

Optimization of Illumina TruSeq Targeted RNA Expression (TREX) Library Quality (Pengoptimuman Ekspresi RNA Sasaran (TREX) Illumina TruSeq Kualiti Perpustakaan)

NADIAH ABU, SWEE KEONG YEAP*, NORAINI NORDIN, SHEAU WEI TAN, WAN YONG HO & NOORJAHAN BANU ALITHEEN

ABSTRACT

RNA-seq has become an essential tool in molecular research. Nevertheless, application of RNA-seq was limited by cost and technical difficulties. Illumina has introduced the cost effective and ease to handle Truseq Targeted RNA Sequencing. In this study, we present the requirements and the optimization procedure for this Truseq Targeted RNA sequencing on cell line. Total RNA was recommended as starting materials but it required optimization including additional purification step and adjusting the AMPure beads ratio to eliminate unwanted contaminants. This can be resolved by using PolyA-enriched mRNA as starting material. TREx is a useful assay to evaluate gene expression. Quality library of TREx can be prepared by adding multiple washing steps or changing input sample to mRNA.

Keywords: Illumina; poly(A) mRNA; targeted expression; TREx

ABSTRAK

RNA-seq telah menjadi teknik penting di dalam kajian molekul. Walau bagaimanapun, kos yang tinggi serta kesukaran daripada segi teknikal telah menyebabkan kurangnya aplikasi RNA-seq. Illumina telah memperkenalkan teknik baru, penjujukan sasaran RNA Truseq yang lebih murah dan senang digunakan. Di dalam kajian ini, kami telah memperkenalkan syarat dan prosedur yang optimum bagi teknik ini. Penggunaan RNA keseluruhan sebagai bahan permulaan adalah digalakkan tetapi ia memerlukan kaedah pengoptimuman seperti pertambahan pembersihan dan penyesuaian nisbah manik 'AMPure' untuk menghapuskan bahan cemar yang tidak diinginkan. Ini boleh diselesaikan dengan menggunakan 'PolyA-enriched' mRNA sebagai bahan permulaan. TREx adalah asai berguna untuk menilai pengekspresan gen. Kualiti perpustakaan TREx boleh disediakan dengan menambah langkah basuhan berganda atau menukar sampel input kepada mRNA.

Kata kunci: Ekspresi sasaran; illumina; poli(A) mRNA; TREx

INTRODUCTION

Transcriptome evaluation is one of the ways to understand the functional elements of a genome. Among the transcript, messenger RNA (mRNA) is the most widely studied RNA for gene expression study. Multiple methods that have been developed to allow the evaluation of mRNA including; hybridization (microarray, nanostring, northern blot), PCR (quantitative reverse transcriptase PCR/qRT-PCR) and sequencing methods (Sanger sequencing + Expressed Sequence Tag, Next generation RNA sequencing) (Wang et al. 2009). However, the major drawbacks of using NGS RNA sequencing comparing to all the other transcriptomic analyses are cost and difficulties with data analysis, which hinder its application in fundamental researches (Hitzemann et al. 2013) and even clinical translation (Long et al. 2014). Furthermore, validation of other high throughput transcriptome data such as microarray and PCR array may need a targeted sequencing rather than whole transcriptome sequencing. To overcome these barriers, the two leading providers of next generation sequencing, Illumina and Thermo Fisher Scientific, have

offer alternative solution of targeted RNA sequencing name as Illumina TruSeq Targeted RNA-Expression kit (TREX) and Ion AmpliSeq RNA targeted panel, respectively. Both systems offer customizable medium to high-plex targeted gene expression profiling. However, Illumina TREX offers more choices of ready-to-use fixed panels for 10 pathways/diseases comparing to Ion AmpliSeq that only offers 2 panels for now (October 2015). Thus, Illumina TREX offers an easier entry and more choices for beginners to utilize this targeted RNA sequencing in their gene expression study. In TREX, read budget for multiplexing of ~100 samples can be achieved with good read depth (more than 1000 reads/assays) for most of the panels (which contain approximate 100+ mRNA targets) in one MiSeq run to generate that equivalent to approximate 10,000 qRT-PCR assays. Kellum et al. (2015) has successfully utilized TREX apoptosis panel to validate *in vitro* mammalian cell micronucleus test in evaluating genotoxicity of clastogenic and aneugenic chemicals in human lymphoblasts TK6 cells. In addition, Hashida et al. (2015) have utilized TREX to identify genes that contributed to the resistance of lung cancer stem cells to afatinib. Other published application of

TREX including evaluation of genes involved in monogenic human cardiomyopathies in a polygenic model (Prestes et al. 2016) and correlation between methylation with gene expression in human fetal brain development (Schneider et al. 2016). However, reports on usage of TREX for medium scale gene expression evaluation are still lacking. Here, we share the optimization of TREX library preparation on the human breast cancer cell line.

MATERIALS AND METHODS

CELL PREPARATION AND TREATMENT

MCF-7 and MDA-MB231 breast cancer cell lines were purchased from American Type Culture Collection (ATCC, USA) and cultured using RPMI1640 medium (Sigma-aldrich, USA) supplemented with 10% Foetal Bovine Serum (FBS) (Gibco, USA) at 37°C, 5% CO₂. Citral is a compound present in essential oils from lemon grass (*Cymbopogon citrates*). Recent report has shown that citral possessed cytotoxicity on MCF-7 and MDA-MB231 breast cancer cell lines (Patel et al. 2015). In this study, untreated and citral treated (10 µg/mL, 48 h) of MCF-7 and MDA-MB231 were prepared and subjected to TREX analysis.

RNA EXTRACTION AND QUALITY CONTROL OF RNA

Total RNA was extracted from each set of the cells using the RNeasy Mini Kit (Qiagen, USA) followed by an on-column DNase treatment (Qiagen, USA). Before proceeding to the library preparation of TREX assay, concentration, integrity and purity of the RNA was checked. The purity of the RNA was checked using a nano-drop Spectrophotometer (Beckman Coulter, USA). The readings of A₂₆₀/A₂₃₀ and A₂₆₀/A₂₈₀ were obtained and the ratio for all of the samples must be more than or equal to 2.0. To check for the concentration of the RNA, the Qubit Fluorometric Quantitation method was employed (Life Technologies, USA). To measure the integrity of the RNA (no degradation of RNA), the Eukaryote Total RNA chip on the Bioanalyzer 2100 was used. The RIN (RNA Integrity Number) for each of the samples were measured and all samples must have a RIN more than 8 before qualifying for the TREX assay.

TREX LIBRARY PREPARATION

There are several major steps that were done in the TREX Library Preparation. After all the samples have passed the quality control checkpoints, the total RNA was diluted to 10 ng/µL using nuclease-free water. Then, cDNA for each samples were synthesized from the diluted RNA using ProtoScript III and RCSI. The protocol for cDNA synthesis is as follow; 25°C for 5 min, 42°C for 15 min, 95°C for 10 min and hold at 4°C. Next step is hybridize oligo pool using the following settings; 70°C for 5 min, 68°C for 1 min, 65°C for 2.5 min, 60°C for 2.5 min, 55°C for 4 min, 50°C for 4 min, 45°C for 4 min, 40°C for 4 min, 35°C for 4 min, 30°C for 4 min and hold at 30°C. Afterwards, the samples were subjected to the quality control as stated below.

QUALITY CONTROL AFTER LIBRARY PREPARATION

After successfully preparing the library for the samples, there are several quality control checkpoints that should be done. First, is the Qubit Fluorometric Quantitation analysis, the amount of the samples will be determined. The recommended concentration must be between 1-10 ng/µL. Then, the region (size of base pairs) and the molarity of the library will be analyzed using the Agilent Technologies DNA 1000 chip. According to the illumina protocol, the region should be between 100-300 bp with no additional peaks. Afterwards, to confirm the library quantitation the KAPA Library Quantification Kit for Illumina platforms was used on CFX-96 BioRad qPCR Thermal Cycler (Bio-rad, USA).

SECOND PURIFICATION STEP

If the samples did not pass the quality control checkpoint due to the presence of primer dimers or adapter dimers, a second purification step should be done. The purification step is done using AMPure beads at a ratio of 1:1.5 from the volume containing the library. The mixture was vortexed on a microplate shaker at 1800 rpm for 2 min. Then the plate was placed on a magnetic stand for 5 min until the liquid appear clear. The supernatant was then removed and the pellet was washed with 80% ethanol twice before leaving it to air-dry for 15 min. After that, the samples were eluted using 15 µL of resuspension buffer and were subjected to quality check again.

MRNA ISOLATION

As the MCF-7 untreated samples were not able to obtain enough quality even after 3rd round of purification, total RNA from this sample was subjected to mRNA isolation using the NEBNext Magnetic Oligo d(T)₂₅. Then, the isolated mRNA was subjected to the same TREX library preparation and quality control for comparison.

MISEQ RUN

The sequencing was run on an Illumina Miseq platform based on Illumina's guidelines. All of the libraries were normalized and pooled. Around 9 pM with 1% phix control (small viral genome used to enable quick alignment and estimation of error rates) of the pooled libraries were loaded into a flowcell.

STATISTICAL ANALYSIS

Means and standard deviations of triplicates for the concentration of the library were calculated by excel. One way analysis of variance (ANOVA) followed by post-hoc Tukey's test using SPSS version 16. $p < 0.05$ was considered significant.

RESULTS

In this study, optimisation of TREX library preparation was performed using both untreated and citral treated breast cancer cells. Citral is a terpenoids that was previously

reported with cytotoxic effect on both MCF-7 and MDA-MB231 cells. Citral treated MCF-7 and MDA-MB231 cells were prepared along with the untreated cells to allow comparison of the TREX library quality prepared from healthy cells and cells that underwent apoptosis. Based on Table 1, after the library preparation, a set of quality control checkpoints were conducted on the final product. The concentration of the final product was measured using the Qubit Fluorometric Quantitation and bioanalyzer technique (fragment size). Additionally, the size of the library should also be measured and this was done using the BioAnalyzer machine (Life Technologies, USA). Per the illumina protocol for this assay, the size of the library should fall between 100 – 300 bp. In this experiment, all QC methods including qubit, bioanalyzer and qPCR provided almost similar values. As illustrated in Figure 1(a), after the first clean-up during library preparation using $1.7 \times$ AMPure beads, there are multiple peaks positioned at both upstream and downstream of the desired peak in all samples. It is important to remove these peaks as it may interfere with the quantification values and result in an improper dilution when pooling the samples. Therefore, we resorted to a second clean-up using a different ratio of AMPure beads. The second clean-up was repeated based on the protocol aligned by Illumina for TREX using a

ratio of $1.5 \times$ of AMPure beads. After the second cleanup, the library was quantified using Qubit Fluorometric Quantitation again. As in Table 1, the value of the readings decreased by $\sim 70\%$ for all of the samples. The bioanalyzer distribution was also repeated using these samples, and most of the samples have a shift in the average size of the peak to a slightly lower value. Based on Figure 1, for all of the samples; Samples 2, 3 and 4 have clean single peaks. Most of the contaminants have been removed from the samples. Each of the samples from the first clean-up and second clean-up were run on the Mi-Seq to compare the results of sequencing. As shown in Table 1, the results of the Cluster PF, percentage of aligned and passing filter and percentage of aligned reads are presented. The % of Aligned is % of reads aligned to the reference sequences provided in the manifest file and only these reads will be used for further differential expression analysis. We noticed that the % Aligned is low for samples which has high level of adapter dimers contamination which means both the adapter dimers and our real libraries were being sequenced at the same time but the adapter dimers sequence cannot be aligned to the reference sequences and for further analysis. According to the manufacturer estimation, to achieve 2500 reads/assay for total of 174 assays, each sample should have a minimum of 435,000

TABLE 1. The pre and post library preparation quality control of total RNA and mRNA samples and values of clusters PF (Passing filter), percentage aligned and total PF and Aligned reads of the samples from first clean up and second clean up. Different roman number among the samples for concentration were indicated statistical significant ($p < 0.05$)

Samples	Qubit reading for concentration (ng/ μ L)		Average size of library (BioAnalyzer)		Concentration conversion to nM *		Presence of contaminants (%)		Proceed for targeted sequencing?		Clusters PF		% Aligned		PF and aligned reads	
	1st Post-QC ^c	2nd Post-QC ^d	1st Post-QC	2nd Post-QC	1st Post-QC	2nd Post-QC	1st Post-QC	2nd Post-QC	1st Post-QC	2nd Post-QC	1st Post-QC	2nd Post-QC	1st Post-QC	2nd Post-QC	1st Post-QC	2nd Post-QC
MCF-7 untreated total RNA	8.94 \pm 1.00 ^{e,iv}	1.97 \pm 0.50 ⁱⁱ	209	208	64.81 \pm 7.25 ^{iv}	14.35 \pm 3.64 ⁱⁱ	28	23.00	No	No	140565	1432183	57.8	46.62	81247	1205755
MCF-7 citral treated total RNA	2.56 \pm 0.6 ⁱⁱⁱ	0.56 \pm 0.10 ⁱ	209	201	18.55 \pm 4.40 ⁱⁱⁱ	4.24 \pm 0.76 ⁱ	18	<0.5	No	Yes	173041	1214129	57.5	84.19	99499	566027
MDA-MB231 untreated total RNA	1.49 \pm 0.4 ⁱⁱ	0.33 \pm 0.10 ⁱ	205	202	11.01 \pm 2.96 ⁱⁱ	2.49 \pm 0.75 ⁱ	12	<0.5	No	Yes	81325	1498743	67.36	93.49	54781	1401175
MDA-MB231 citral treated total RNA	1.10 \pm 0.4 ⁱⁱ	0.31 \pm 0.20 ⁱ	203	201	8.21 \pm 2.99 ⁱⁱ	2.26 \pm 1.46 ⁱ	9	<0.5	No	Yes	206538	1420827	84.29	81.33	174091	1155559
MCF-7 untreated mRNA	26.40 \pm 4.00 ^v	-	204	-	196.07 \pm 29.71 ^v	-	1.10	-	Yes	-	-	-	94.66	-	-	-

^aConcentrations of the samples were obtained from the readings of the Qubit Fluorometric Quantitation. Values are presented after including the dilution factor

^cQuality control post library preparation

^dQuality control after second step of purification using 1.5X AMPure Beads

^eConversion to nM was done based on the values obtained from the Qubit reading and the average size of the library

aligned and PF reads (174×2500). The samples from the first clean-up did not manage to pass the minimum number of required reads. The second clean-up managed to produce probable number of reads that can be used to analyze the results. Based on Illumina's recommendations, starting samples using mRNA has not been optimized yet and is not currently supported (October 2015). However, based on our observation and the results that we received, using mRNA as input for TREX assay could eliminate all the contaminants and result in a higher yield. Therefore, we decided to use one mRNA sample in the TREX library preparation to evaluate the difference. Based on Table 1, the quality of the RNA used before fishing out the mRNA was checked. The integrity, purity and concentration passed the minimum requirements; therefore we proceeded to mRNA isolation using NEBNext mRNA Isolation Protocol. The percentage of ribosomal RNA contamination was determined to estimate the efficiency and purity of the mRNA isolation method and below 5% of rRNA was detected indicating a good quality of mRNA was isolated. After obtaining the mRNA, the library preparation was performed using 5 ng of mRNA. Table 1 illustrates the results of the final product based on the qubit reading, size of the library and presence of contaminants. The concentration of mRNA MCF-7 untreated library (26.40 ± 4.00 ng/ μ L) had a significantly higher yield than the total RNA MCF-7 untreated library after 1st cleaning (8.94 ± 1.00 ng/ μ L) and second cleaning (1.97 ± 0.50 ng/ μ L). In Figure 1(B), based on the bioanalyzer results, there were no additional peaks besides the library peak.

DISCUSSION

During NGS library preparation, multiple quality control steps are needed to ensure the quality of the library before enter into sequencing. Among the quality control steps, Qubit, bioanalyzer and qPCR are the three most commonly used methods. Qubit has been proven to be more sensitive and accurate than Nanodrop spectrophotometer and is the standard measuring procedure to determine library concentration in the NGS-based workflow. Nevertheless, even though Qubit quantification is recommended, before proceeding to sequencing, it is advisable to reconfirm the concentration using qPCR (Simbolo et al. 2013). In this experiment, we found that both Qubit and qPCR quantification obtained similar results for the library concentration indicating that both Qubit and qPCR are suitable for TREX library concentration determination.

Another major issue that needs to be addressed is the presence of contaminants in the final library product, which can be identified by fragment analysis by bioanalyzer. Based on the bioanalyzer results, there should be one single clean peak at 100-300 bp only, if there are any other peaks, the samples are deemed contaminated. There are various types of contaminants such as primer dimers, adapter dimers and non-specific DNA (Head et al. 2014). Based on our results, we found that according to the Illumina recommended protocol, different degree of none-specific fragment still can be observed across

the samples. Variation of the contaminants across the samples maybe contributed by the different level of the targeted mRNA to interact with the adapters in the samples. Thus, different levels of repeating purification steps maybe needed to remove the excessive adapter dimers particularly in the untreated samples. However, this approach reduced the concentration of the library drastically. Thus, we proposed the changed the starting materials from total RNA to mRNA. Previously, Bustin and Nolan (2004) have suggested that used of enriched mRNA may help to improve the efficiency of qPCR on very low target mRNAs by improving primer-binding efficiency and improve the quantity of unrelated template present during primer/target annealing. This concept maybe applicable in the TREX since it also involves PCR step during library preparation. Furthermore, some of the target genes in the panel maybe just low/very low abundant in their mRNA copies. Our result indicated that using mRNA as the starting input could significantly yield in higher yield and purity. However, use of mRNA as starting material may still associate with several level of pitfall as bias may occur due to loss of some mRNA during isolation (Bustin & Nolan 2004). Although the usage of mRNA has not been established and this isolation step may lead to loss of some mRNA, the projection of number of reads using this sample could be much higher. Instead of repeating the cleaning step to obtain valid library results, isolating the mRNA may be a more cost and time-effective way of doing TREX.

CONCLUSION

TREX is a promising tool that could simplify researchers who are dealing with NGS-based assays. This assay could be used to validate large results in a timely manner. Nevertheless, there is a lot of optimization required before obtaining valid results. Library preparation is the most important step to ensure quality results and must be done thoroughly. Based on our results, there were multiple tweaks that needed to be done to the system such as; additional purification step, adjusting the AMPure beads ratio and changing the input sample to mRNA. This manuscript is intended to ease researchers who are planning to use the TREX assay in the future.

ACKNOWLEDGEMENTS

The authors wish to thank National Cancer Council Malaysia, (MAKNA) (vote no: 6300123) for funding this study. Authors would like to thanks Miss Boo Sook Yee (Science Vision, Malaysia), Miss Liew Woan Charn (IBS, UPM, Malaysia) and Dr Heidi Cheng (Illumina, Singapore) to provide technical support of this optimization. The authors declare that they have no conflict of interest.

REFERENCES

- Bustin, S.A. & Nolan, T. 2004. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. *J. Biomol. Tech.* 15: 155-166.

- Hashida, S., Yamamoto, H., Shien, K., Miyoshi, Y., Ohtsuka, T., Suzawa, K., Watanabe, M., Maki, Y., Soh, J., Asano, H., Tsukuda, K., Miyoshi, S. & Toyooka, S. 2015. Acquisition of cancer stem cell-like properties in non-small cell lung cancer with acquired resistance to afatinib. *Cancer Sci.* 106: 1377-1384.
- Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R. & Ordoukhanian, P. 2014. Library construction for next-generation sequencing: overviews and challenges. *BioTechniques* 56(2): 61-77.
- Hitzemann, R., Bottomly, D., Darakjian, P., Walter, N., Iancu, O., Searles, R., Wilmot, B. & McWeeney, S. 2013. Genes, behavior and next-generation RNA sequencing. *Genes Brain Behav.* 12: 1-12.
- Kellum, S.N., Glover, K., Myhre, A., Faranda, A. & Donner, M. 2015. Multi-endpoint evaluation of known apoptosis inducers and cytogenetic chemicals in the *in vitro* micronucleus assay with TK6 cells. In *Genetic Toxicology Association Meeting*, edited by Clayton, J.M. Hall Conference Center, University of Delaware: Current Developments in Genetic Toxicology. <http://www.gta-us.org/scimtg/2015Meeting/posters2015.html>.
- Long, Q., Xu, J., Osunkoya, A.O., Sannigrahi, S., Johnson, B.A., Zhou, W., Gillespie, T., Park, J.Y., Nam, R.K., Sugar, L., Stanimirovic, A., Seth, A.K., Petros, J.A. & Moreno, C.S. 2014. Global transcriptome analysis of formalin-fixed prostate cancer specimens identifies biomarkers of disease recurrence. *Cancer Res.* 74: 3228-3237.
- Patel, P.B., Thakkar, V.R. & Patel, J.S. 2015. Cellular effect of curcumin and citral combination on breast cancer cells: Induction of apoptosis and cellular cycle arrest. *J. Breast Cancer* 18: 225-234.
- Prestes, P., Marques, F., Curl, C., Lewandowski, P., Delbridge, L., Harrap, S. & Charchar, F. 2016. The effect of genes involved in monogenic human cardiomyopathies in a polygenic model of cardiac hypertrophy. *J. Hypertens.* 34: PS 01-07.
- Schneider, E., Dittrich, M., Bock, J., Nanda, I., Muller, T., Seidmann, L., Tralau, T., Galetzka, D., Hajj, N.E. & Haaf, T. 2016. CpG sites with continuously increasing or decreasing methylation from early to late human fetal brain development. *Gene* 592: 110-118.
- Simbolo, M., Gottardi, M., Corbo, V., Fassan, M., Mafficini, A., Malpeli, G., Lawlor, R.T. & Scarpa, A. 2013. DNA qualification workflow for next generation sequencing of histopathological samples. *PLoS ONE* 8: e62692.
- Wang, Z., Gerstein, M. & Snyder, M. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10: 57-63.
- Nadiah Abu, Noraini Nordin & Noorjahan Banu Alitheen
Department of Cell and Molecular Biology
Faculty of Biotechnology and Biomolecular Science
Universiti Putra Malaysia
43400 UPM Serdang, Selangor Darul Ehsan
Malaysia
- Nadiah Abu
UKM Molecular Biology Institute (UMBI)
UKM Medical Centre, Jalan Yaa'cob Latiff
Bandar Tun Razak, 56000 Cheras
Kuala Lumpur, Wilayah Persekutuan
Malaysia
- Swee Keong Yeap* & Sheau Wei Tan
Institute of Bioscience, Universiti Putra Malaysia
43400 UPM Serdang, Selangor Darul Ehsan
Malaysia
- Swee Keong Yeap*
China-ASEAN College of Marine Sciences
Xiamen University Malaysia
Jalan Sunsuria, Bandar Sunsuria
43900 Sepang, Selangor Darul Ehsan
Malaysia
- Wan Yong Ho
School of Biomedical Sciences
The University of Nottingham Malaysia Campus, Jalan Broga
43500 Semenyih, Selangor Darul Ehsan
Malaysia

*Corresponding author; email: skyeap2005@gmail.com

Received: 6 September 2016

Accepted: 2 August 2017