iScript™ Select cDNA Synthesis Kit Has Highest Fidelity as Determined by Next-Generation Sequencing

Xiangdong Meng, Cindy Wan, Jennifer Dennis, Xiao-Song Gong
Bio-Rad Laboratories, Inc., 2000 Alfred Nobel Drive, Hercules, CA 94547

Abstract

Here we describe a high-throughput method for determining reverse transcriptase (RT) fidelity using next-generation sequencing (NGS) technology. This method integrates 12 randomized bases of barcodes in the primers for cDNA synthesis. This barcoding strategy was used to filter out PCR amplification and sequencing errors from those generated during the reverse transcription reaction, making the detection of reverse transcription mutations more accurate and reliable than previously established techniques. The error rates of DNA polymerase and MMLV RT detected by our NGS method were comparable to previous reports using similar approaches. This method was used to compare the fidelity of six commonly used RT kits. Our results demonstrate that the Bio-Rad iScript Select cDNA Synthesis Kit has the highest fidelity of the RT kits examined, making it ideal for applications such as cloning, sequencing, and RNA-Seq.

Introduction

Reverse transcriptases (RTs) transcribe RNA into cDNA and play an important role in molecular biology research because they enable applications such as gene expression analysis, cloning, and RNA-Seq. RTs from Moloney murine leukemia virus (MMLV) and avian myeloblastosis virus (AMV) are commonly used in cDNA synthesis. Because RTs lack 3' to 5' proofreading capability, they can introduce errors during cDNA synthesis (Menendez-Arias 2009). Additionally, because the relative error rate of RTs is much higher than the error rate of DNA polymerases, it is especially important to choose a high-fidelity RT for RT-PCR applications requiring high fidelity. RT fidelity defines the extent of mutations/errors introduced during cDNA synthesis and is quantified as the number of errors per total number of bases transcribed (error rate). High-fidelity/low error rate is a highly desirable feature for RTs used in cloning, cDNA library preparation, cDNA probe preparation for microarray experiments, sequence validation, and RNA-Seq applications, as even single base errors could have dramatic effects on downstream analyses.

One of the traditional assays used to quantify RT error rate is the M13 *lacZ* forward mutation assay (Menendez-Arias 2009, Roberts et al. 1988). In this assay, gapped duplex DNA is generated by an RT from a single-stranded M13mp2 template. Errors that occur during this process are quantified by analyzing the phenotypical color change of plaques generated after infection of host bacteria. This phenotypic assay is tedious, as it is low throughput and dependent on phenotypical plaque counting. Additionally, silent mutations that do not result in a plaque color change are not detected by this assay. It is also a challenge to achieve a large number of white/blue plaques that can be counted reproducibly and reliably. Therefore, we sought to develop a new, more accurate and higher throughput approach to assess RT fidelity.

NGS is widely used to identify rare mutations in genomes and misincorporation during DNA synthesis as it can generate a large number of sequence reads rapidly (Ellefson et al. 2016, Lee et al. 2016). We adapted a DNA polymerase fidelity assay (Lee et al. 2016) to generate a faster and high-throughput approach to determine RT fidelity using NGS. Our results demonstrate that the iScript Select Kit is highly accurate, making it ideal for applications that require high fidelity.

BIO-RAD

Materials and Methods

RNA transcribed from a plasmid containing a 217 bp DNA template of the ampicillin resistance gene with a T7 promoter was used to synthesize cDNA. Primers containing barcodes of 12 randomized bases were used for cDNA synthesis (Figure 1, Figure 2). Bio-Rad's iScript Select Kit, containing a recombinant enzyme with a proprietary reaction buffer, was compared to five RT kits from other vendors, including a wild-type MMLV RT, in NGS-based fidelity assays. One microgram of RNA was used in a 20 µl reverse transcription reaction using 250 nM of gene-specific primers following manufacturer protocols. Four replicate reverse transcription reactions were performed. The resulting cDNA templates were used for primer extension using a high-fidelity DNA polymerase (Bio-Rad's iProof™ High-Fidelity Master Mix) and primers containing barcodes of another 12 randomized bases, and the resulting products were used for Illumina sequencing library preparation. iProof High-Fidelity Master Mix was used for library amplification for at least 20 amplification cycles to ensure enough reads with over three replicates. Fidelity of iProof DNA Polymerase was also determined using the same template of DNA to validate our sequencing results and to remove mutations generated by PCR from the calculated RT error rate. After sequencing with a MiSeq Desktop Sequencer (Illumina) using the MiSeq V2 PE2x150 300 Cycle Kit, the paired-end FASTQ sequencing reads were analyzed using sequencing analysis software developed in-house. After removing reads that failed QC and sequence alignment, reads with the same unique barcodes successfully obtained at least three times were counted (these were called valid reads) and mutations were scored as RT error only when the unique reads all had the same mutations. The errors from PCR amplification and sequencing were filtered out through this step. Mutation types, such as substitution, insertion, and deletion, were also analyzed. The calculation of total errors divided by total valid bases resulted in the error rate, and fidelity was calculated as the inverse of error rate. Means and standard errors from four replicates were used to compare the error rate and fidelity of all results.

DNA template

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNttcggtcctccgatcgttgtcagaagtaagttggccg
cagtgttat cactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagtt
gctcttgcccggc gtcaacacgggataataccgcgccacatagcagaaNNNNNNNNNNNNTCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTGNNNNNNTAGAG
CATACGGCAGAAGACGAAC

Gene-specific primers used for cDNA synthesis and primer extension

5'-TCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNttctgctatgtggcgcggta-3'
5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNttcggtcctccgatcgttgt-3'

Primers for Illumina library construction

P5-RD1: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
P7-index-RD2: 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'

Fig. 1. DNA template sequence, Illumina sequencing template, and primer sequences. Sequence used for error rate determination (—); target sequences for gene specific primers (—); sequencing adaptors are capitalized; The 177 bp sequence excluding the primer target sequence is the template used for sequencing alignment; 12 bp N barcodes (—); 6 bp N index barcode (—); Illumina P5-RD1 complementary sequence and P7-index-RD2 complementary sequence (—).

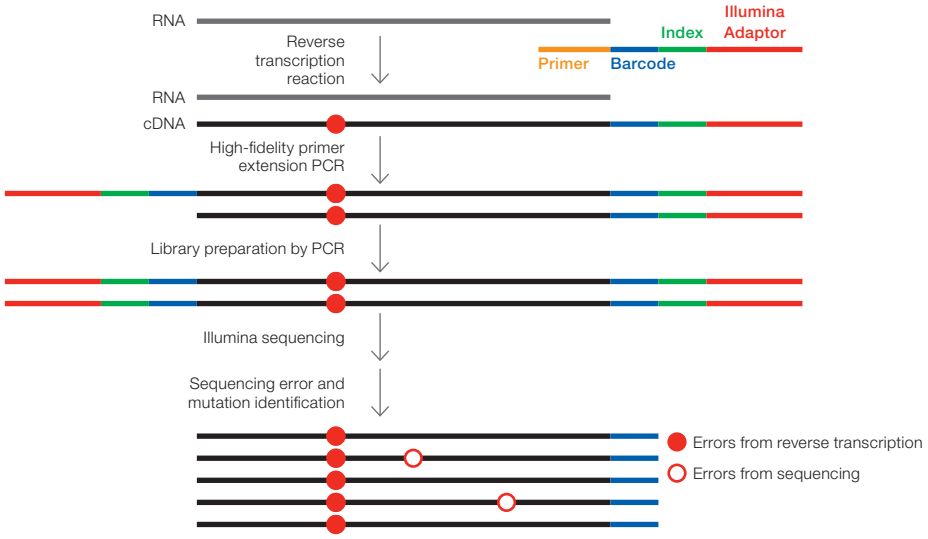

Fig. 2. Workflow diagram for RT fidelity determination. Sequencing reads with the same barcodes are binned to identify the same errors and filter out sequencing errors. Only errors from reads with the same barcodes found in three replicates are used to calculate RT error rates.

Results

We developed a fast and high-throughput method to quantify RT fidelity using NGS (Figure 2). RNA templates were reverse transcribed into cDNA by RTs using primers containing 12 bp molecular barcodes to tag each strand of cDNA with a unique barcode. This unique barcode allowed identification of mutations generated during the cDNA synthesis when more than three sequencing reads had the same barcode and mutations, as determined by NGS. This method allowed us to calculate error rates for different RTs.

Table 1 compares fidelity results obtained in these experiments as well as published fidelity results for similar enzymes. Bio-Rad's iProof High-Fidelity Master Mix DNA Polymerase demonstrated an error rate of 2.70 x 10^{-6}, in line with the published error rate of New England BioLabs' Q5 High-Fidelity DNA Polymerase. Furthermore, the MMLV RT error rates obtained using the method described here are comparable to those previously reported (Ellefson et al. 2016, Lee et al. 2016, McInerney et al. 2014). Taken together, the comparable error rates observed across the different experiments support the validity of the fidelity experiment described here. The small standard errors obtained in our experiments indicate the high reproducibility of these assays. An additional benefit to this fidelity assessment is that our approach also provides accurate information for all different mutation types.

The fidelity of the Bio-Rad iScript Select Kit was then compared to several commonly used RT kits. The results demonstrated that Bio-Rad's iScript Select Kit had the lowest error rate of all the RT kits studied (Table 2, Figure 3). Our data also revealed that the majority of mutations in samples generated using the iScript Select Kit were substitutions, while other RT kits had a high occurrences of insertions and deletions (Table 2). High occurrences of insertions and deletions are more likely to result in changes at the protein level, a serious problem when cloning cDNA into expression cassettes. These results confirm that Bio-Rad's iScript Select Kit is a high-fidelity RT kit that is suitable for applications that require high accuracy, such as cloning, cDNA library preparation, and RNA-Seq.

Table 1. Comparison of RT and DNA polymerase error rates.

Enzymes	Data Sources	Method	Total Bases	Total Errors	Substitutions	Insertions	Deletions	Error Rate, bp	SE
iProof DNA Polymerase	Our data	NGS	12,294,720	33	33	0	1	2.70 x 10^{-6}	1.02 x 10^{-7}
Q5 DNA Polymerase	Lee et al. 2016	NGS	15,347,568	46	45	0	1	3.00 x 10^{-6}	Not reported
MMLV RT	Our data	NGS	18,466,230	18,693	2,669	15,996	28	1.01 x 10^{-3}	8.90 x 10^{-5}
MMLV RT	Ellefson et al. 2016	NGS	1,120,000	543	536	7 (Indel)	Not reported	4.86 x 10^{-4}	Not reported

SE, standard error.

Table 2. Error rates of different reverse transcriptases detected by NGS.

Reverse Transcriptase Kits	Overall Error Rate, bp	SE	Total Valid Bases Sequenced	Total Errors	Substitutions	Insertions	Deletions	Sub%	Ins%	Del%
iScript Select	2.47 x 10^{-4}	3.38 x 10^{-5}	9,768,480	2,534	2,509	21	4	99.05%	0.74%	0.21%
RT1	2.73 x 10^{-3}	3.65 x 10^{-5}	13,905,420	37,910	29,548	8,194	169	77.87%	21.68%	0.44%
RT2	2.33 x 10^{-3}	6.92 x 10^{-5}	12,111,540	28,420	20,024	8,115	281	72.26%	26.81%	0.93%
RT3	2.05 x 10^{-3}	6.12 x 10^{-5}	12,186,930	25,243	19,687	5,242	314	77.64%	21.12%	1.24%
RT4	1.46 x 10^{-3}	3.44 x 10^{-4}	13,141,290	24,265	814	23,443	8	4.84%	95.12%	0.04%
MMLV	1.01 x 10^{-3}	8.90 x 10^{-5}	18,466,230	18,693	2,669	15,996	28	14.56%	85.30%	0.15%

SE, standard error; Sub, substitution; Ins, insertion; Del, deletion.

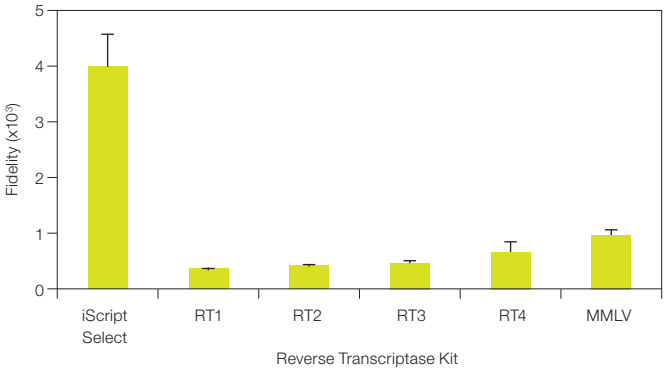

Fig. 3. iScript Select Kit demonstrates superior fidelity. A partial fragment of the ampicillin resistance gene was reverse transcribed using gene-specific primers with barcodes and one of six different commonly used reverse transcriptase kits according to manufacturer's instructions. PCR was performed using Bio-Rad's iProof High-Fidelity Master Mix. Resulting amplicons were sequenced using the Illumina MiSeq platform. For each sample, approximately 10 million bases of valid reads with three or more repeats were obtained. Fidelity of each reverse transcriptase was determined by taking the inverse of the error rate. Mean fidelity of four replicates is shown on the y-axis. Error bars are standard errors. Reverse transcriptase kits are indicated on the x-axis.

Conclusion

We developed a rapid, accurate, and high-throughput approach to determine RT fidelity using NGS. This approach can reliably detect mutations from cDNA synthesis. Our fidelity results are comparable to those obtained in other independent reports, which validates the method used for these experiments. We show that Bio-Rad's iScript Select Kit has the highest fidelity and lowest insertion and deletion occurrences among the RT kits assessed. This suggests that pairing the iScript Select cDNA Synthesis Kit with a high-fidelity DNA polymerase, such as the iProof High-Fidelity Master Mix Polymerase, enables researchers to minimize error rates when generating cDNA for downstream studies.

References

Ellefson JW et al. (2016). Synthetic evolutionary origin of a proofreading reverse transcriptase. Science 352, 1,590–1,593.

Lee DF et al. (2016). Mapping DNA polymerase errors by single-molecule sequencing. Nucleic Acids Res 44, e118.

McInerney P et al. (2014). Error rate comparison during polymerase chain reaction by DNA polymerase. Mol Biol Int 2014, 1–8.

Menendez-Arias L (2009). Mutation rates and intrinsic fidelity of retroviral reverse transcriptases. Viruses 1, 1,137–1,165.

Roberts JD et al. (1988). The accuracy of reverse transcriptase from HIV-1. Science 242, 1,171–1,173.

Visit **bio-rad.com/fidelity** for more information.

BIO-RAD

Bio-Rad Laboratories, Inc.

Life Science Group

Web site *bio-rad.com* ***USA*** *1 800 424 6723* ***Australia*** *61 2 9914 2800* ***Austria*** *43 01 877 89019* ***Belgium*** *32 03 710 53 00* ***Brazil*** *55 11 3065 7550* ***Canada*** *1 905 364 3435* ***China*** *86 21 6169 8500* ***Czech Republic*** *36 01 459 6192* ***Denmark*** *45 04 452 10 00* ***Finland*** *35 08 980 422 00* ***France*** *33 01 479 593 00* ***Germany*** *49 089 3188 4393* ***Hong Kong*** *852 2789 3300* ***Hungary*** *36 01 459 6190* ***India*** *91 124 4029300* ***Israel*** *972 03 963 6050* ***Italy*** *39 02 49486600* ***Japan*** *81 3 6361 7000* ***Korea*** *82 2 3473 4460* ***Mexico*** *52 555 488 7670* ***The Netherlands*** *31 0 318 540 666* ***New Zealand*** *64 9 415 2280* ***Norway*** *47 0 233 841 30* ***Poland*** *36 01 459 6191* ***Portugal*** *351 21 4727717* ***Russia*** *7 495 721 14 04* ***Singapore*** *65 6415 3188* ***South Africa*** *36 01 459 6193* ***Spain*** *34 091 49 06 580* ***Sweden*** *46 08 555 127 00* ***Switzerland*** *41 0617 17 9555* ***Taiwan*** *886 2 2578 7189* ***Thailand*** *66 2 651 8311* ***United Arab Emirates*** *971 4 8187300* ***United Kingdom*** *44 01923 47 1301*

Bulletin 7076 Ver A US/EG *18-0283 0418 Sig 0118*