

NLR Gene Resequencing and Expression Measurement Using cDNA RenSeq

ABSTRACT

Key components of the plant pathogen defense system are disease resistance genes encoding intracellular immune receptors with nucleotide-binding and leucine-rich repeat (NLR) domains. Resistance gene transcript enrichment sequencing (cDNA RenSeq) is a cost-effective method of discovering and quantifying even lowly-expressed NLR genes in various tissues and genotypes. In this application note, we demonstrate the power of cDNA RenSeq using the myBaits® RNA-Seq capture system in identifying and quantifying expressed NLR gene transcripts in various wheat tissues. We show how the system uses a simple workflow to accurately, efficiently, and consistently retrieve comparable NLR gene information to transcriptome-wide RNA-Seq at a fraction of the cost.

WHY cDNA RenSeq

Due to significant diversifying selection, many NLR genes display extreme intraspecific diversity with ubiquitous presence-absence polymorphisms and sequence and structural variations, which makes it impossible to obtain full NLR gene repertoire from computationally predicted gene models [1, 2]. While genomic RenSeq focuses on obtaining NLR gene sequences, cDNA RenSeq, on the other hand, can obtain a better picture of functioning NLR genes by capturing expressed NLR transcripts and avoiding non-expressed paralogues. Furthermore, cDNA RenSeq enables the discovery of very lowly-expressed NLR genes, identifying novel transcripts and rare isoforms that were not annotated or incorrectly annotated by the automated gene prediction software [3]. Importantly, as a cost-effective alternative to deep transcriptome-wide RNA-Seq, cDNA RenSeq can also generate high-resolution quantitative datasets to assess changes in the abundance of NLRs genes. Undoubtedly, these advantages not only make cDNA RenSeq a useful tool to studying NLR gene diversity, but also an indispensable highly efficient approach to study the dynamics of NLR gene expression.

MATERIAL AND METHODS

To demonstrate cDNA RenSeq's accuracy and efficiency, we analyzed data generated using a custom Triticeae RenSeq bait-library containing 60,000 baits designed from predicted NLR genes present in eight Triticeae species (as described in Steuernagel *et al.* 2016 [4]). The myBaits® cDNA RenSeq workflow is illustrated in Figure 1A. In brief, RNA was extracted from different tissues of wheat at various growth stages. cDNA libraries were prepared with KAPA mRNA-Seq Kit following the manufacturer's protocol. Targeted enrichment was performed according to the standard myBaits protocol version 3. Both enriched, and non-enriched cDNA libraries were sequenced on an Illumina HiSeq 2500, and the downstream analytic bioinformatic pipeline was implemented as depicted in Figure 1B.

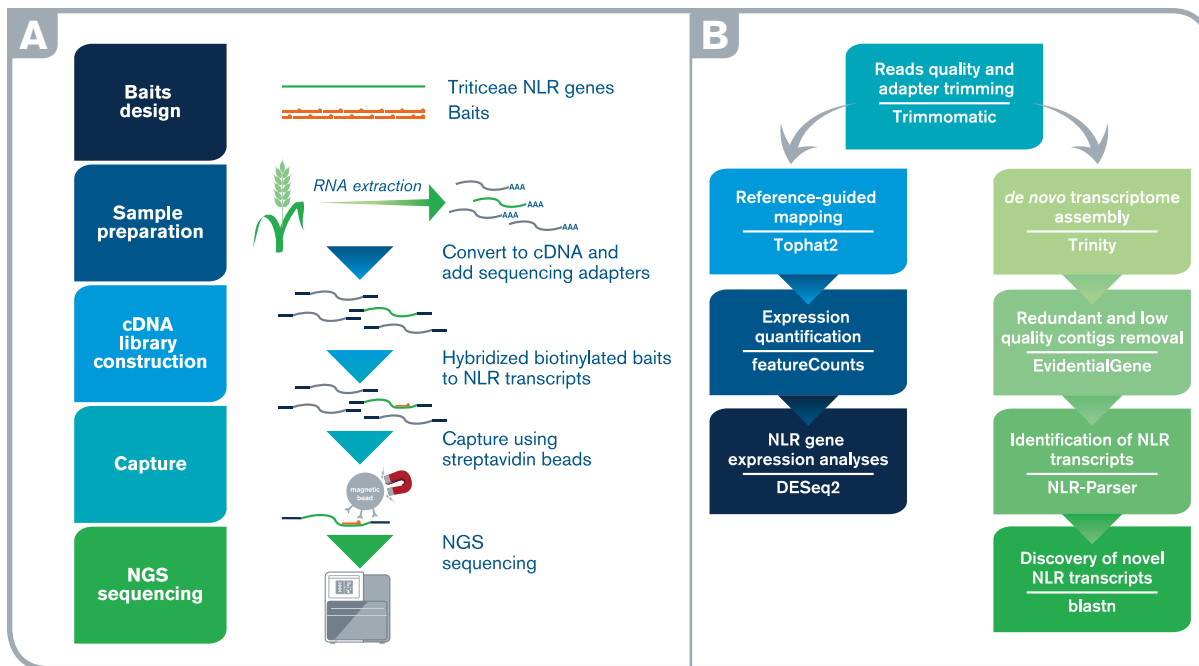


Figure 1. Simple and easy to implement. Schematic illustration of the myBaits® cDNA RenSeq workflow (A) and bioinformatic pipelines (B).

RESULTS/DISCUSSION

High enrichment efficiency of myBaits® cDNA RenSeq

To assess enrichment efficiency, we analyzed data reported in Steuernagel *et al.* 2018 [5]. Here, cDNA RenSeq libraries from young leaf samples were compared to their corresponding non-enriched RNA-Seq libraries (referred to as deep RNA-Seq here). Deep RNA-Seq of roughly 100 Gbp per sample only yielded 0.16-0.31% of reads aligning to NLR genes, while the cDNA RenSeq data yielded 47.46–55.26% of reads aligning (Figure 2). This roughly **222-fold enrichment rate** translates to an average **4.3X improvement** in NLR coverage depth using just 3% of the sequencing in deep RNA-Seq (Figure 4A).

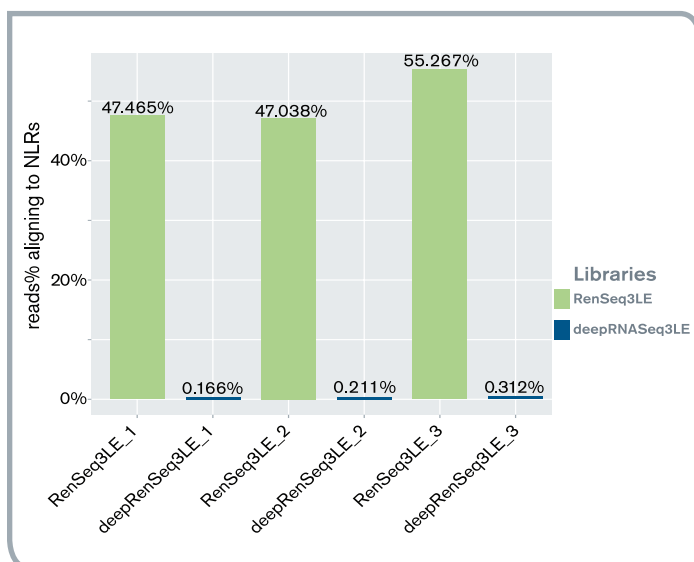


Figure 2. Orders of magnitude target enrichment. Percentage of reads aligning to NLR genes in cDNA RenSeq and RNA-Seq libraries.

To better illustrate the benefits of cDNA RenSeq, we included comparisons to coverage obtained after down-scaling the deep RNA-Seq data to the same sequencing depth as the cDNA RenSeq (referred to down-scaled RNA-Seq hereinafter). After down-scaling, the significant increase in read depth per transcripts achieved with cDNA RenSeq is even clearer. The median log2 read depth calculated from down-scaled RNA-Seq, deep RNA-Seq and cDNA RenSeq dataset were 0.3, 3.58 and 5.67 respectively (Figure 4A, D). In deep RNA-Seq, analysis of read depth changes reveals only an average of 24-fold increase gained with 30X deeper sequencing, while an average of 212-fold increase with switching to cDNA RenSeq (Figure 4B).

MORE NLR TRANSCRIPTS DETECTABLE AT COMPARABLE READ DEPTH

cDNA RenSeq is able to detect low-abundance NLR transcripts and simultaneously increase the read depth per transcript. As shown in Figure 3, cDNA RenSeq detected 319 (13%) more NLR transcripts than the deep RNA-Seq and 916 (49%) more than the down-scaled RNA-Seq.

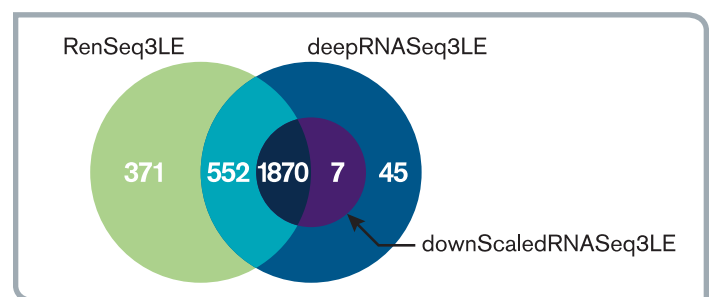
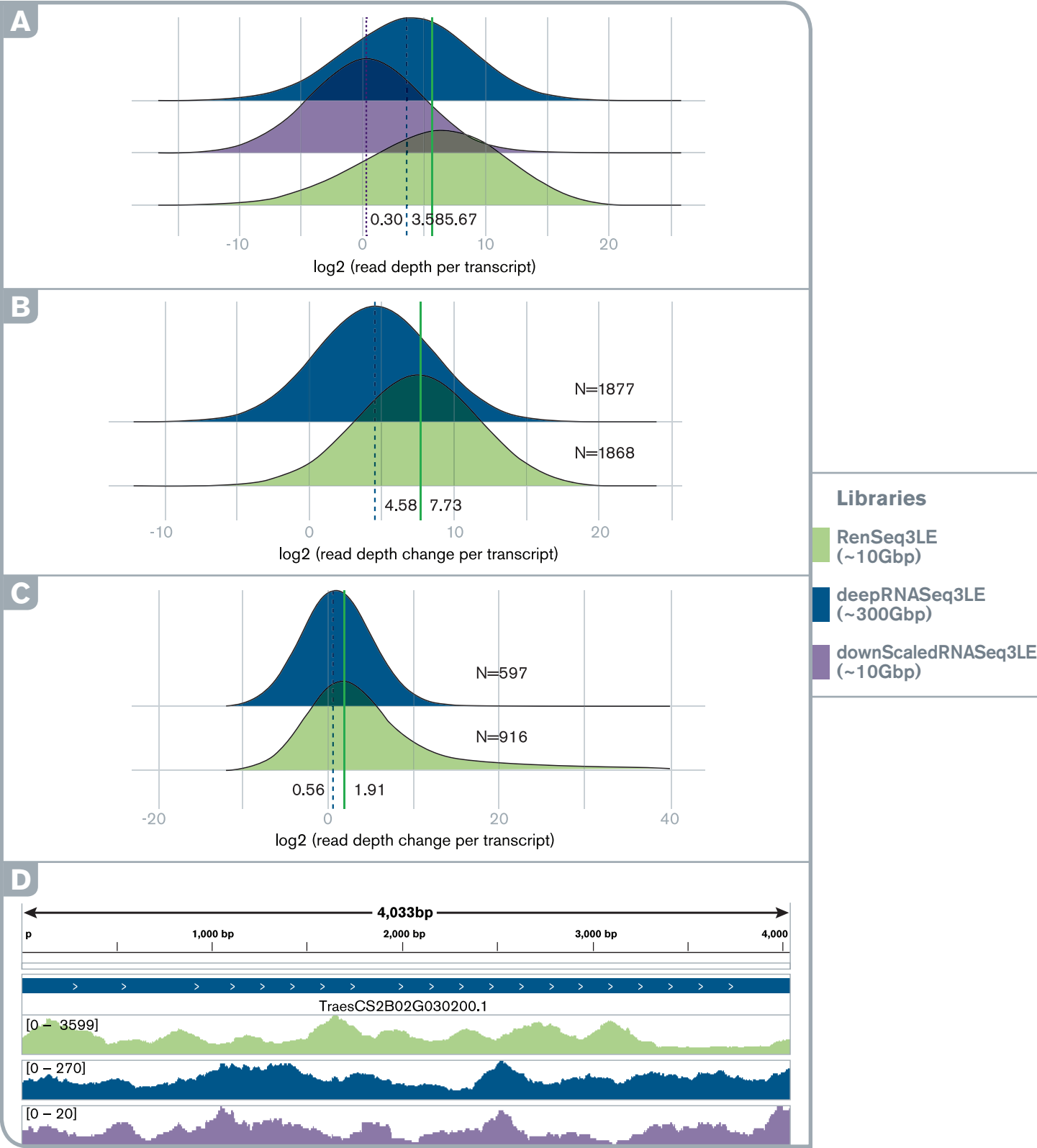


Figure 3. Discover lowly-expressed transcripts. Number of NLR transcripts detected in cDNA RenSeq, deep RNA-Seq and downScaled RNA-Seq libraries.

As for low-abundance transcripts that were not detected by down-scaled RNA-Seq, cDNA RenSeq not only was able to detect 53% more low-abundance transcripts but also achieved an average read depth 2.5-fold higher than deep RNA-Seq (Figure 4C). Taken together, it is substantially more efficient and cost-effective to use cDNA RenSeq rather than increasing RNA-Seq depth to retrieve NLR transcripts.

Figure 4. Significant improvement of read depth. (A) \log_2 read depth of NLR transcripts detected in each library. (B) \log_2 read depth change when switching to cDNA RenSeq versus increasing sequencing depth (C) \log_2 read depth of low-abundance NLR transcripts detected in cDNA RenSeq or deep RNA-Seq but not in downScaled RNA-Seq. N: number of NLR transcripts. (D) Visualization of read depth of an NLR gene locus in cDNA RenSeq, deep RNA-Seq and downScaled RNA-Seq.



HIGH EXPRESSION MEASUREMENT ACCURACY AND REPRODUCIBILITY

In order to examine how well enriched libraries reflect the expression levels measured with standard RNA-Seq, we investigated the correlation between log2 read counts (expression value) of NLR genes in the cDNA RenSeq datasets and their corresponding unenriched deep RNA-Seq datasets. Here we observed an excellent correlation ($R=0.92$; Figure 5A) in transcript relative coverage for a spectrum of expression levels.

In addition, hierarchical clustering analysis using the normalized log2 read counts of NLR genes reveals that the libraries cluster based on tissue type rather than sequencing method (direct vs. enriched), indicating that the capture introduces no systematic bias in expression values of NLR genes (Figure 5B). Reproducibility was tested by measuring the correlation coefficient of normalized log2 read counts of NLRs among biological replicates, which indicate a high average pairwise correlation $R = 0.95$. Altogether, the validation analyses demonstrate that myBaits® cDNA RenSeq is accurate and reproducible, and is a powerful method for efficient NLR expression profiling.

IDENTIFICATION OF FUNCTIONAL AND NOVEL NLR TRANSCRIPTS

As a case study, with a de novo assembly-based pipeline, the combined multi-tissue cDNA RenSeq successfully identified 115 novel NLR-associated transcripts that were not annotated in the reference genome, and 581 NLR-associated transcripts which have low similarity to annotated transcripts, demonstrating that it is highly efficient to employ myBaits® cDNA RenSeq for detecting novel NLR transcripts and/or rare isoforms that were not picked up by computational gene model prediction tools.

CONCLUSION

This note demonstrates the power of myBaits® cDNA RenSeq in retrieving functional NLR repertoire and the capacity to be used for quantitative gene expression applications. As a highly efficient, reliable and affordable approach, implementing cDNA RenSeq would undoubtedly increase the number of annotated NLR genes and will greatly accelerate functional study of disease resistance genes, thus establishing a solid foundation for more targeted and specific resistance breeding strategies for agronomically important species with more complex genomes.

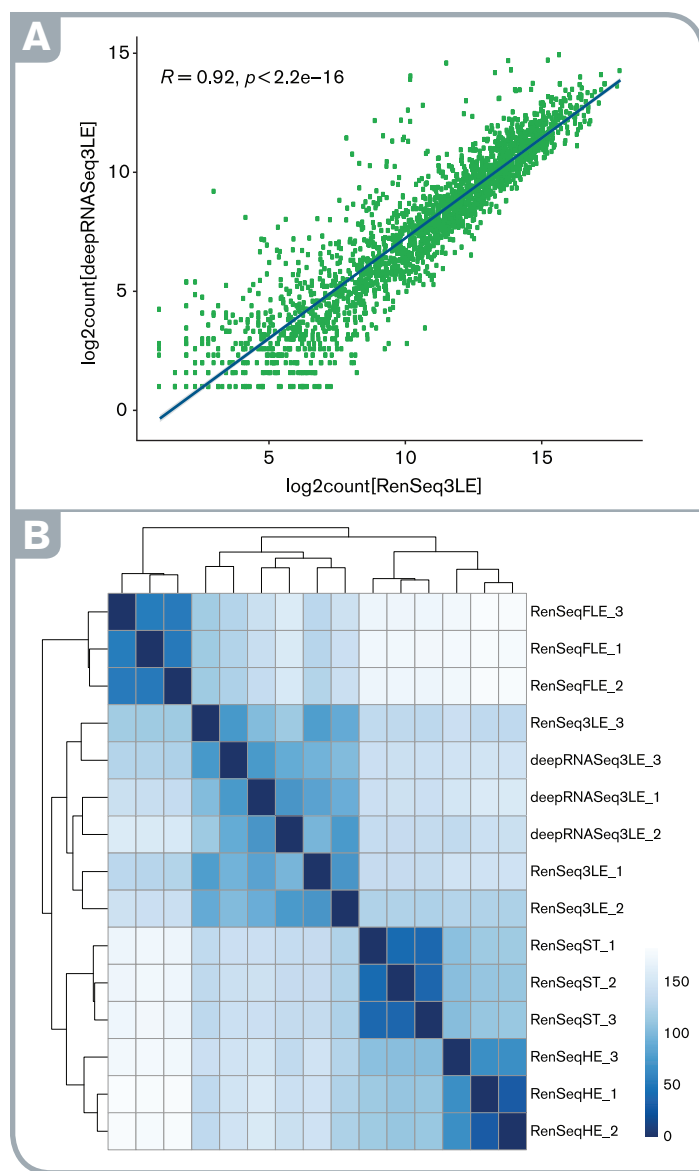


Figure 5. Accuracy and reproducibility of cDNA RenSeq. (A) A scatter plot of log2 read counts of NLR genes shows a high correlation between cDNA RenSeq and deep RNA-Seq libraries. (B) Hierarchical clustering of gene expression in different tissues.

REFERENCES

1. Van de Weyer, A. L. *et al.* (2019) *Cell*.
2. Ma, Y. *et al.* (2019) *BMC genomics*.
3. Andolfo, G. *et al.* (2014) *BMC plant biology*.
4. Steuernagel, B. *et al.* (2016) *Nature Biotechnology*.
5. Steuernagel, B. *et al.* (2018) *bioRxiv*.

