced several cDNA inserts, revealing them to be concatamers of direct repeats of the TS oligo (Figure 3A). This concatamerization of TS oligo occurs in cDNA syntheses from biological samples containing very small amounts of RNA (Figure 2A, lanes

4-6), but appears to be reduced as the size of the biological sample increases (Figure 2A, lanes 1-3). In samples containing either no RNA or very low numbers of mRNAs to serve as template for reverse transcription, RT activity is spuriously primed via primer-primer annealing (Figure 1A). Although the efficiency of MMLV-RT terminal transferase activity toward uncapped RNA templates under the conditions used has been shown to be much lower than for capped transcripts (activity with DNA templates was not examined) [17], the low abundance of transcript templates increases the likelihood of terminal transferase addition of dC residues to the growing cDNA strand at the ends of any available template, including non-capped templates. After MMLV-RT has switched templates to the DNA TS oligo, the process is repeated and TS oligo is added indeterminately to the 3' end of the growing cDNA strand (see Figures 1A/B and 3).

This background cDNA synthesis in the absence of RNA or in the presence of low amounts of RNA and approaches to eliminate it were not mentioned in the original or subsequent publications on the TS method [12]. The original TS patent (U.S. Pat. No. 5,962,272) mentions a background problem, but does not describe or speculate as to its cause. However, this problem presents a serious obstacle to effective construction of cDNA libraries representing biological transcripts rather than synthesis artefacts, especially for very small RNA samples originating from just a few cells. One approach to circumvent this problem is to incorporate restriction endonuclease sites into the oligos used during cDNA synthesis. This facilitates ligation into cloning vectors

Sequence example A

(20 repeats of TS oligo)

>02_ssn3_T7_171962

```
GCNNTATCATGCTCGAGCGGCCAGTGTGATGGATATCTGCAGAATTCGCCCTT AAGCAGTGNATC
AAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAG
TACGCGGG AAGCAGTGGTAACAACGCAGAG TACGCGGG AAGCAGTGGTAACAACGCAGAGTACGCGGG AA
GCAGTGGTAACAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAGTACGCGGG AAGCAGTGGTAT
CAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAG
TACGCGGG AAGCAGTGGTATCAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAGTACGCGGG AA
GCAGTGGTAACAACGCAGAGTACGCGGG GAGCAGTGGTATCAACGCAGAGTACGCGGG AAGCAGTGGTAA
CAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAG
TACGCGGG AAGCAGTGGTATCAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAGTACGCGGG AA
GCAGTGGTAACAACGCAGAGTACGCGGG AGATGACTTGG GCAAAAAAAAAAAGAAAAAAAAACAAACATG
TCGGCCGCCTCGGCCCTCTA
```

Sequence example B

(9 repeats of TS oligo attached to tomato sequence)

>70_f6_T7_172122

```
GCNTNANCTGCTCGAGCGGCCAGTGTGATGGATATCTGCAGAATTCGCCCTT AAGCGTGTNTCNACG
CAGAGTACGCGGG AAGCAGTGGTATCAACGCAGAGTACGCGGG AAGCAGTGGTATCAAC GCAGAGTACGC
GGGAAGCAGTGGTATCAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAGTACGCGGG AAGCAGT
GGTAACAACGCAGAGTACGCGGG AAGCAGTGGTAACAACGCAGAGTACGCGGG AAGCAGTGGTAACAACG
CAGAGTACGCGGG AAGCAGTGGTATCAACGCAGAGTACGCGGG CGCGAGAAACTTTTGGGAACCGGGGCA
CGATCAAATGCTTACACAGATTTTCATCATACTGT GTTCCACATCCATTACCCACTAGCACAAAGGGCT
CAGAAGGCGATTGTCTCCAGTGGTCTTAGATGCTCTGAATGAGATCGCTTCCATCCTAAGTTTCTCGC
TTCCGCGAGTTG AGAAAAAAAAAAAAAAAAAAAAAAAAAAC
```

Figure 3 Examples of sequencing results for cDNA clones having concatamers of TS oligo. Sequence example A reflects a cDNA clone with an insert that appears to be comprised entirely of a concatamer of 20 direct repeats of TS oligo (highlighted in alternate shades of gray) associated with a single oligo-dT reverse primer (highlighted in turquoise). The 11 intervening bases are of unknown origin. See Figure 1A for a diagrammatic representation of the process whereby this type of concatamer may arise. Sequence example B shows 9 repeats of TS oligo concatamerized to a 178 base sequence with homology to nucleotide sequences from *Vitis vinifera* and *Arabidopsis thaliana* potentially representing a plant metalloendopeptidase. The oligo-dT reverse primer is also shown. See Figure 1B for a representation of the mechanism by which this type of concatameric sequence would be generated.

for library construction and reduces concatamerization artefact contamination of the final cDNA sample. However, cDNA size fractionation does not eliminate all of the small cleaved fragments and they contaminate the final cDNA library, resulting in a lower proportion of transcript-based cDNAs available for ligation into a cloning vector (Figure 2A, lanes 3,6). Moreover, overall yield of high quality cDNA representing biological transcripts is also lower when the reverse transcriptase spends a significant proportion of its time synthesizing concatamers as opposed to unique cDNAs.

We have developed a unique method to eliminate this concatamer-generated background problem that takes advantage of the expanded genetic alphabet [18,19] to prevent MMLV-RT from reaching the 5' end of the TS oligo template after the initial template switch, thus preventing terminal transferase-mediated addition of dC bases at the end of the sequence complementary to the TS oligo template and subsequent concatamerization of TS oligo sequences (Figure 1C). Stereoisomers of deoxycytosine and deoxyguanosine, referred to as isocytosine (iso-C) and isoguanosine (iso-G), have the properties of natural nucleotide bases but have altered hydrogen bonding abilities which cause them to form nonstandard

base pairs specifically with each other and discriminate against their naturally occurring analogs [17,19]. Until now, application of this expanded alphabet has been extended only to hybridization-based molecular diagnostic assays [20].

We designed a modified TS oligo (called iso₃TS, see Materials and Methods) that incorporated iso-C and iso-G bases at the 5' end of the oligo to test their ability to inhibit the MMLV-RT terminal transferase activity in a first strand synthesis mix that lacked the complementary iso-dNTPs. We designed a series of three isomeric nucleotide bases to discourage any potential infidelity that MMLV-RT might exhibit. The performance of the iso₃TS oligo in TS-PCR cDNA synthesis was tested in parallel with an identical TS oligo lacking the isomeric bases to determine the effect of the modified oligo on the production of artefactual cDNA in both negative control cDNA synthesis having no RNA present in the reverse transcription step as well as a test RNA sample isolated from a small number of tomato glandular trichomes. As shown in Figure 2B, performing cDNA synthesis using the iso₃TS oligo greatly reduces or eliminates artefactual cDNA that resulted from concatamer formation and amplification in the cDNA synthesis, as

occurred when using the unmodified TS oligo either with a small quantity of RNA template or without RNA template. It was difficult to distinguish between the intensity or size range of artefactual cDNA synthesized in the latter two cases, suggesting that a large proportion of the cDNA synthesized for the biological sample may consist of artefactual cDNAs while a smaller proportion of the cDNA population represents actual transcripts. This conclusion was supported by the results of subsequent *Sfi*I digestion of the cDNA samples in preparation for their ligation into the library cloning vector (Figure 2C), where cDNA samples prepared using unmodified oligo showed a strong shift toward lower molecular weight fragments after restriction digestion, a shift that was not evident for cDNA synthesized using iso₃TS oligo.

Evidence of the efficacy of the iso₃TS oligo can be observed in comparisons of several of the EST libraries we have recently made and sequenced using 454 technology. For instance, in the *Salvia divinorum*, *Solanum lycopersicum*, and *Solanum habrochaites* glandular trichome libraries, we found that 20.6%, 17.9%, and 11.6% of the reads, respectively, contained 2 or more concatamers of the SMART IV-derived oligo (TS oligo). Many of these reads contained multiple concatamers as illustrated in Figure 3. This number of concatamers is dramatically reduced or eliminated via the use of the iso₃TS oligo as seen in the *Equisetum hyemale*, *Solanum habrochaites* and *Isatis tinctoria* 454 libraries which possess 0.02%, 1.5%, 3.1% of reads possessing 2 or more concatamers of the SMART IV-derived oligo (TS oligo) (Table 1), with the vast majority of these possessing only 2 concatamers. In all of these cases, the same method for cDNA synthesis, 454 library construction and sequencing was used, and involved the same amount of template RNA. With the modified method, we also produced high quality cDNA from rhizome tips and elongation zones of scouring rush and red rice that was suitable for next generation sequencing using the

454 GS-FLX instrument with Titanium technology (Figure 2D).

Discussion

The use of oligos containing isomeric nucleotide bases to inhibit the enzymatic activity of MMLV-RT at a discrete point in reverse transcription of a template represents a novel application of the expanded genetic alphabet. The use of this approach to modify standard TS-PCR cDNA synthesis methods permits elimination of TS oligo concatamer formation, dramatically changing the composition of the first strand cDNA population used as template for the LD PCR step in TS-PCR based cDNA synthesis by removing a huge proportion of sequence comprised only of direct repeats of the TS oligo. Amplification of these concatamers in the PCR step of the TS-PCR generates a substantial quantity of artefactual sequence in the amplified population of cDNA. Upon digestion, TS oligo fragments can potentially represent a vast molar majority of the DNA fragments in the cDNA population, making downstream cloning and sequencing inefficient and more costly. Reduction or elimination of TS concatamer formation via use of the iso₃TS oligo and the resultant increase in the efficiency of cDNA synthesis from the biological sample itself increases the signal to noise ratio and restores the biological integrity of cDNA samples synthesized from very limited samples.

Performing first strand cDNA synthesis using the iso₃TS oligo while providing only the four naturally occurring nucleotide bases apparently causes MMLV-RT to prematurely stall before reaching the 5' end of the iso₃TS oligo template (see Figure 1). The great reduction of concatamer formation in these reactions suggested that misincorporation of natural bases by MMLV-RT opposite the synthetic bases was inhibited and that MMLV-RT terminal transferase activity was reduced, although we did not measure actual terminal transferase activity. However, we believe that our data only support a

Table 1 Comparison of sequencing data from different cDNA libraries generated by standard TS oligo vs. iso₃TS oligo

cDNA synthesis method	Species	Avg length of read (nt)	Num. 454 reads (Mb)	Num. 2+ SMART oligos (concatamers)	% of concatamers
Standard TS oligo	<i>Salvia divinorum</i>	197	342263 (77.6)	70549	20.6
	<i>Solanum lycopersicum</i>	207	336745 (70.0)	60173	17.9
	<i>Solanum habrochaites</i>	165	329070 (54.4)	38231	11.6
Modified TS oligo (iso ₃ TS oligo)	<i>Solanum habrochaites</i>	199	344190 (68.6)	5032	1.5
	<i>Isatis tinctoria</i>	345	970893 (335.9)*	29734	3.1
	<i>Equisetum hyemale</i>	324	871579 (282.2)*	159	0.02

* data from 454 Titanium sequencing run

mechanism for concatamer formation that involves the terminal transferase activity of MMLV-RT. Because the “monomers” are all added in the same orientation, a mechanism involving addition of primer-dimers is precluded. The only other potential mechanism for concatamer formation might involve blunt ligation of single stranded TS oligos to the end of the template followed by replication by MMLV-RT. However, no enzyme that could perform such a ligation was present in the assays. Moreover, the elimination or dramatic reduction in concatamer formation that was observed when the iso₃TS oligo was used but not when the standard TS oligo was used fully supports our proposed mechanism. The fact that some concatamers (almost exclusively just single concatamer additions) were observed when the iso₃TS oligo was used suggests that a residual level of terminal transferase activity was still present. However, our results clearly show that this activity was greatly reduced when compared to conditions using the original TS oligo. Perhaps using more than three iso-C/iso-G bases and/or addition of Mn²⁺ ions would help reduce this activity further as the presence of Mn²⁺ appears to have a very dramatic effect on terminal transferase activity and generation of CCC tail [17].

The increased efficiency of TS-PCR achieved by use of the iso₃TS oligo design enhances the amenability of small or rare biological samples to genomics investigation and may facilitate comparative functional genomics approaches to understanding processes such as cancer pathogenesis or stem cell differentiation at very early stages. Very small biological samples, on the order of 10 s to 100 s of cells, would be sufficient to generate cDNA for cloning and sequencing, potentially opening new avenues for investigation of important biological processes.

Conclusions

The procedure presented here was developed to permit synthesis of cDNA using the TS-PCR approach from very small quantities of starting material, such as manually selected glandular trichomes or any sample of a particularly limited nature. cDNAs are synthesized using an adaptation of the SuperSMART approach relative to the sample handling. However, the primers used are similar to those from the SMART cDNA synthesis kit to enable *Sfi*I digestion of the cDNA and ligation into the pDNR-LIB vector for library propagation in *E. coli* and subsequent Sanger sequencing. Digestion with *Sfi*I also allows for cDNA produced in this way have most of the primer sequence removed prior to use in massively parallel sequencing approaches (next-generation sequencing) such as 454 Life Sciences GS FLX Titanium series sequencing. Validation of the utility of this method in such approaches was provided by successful construction and 454 sequencing of cDNA from rhizome

samples of two divergent plant species: scouring rush (fern ally, “primitive plant”) and red rice (monocot, “advanced plant”). We have continued to use this method to produce cDNA samples from several other plant species with similarly successful results in next-generation sequencing.

Methods

Tissue collection

Glandular trichomes (glands) were selected by hand from intact leaves of various *Solanum* and other species. Briefly, very fine, stretched glass microcapillaries were drawn by hand and the ends sealed. These glass microcapillary probes were subsequently used to touch and remove specific types of glandular trichomes from the surfaces of the leaves. A counted number of glands was collected on a probe over a short period of time, and the end of the probe with the glands was transferred to a microfuge tube containing 100 µl of RNAlater (Ambion, Austin, TX, USA). RNA isolation was performed after the desired number of glands had been accumulated (see Results for number of glands collected for specific experiments). Rhizome tips and elongation zones (~10 - 100 mg tissue) were dissected and collected from fresh rhizomes of rhizomatous species, scouring rush and red rice, and immediately frozen in liquid nitrogen.

RNA isolation

RNA was isolated from glandular trichome samples using Machery-Nagel (Clontech, Mountain View, CA) NucleoSpin RNA II kits, following the manufacturer’s instructions, with the following exceptions: 1) 350 µl of buffer RA1 (with B-mercaptoethanol added) was added to manually selected glands contained in microfuge tube in 100 µl of RNAlater (Ambion); 2) glands were disrupted using a probe sonicator (Branson Sonifier 450, setting #4 for ~30 s using microprobe); 3) RNA elution was done in 2 steps, using 2 volumes of 27 µl of nuclease-free water, into the same microfuge tube, yielding a total eluate volume of slightly more than the 52.5 µl for use as input for the first strand cDNA synthesis step below. For rhizomatous species, RNA was isolated from rhizome tips and elongation zones using RNeasy Plant Mini Kit (Qiagen, Valencia, CA) according to the manufacturer’s instructions. These different RNA isolation kits were used to ensure that RNA isolated by different approaches would work equally well in the modified cDNA synthesis procedure outlined below.

Primer sequences

Primers for first strand synthesis: TS oligo (*Sfi*I site **bold**) 5'-AAGCAGTGGTATCAACGCAGAGT**GGC-CATTACGGCC**rGrGrG-3' (rG = ribonucleotide G);

iso₃TS oligo (*Sfi*I site **bold**) 5'-iCiGiCAAGCAGTGG-TATCAACGCAGAGTGGCCATTACGGCCrGrGrG-3' (iC = iso-dC, iG = iso-dG); Oligo-dT primer (*Sfi*I site **bold**) 5'-TAGAGGCCGAGGCCGCCGACATGTTT TGT TTT TTT TTT TCT TTT TTT TTT TTT VN-3'. Primers for LD-PCR: iso₃TS oligo as 5'PCR primer; 3'PCR primer: 5'-TAGAGGCCGAGGCCGCCGAC-3'.

First strand cDNA synthesis

First strand cDNA synthesis was performed in 106 µl reactions assembled in the following manner. For each sample, the following components were placed in a 200 µl PCR tube on ice: 52.5 µl RNA sample, 8.4 µl oligo dT primer (10 µM), 8.4 µl TS oligo or iso₃TS oligo (10 µM), 69.3 µl total volume. RNA and primer mixes were denatured at 65°C for 2 min and then removed to ice. The following reagents were then added to each reaction tube: 20 µl 5× first strand buffer (Powerscript [Clontech] RT buffer: 250 mM Tris (pH 8.3), 30 mM MgCl₂, 375 mM KCl, 20 mM DTT), 2 µl DTT (100 mM), 10 µl 50× dNTP (10 mM), 5 µl SuperScript II (Invitrogen, Carlsbad, CA). First strand reactions were incubated at 42°C for 2 hr.

First strand cDNA cleanup

A cleanup step was performed for first strand cDNA samples using Machery-Nagel NucleoSpin Extract II kits following the manufacturer's protocol and the total recovered elution volume was 80-85 µl per sample.

cDNA amplification by Long-Distance PCR (LD PCR)

For PCR amplification of full length cDNAs, 80 µl of cleaned-up first strand cDNA was used as template in 100 µl PCR reactions including the following reagents: 1× Advantage 2 PCR Buffer (Clontech), 0.2 mM dNTPs, 0.12 µM 5'PCR primer (iso₃TS oligo), 0.12 µM 3'PCR primer, and 1× Advantage 2 Polymerase Mix. PCR amplifications were performed using an MJ Research Dyad (MJ Research, Waltham, MA, USA) with the following program parameters: an initial hold at 95°C for 1 min, followed by 27 cycles of 95°C for 5 s, 65°C for 5 s, and 68°C for 6 min. After completion of a PCR reaction, a 5 µl aliquot of each sample was removed and run on a 1.1% agarose/TAE gel containing 0.5 µg/ml ethidium bromide alongside a DNA MW marker such as the Fermentas Mass Ruler (Fermentas, Glen Burnie, MA, USA).

LD PCR reaction clean up

LD PCR amplified cDNA samples were cleaned up using NucleoSpin II kits following the manufacturer's protocol and the total combined eluate was approximately 32 µl.

*Sfi*I digestion of cDNA

*Sfi*I digestion was performed in 40.5 µl total volumes containing 32 µl cDNA, 4.05 µl 10× *Sfi*I buffer 4.05 µl

*Sfi*I enzyme, and 0.4 µl 100× BSA. Digestions were incubated at 50°C for 2 hrs before proceeding to cDNA size fractionation. *Sfi*I buffer and enzyme were obtained from NEB (Ipswich, MA, USA).

cDNA size fractionation by gel electrophoresis

cDNA was fractionated by molecular weight using preparative DNA gel electrophoresis after *Sfi*I digestion. Digested cDNA samples were loaded (using a loading buffer containing bromophenol blue) on a 2% low melt agarose (NuSieve GTG, Cambrex, Rockland, ME, USA) gel containing 0.5 µg/ml ethidium bromide. Electrophoresis was performed at a low voltage setting (5-8 V/cm) until the bromophenol blue dye had migrated approximately 4 cm. Gels were photographed and the lanes excised from the 400 bp molecular weight marker up to the sample well. DNA was isolated from the gel slice using a Qiagen MinElute kit (Valencia, CA, USA) and 6 sample volumes of buffer QG. Elution was performed using 10 µl of elution buffer. Ligations into pDNR-LIB (Clontech) were performed immediately following cDNA size fractionation and elution according to the method specified in the SMART cDNA library construction manual. Sequencing of gland-derived samples was accomplished using traditional Sanger sequencing of individual clones. The digested, fractionated cDNA samples from rhizome tips and elongation zones were purified using QIAquick PCR purification kit (Qiagen). 5~10 µg purified cDNA from each rhizome sample was used in construction of a 454 Genome Sequencer FLX [21-23] fragment library using Titanium series reagents <http://www.454.com> followed by sequencing on a GS FLX instrument.

Acknowledgements

We thank Shari Tjugum-Holland for assistance with the 454 sequencing, Dr. Cari Soderlund and Anne Descour for their help with bioinformatic analyses. This work was supported by US National Science Foundation grants (DBI-0604336 and DBI-0820346).

Author details

¹School of Plant Sciences and BIO5 Institute, The University of Arizona, 1657 E Helen Street, Tucson, AZ 85721, USA. ²Institute of Biological Chemistry, Washington State University, PO Box 646340, Pullman, WA 99163, USA.

Authors' contributions

JK and RH carried out the experiments, analyzed the data and wrote the manuscript. ETM participated in sample collection and library preparation. DRG designed and coordinated the study, and edited the manuscript. All authors read and approved the final manuscript.

Received: 11 December 2009 Accepted: 2 July 2010

Published: 2 July 2010

References

1. Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, Liotta LA: *Laser Capture Microdissection. Science* 1996, **274**(5289):998-1001.

2. Cloonan N, Forrest ARR, Kolle G, Gardiner BBA, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, et al: **Stem cell transcriptome profiling via massive-scale mRNA sequencing.** *Nature Methods* 2008, **5**(7):613-619.
3. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nature Methods* 2008, **5**(7):621-628.
4. Salehi-Ashtiani K, Yang XP, Derti A, Tian WD, Hao T, Lin CW, Makowski K, Shen L, Murray RR, Szeto D, et al: **Isoform discovery by targeted cloning, 'deep-well' pooling and parallel sequencing.** *Nature Methods* 2008, **5**(7):597-600.
5. Shendure J: **The beginning of the end for microarrays?** *Nature Methods* 2008, **5**(7):585-587.
6. Iscove NN, Barbara M, Gu M, Gibson M, Modi C, Winegarden N: **Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA.** *Nature Biotechnology* 2002, **20**(9):940-943.
7. Petalidis L, Bhattacharyya S, Morris GA, Collins VP, Freeman TC, Lyons PA: **Global amplification of mRNA by template-switching PCR: linearity and application to microarray analysis.** *Nucleic Acids Res* 2003, **31**(22):e142.
8. Edery I, Chu LL, Sonenberg N, Pelletier J: **An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture).** *Mol Cell Biol* 1995, **15**(6):3363-3371.
9. Kato S, Sekine S, Oh SW, Kim NS, Umezawa Y, Abe N, Yokoyamakobayashi M, Aoki T: **Construction of a human full-length cDNA bank.** *Gene* 1994, **150**(2):243-250.
10. Carrinci P, Westover A, Nishiyama Y, Ohsumi T, Itoh M, Nagaoka S, Sasaki N, Okazaki Y, Muramatsu M, Schneider C, et al: **High efficiency selection of full-length cDNA by improved biotinylated cap trapper.** *DNA Res* 1997, **4**(1):61-66.
11. Maruyama K, Sugano S: **Oligo-Capping - a simple method to replace the cap structure of eukaryotic messenger-RNAs with oligoribonucleotides.** *Gene* 1994, **138**(1-2):171-174.
12. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD: **Reverse transcriptase template switching: A SMART (TM) approach for full-length cDNA library construction.** *Biotechniques* 2001, **30**(4):892-897.
13. Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatchenko L, Chenchik A: **Amplification of cDNA ends based on template-switching effect and step-out PCR.** *Nucleic Acids Res* 1999, **27**(6):1558-1560.
14. Zhu Y, Chenchik A, Li R, Diatchenko L, Naniselli A, Siebert P: **CAPswitch based PCR technology for full-length cDNA library construction from small amount of RNA.** *FASEB Journal* 1996, **10**(6):738-738.
15. Zhu Y, Chenchik A, Li R, Kitts P, Siebert P: **A novel system based on m-mlv catalyzed template switching to facilitate full-length cDNA library construction.** *FASEB Journal* 1997, **11**(9):1982.
16. Zhu Y, Chenchik A, Li R, Siebert P: **Representative cDNA synthesis from nanogram level of total RNA: A novel method using the template switching reaction catalyzed by M-MLV reverse transcriptase.** *Molecular Biology of the Cell* 1996, **7**:929-929.
17. Schmidt WM, Mueller MW: **CapSelect: a highly sensitive method for 5' CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs.** *Nucl Acids Res* 1999, **27**(21):e31.
18. Piccirilli JA, Krauch T, Moroney SE, Benner SA: **Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet.** *Nature* 1990, **343**(6253):33-37.
19. Switzer C, Moroney SE, Benner SA: **Enzymatic incorporation of a new base pair into DNA and RNA.** *Journal of the American Chemical Society* 1989, **111**(21):8322-8323.
20. Nolte FS, Marshall DJ, Rasberry C, Schievelbein S, Banks GG, Storch GA, Arens MQ, Buller RS, Prudent JR: **MultiCode-PLx system for multiplexed detection of seventeen respiratory viruses.** *Journal of Clinical Microbiology* 2007, **45**(9):2779-2786.
21. Droege M, Hill B: **The genome sequencer FLX (TM) system-longer reads, more applications, straight forward bioinformatics and more complete data sets.** *Journal of Biotechnology* 2008, **136**(1-2):3-10.
22. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen ZT, et al: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**(7057):376-380.
23. Rothberg JM, Leamon JH: **The development and impact of 454 sequencing.** *Nature Biotechnology* 2008, **26**(10):1117-1124.

doi:10.1186/1471-2164-11-413

Cite this article as: Kapteyn et al.: Incorporation of non-natural nucleotides into template-switching oligonucleotides reduces background and improves cDNA synthesis from very small RNA samples. *BMC Genomics* 2010 **11**:413.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

