# Obtain superior NGS library performance with lower input amounts using the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina®

## Substantial advances in RNA strand-specific library construction in a **poly(A) mRNA enrichment** workflow

This is one of four technical notes available that address directional and non-directional RNA library preparation for both poly(A) mRNA enrichment and rRNA depletion workflows. Additional tech notes:

- NEBNext Ultra II Directional RNA Library Prep Kit in a rRNA depletion workflow
- Non-directional NEBNext Ultra II RNA Library Prep Kit in a poly(A) mRNA enrichment workflow
- Non-directional NEBNext Ultra II RNA Library Prep Kit in a rRNA depletion workflow

### Introduction

RNA-seq has become the most popular method for transcriptome analysis and is widely used to study gene expression, and to detect mutations, fusion transcripts, alternative splicing, and post-transcriptional modifications. It is becoming the method of choice to detect genetic alterations causing diseases, to provide insights on the various molecular pathways perturbed by changes in the transcriptome and to study their implications. As RNA-seq is adopted for this growing range of applications, the need for good quality, reproducible library preparation methods using very low amounts of RNA input, or precious clinical samples, is increasing.

To meet these challenges, we have reformulated each step of the RNA library prep workflow to create the **NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB #E7760)**. This new kit utilizes a fast, streamlined, automatable workflow for high yield production of superior quality libraries, from as little as 10 ng total RNA input in a poly(A) mRNA enrichment workflow.

Strand specificity is important for correct annotation of genes, identification of antisense transcripts with potential regulatory roles, and accurate determination of gene expression levels in the presence of antisense transcripts. Enhanced sensitivity to detect transcripts with uniform coverage across their length offers a non-biased approach for accurate quantitation of transcript levels.

For removal of ribosomal RNA, the kit is compatible with both poly(A) mRNA enrichment and rRNA depletion. **Here we demonstrate the utility of the NEBNext Ultra II Directional RNA Library Prep Kit for library construction in a poly(A) mRNA enrichment workflow, with a broad range of input amounts.**

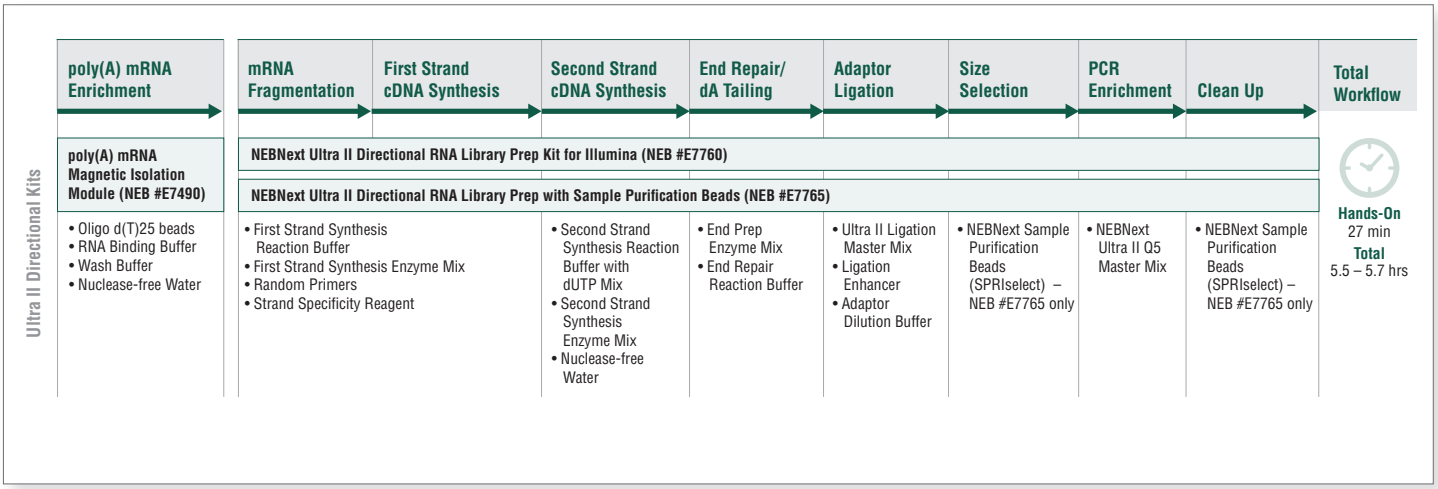
For information on performance in a ribosomal RNA depletion workflow, please refer to the separate technical note on that topic.

### The NEBNext Ultra II Directional RNA Workflow with poly(A) mRNA Enrichment

The workflow combines enrichment of mRNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #7490) and library construction using the NEBNext Ultra II Directional RNA Library Prep Kit (NEB #E7760/E7765).

The library prep kit's new reverse transcriptase master mix improves first strand synthesis, and the inclusion of Actinomycin D in the kit (in the Strand Specificity Reagent) increases the user-friendliness of this step. As in the Ultra II DNA kit (NEB #E7645), combining the End Repair and dA-Tailing steps and minimizing clean up steps makes the workflow both fast (~ 6 hours) and easy to use (Figure 1). The protocol can accommodate 10 ng to 1 µg of total RNA for the poly(A) mRNA enrichment workflow. As little as 1 ng of previously isolated mRNA can be used directly with the NEBNext Ultra II Directional RNA Library Prep Kit. The protocol is compatible with adaptors and primers from the NEBNext product line ("NEBNext Oligos") or from other sources.

FIGURE 1:
**NEBNext Ultra II Directional RNA workflow with poly(A) mRNA enrichment**



## Library Yields

One measure of the success of library preparation is the yield of the final library. The NEBNext Ultra II Directional RNA Kit produces substantially higher yields compared to other commercially available kits (Figure 2), and compared to NEB's original Ultra Directional RNA kit (Figure 3). The increased reaction efficiencies with the Ultra II kit mean that sufficient library yields can be obtained even with low input amounts, and with fewer PCR cycles.

FIGURE 2: **NEBNext Ultra II Directional RNA produces the highest yields, from a range of input amounts**

Poly(A)-containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000), and libraries were made using the NEBNext Ultra II Directional RNA Kit (plus the NEBNext Poly(A) mRNA Magnetic Isolation Module), Kapa Stranded mRNA-Seq Kit, Kapa mRNA HyperPrep Kit and Illumina TruSeq Stranded mRNA Kit. The input RNA amount and number of PCR cycles are indicated. Library yields from an average of three replicates are shown.

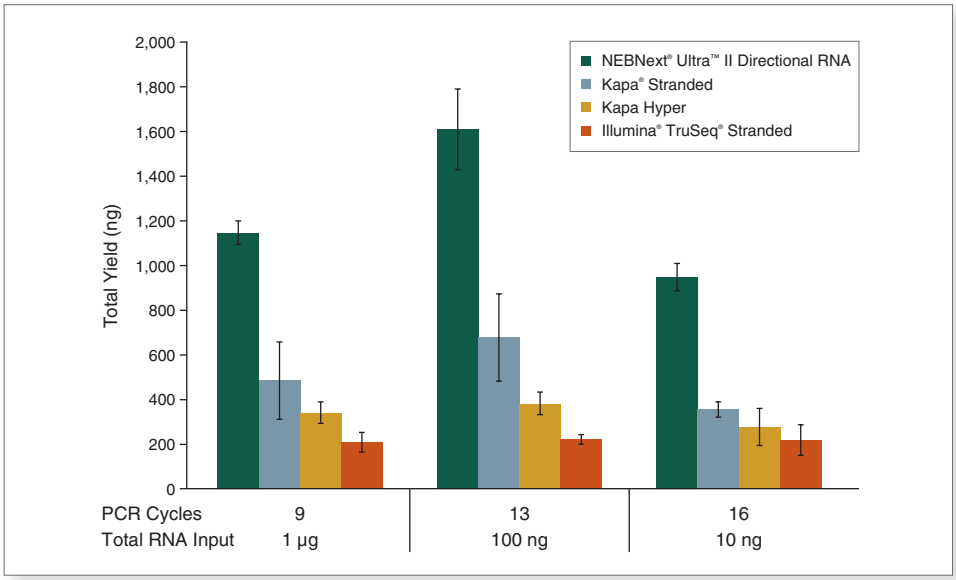
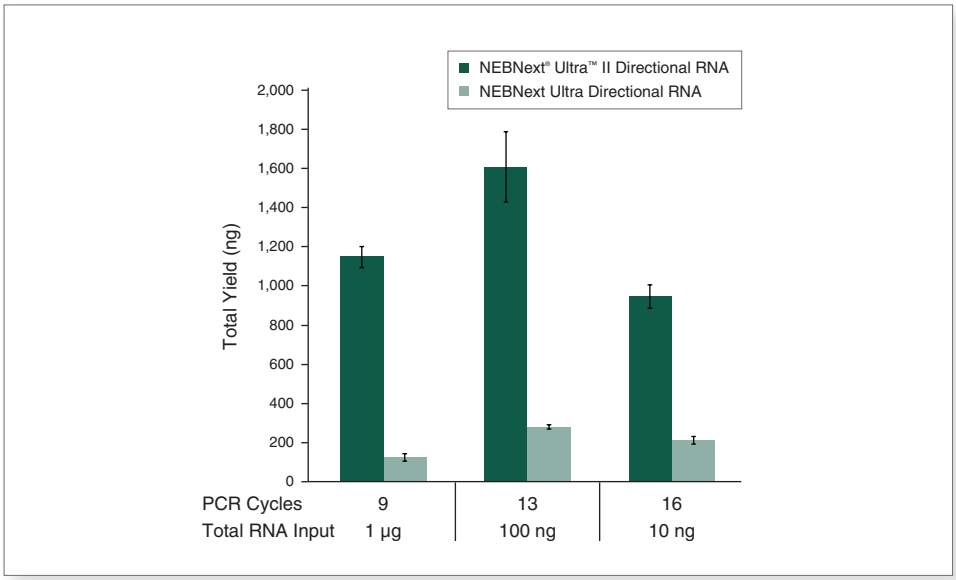



FIGURE 3: **NEBNext Ultra II Directional RNA produces yields several fold higher than the original Ultra Directional RNA Kit**

Poly(A) containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000) using the NEBNext Poly(A) mRNA Magnetic Isolation Module, and libraries were made using either the Ultra II Directional RNA Kit or the original Ultra Directional RNA Kit. Significantly higher yields were achieved with the Ultra II Directional RNA Kit than with the original Ultra Directional RNA Kit. The input RNA amount and number of PCR cycles are indicated. Library yields from an average of three replicates are shown.



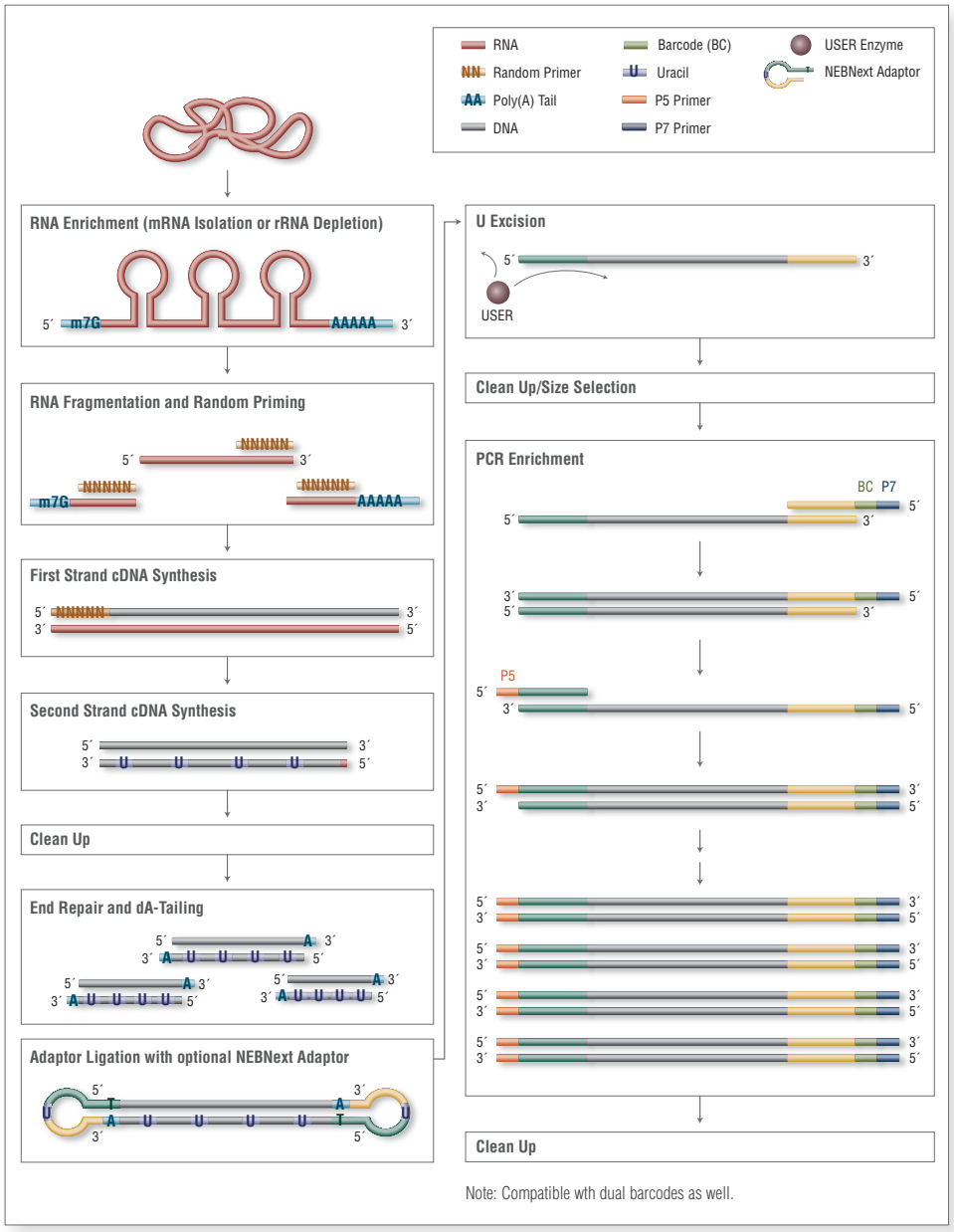
## Library Quality

While sufficient yield of a library is required for successful sequencing, quantity alone is not enough. The quality of a library is also critical, regardless of the input amount or GC content of the sample RNA. A high-quality library will have uniform representation of the RNA of interest in the original sample, as well as even coverage across the GC spectrum.

## Library Directionality/ Strand-specificity

The NEBNext Ultra II Directional RNA Library Prep Kit derives its directionality from the "dUTP method" for strand-specificity, the proven gold standard method for this application (1,2). Labeling of the second strand cDNA using dUTP enables subsequent selective destruction of that strand, with the result that only one strand is incorporated into the final library, thus providing directionality as depicted in Figure 4.

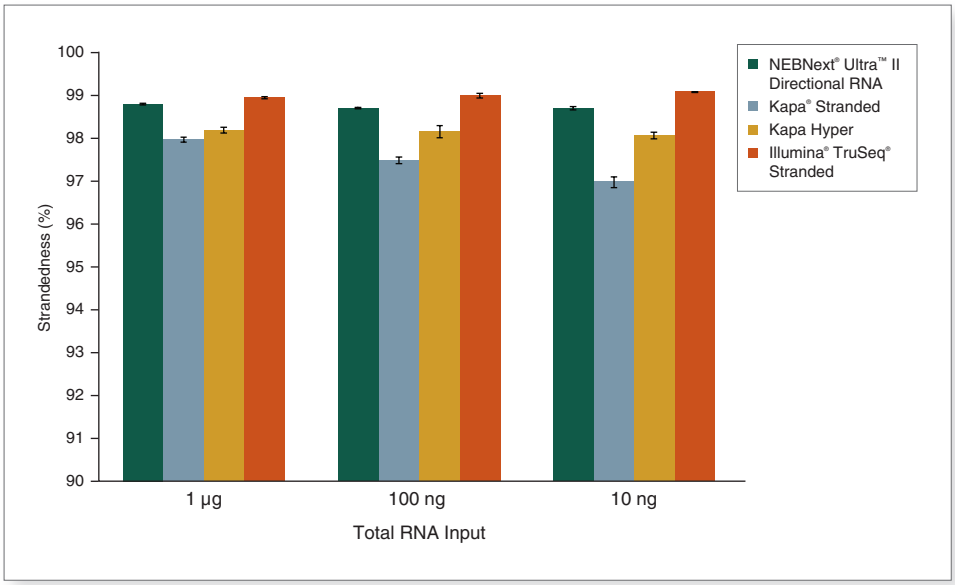
FIGURE 4:
**NEBNext Ultra II Directional RNA Library Prep Kit workflow**




1. Parkomchuk, D., et al. (2009) *Nucleic Acids Res.* 37. e123.
2. Levin, J. Z., et al. (2010) *Nature Methods* 7. 709–715.

FIGURE 5: **NEBNext Ultra II Directional RNA libraries have high directionality/strandedness**

Poly(A)-containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000), and libraries were made using the NEBNext Ultra II Directional RNA Kit (plus the NEBNext Poly(A) mRNA Magnetic Isolation Module), Kapa Stranded mRNA-Seq Kit, Kapa mRNA HyperPrep Kit and Illumina TruSeq Stranded mRNA Kit. Libraries were sequenced on an Illumina NextSeq® 500 using paired-end mode (2x76 bp). Reads were mapped to the hg19 reference genome using Hisat 2.0.3 and directionality was calculated using the infer experiment tool from the RSeQC Quality Control Package.

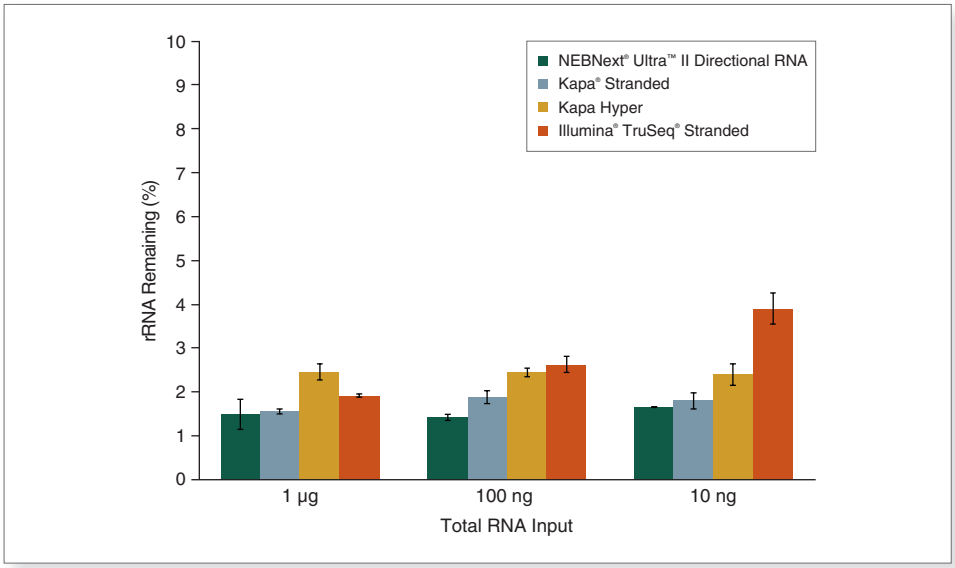



## Levels of Ribosomal RNA Remaining After poly(A) mRNA Enrichment

Ribosomal RNAs (rRNAs) are extremely abundant, constituting 80–90% of total RNA, and so one measure of the efficiency of enrichment of poly(A) mRNAs is the level of rRNA present in the sample after enrichment. Low levels of rRNA remaining in the sample is an indicator of the efficiency and specificity of the mRNA isolation method.

FIGURE 6: **NEBNext Ultra II Directional RNA with NEBNext Poly(A) mRNA isolation results in the lowest remaining ribosomal RNA levels**

Poly(A)-containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000). Libraries were made using the NEBNext Ultra II Directional RNA Kit (plus the NEBNext Poly(A) mRNA Magnetic Isolation Module), Kapa Stranded mRNA-Seq Kit, Kapa mRNA HyperPrep Kit and Illumina TruSeq Stranded mRNA Kit. Libraries were sequenced on an Illumina NextSeq 500 using paired-end mode (2x76 bp). Read pairs were assessed to be ribosomal RNA (rRNA) if they contain 6 or more 32 base matches to 18S, 28S, 5S, 5.8S, 16S or 12S human rRNA sequences (mirabait 4.9). Percent rRNA remaining was calculated by dividing rRNA reads by the total number of reads passing instrument quality filtering. Average percent rRNA remaining is shown for three replicates.



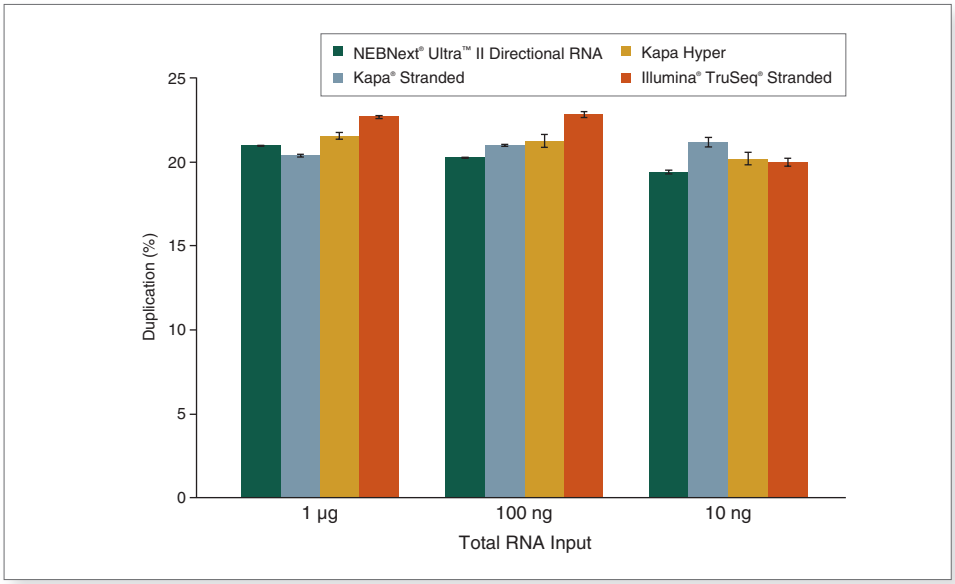
## Duplication Rates

"Duplication rates" refers to the percentage of sequence reads that have identical start and end sites. These have most often arisen from preferential PCR amplification of the same molecule. Low quality libraries can result in high duplication rates, and duplication rates are often greater with libraries constructed from very low input amounts.

The low duplication rates achieved with the NEBNext Ultra II Directional RNA Library Prep Kit, even with the high yields obtained (see Figure 2) indicate the high quality of the libraries produced, and the opportunity to minimize PCR cycles.

FIGURE 7: **NEBNext Ultra II Directional RNA with NEBNext Poly(A) mRNA isolation results in lower duplication rates**

Poly(A)-containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000). Libraries were made using the NEBNext Ultra II Directional RNA Kit (plus the NEBNext Poly(A) mRNA Magnetic Isolation Module), Kapa Stranded mRNA-Seq Kit, Kapa mRNA HyperPrep Kit and Illumina TruSeq Stranded mRNA Kit. Libraries were sequenced on an Illumina NextSeq 500 using paired-end mode (2x76 bp). Reads were down sampled to 9 million read pairs and mapped to the hg19 reference genome. Duplication rates were calculated as a fraction of uniquely mapped reads using the 'Read Duplication' tool of RSeQC where reads mapping to the same genomic location are regarded as duplicated reads.

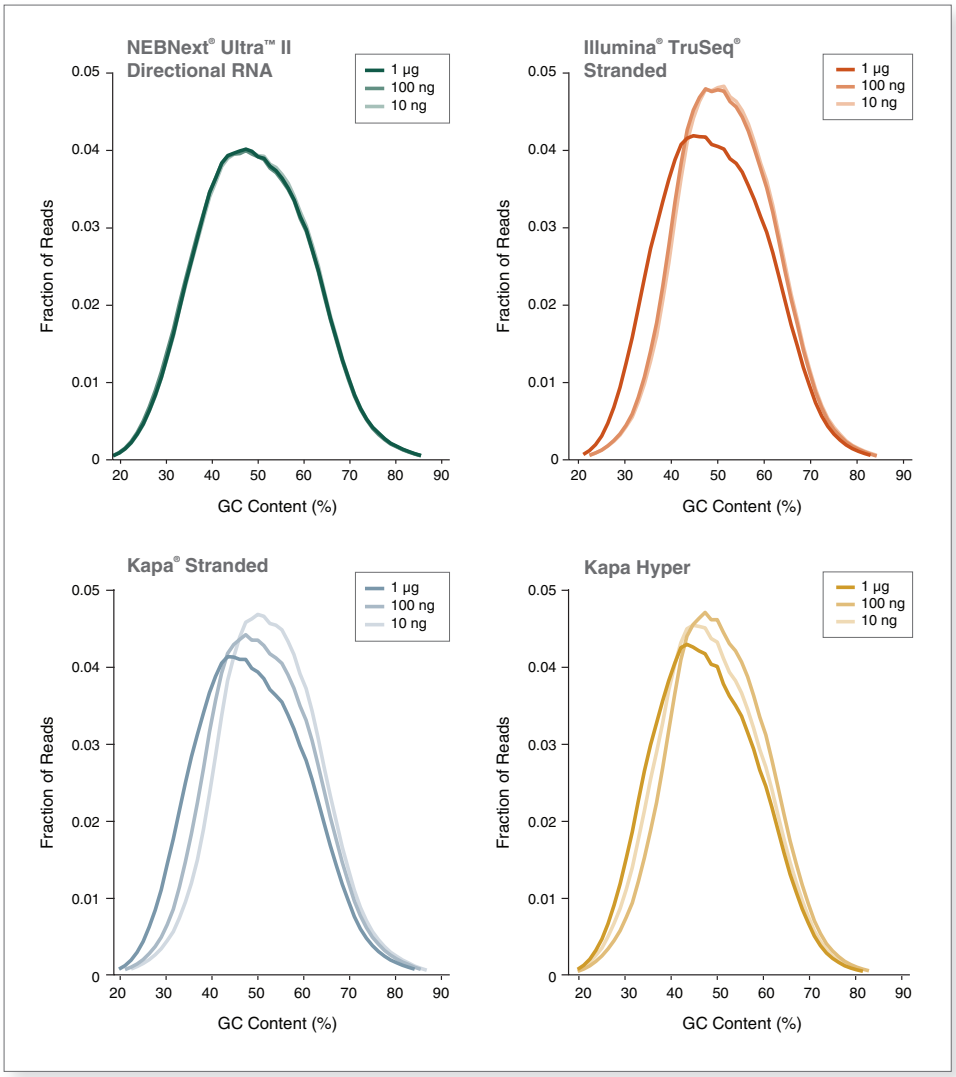


### Uniformity of GC Content Distribution

During the entire library construction workflow, and especially when amplification is required to obtain sufficient library yields, it is important to ensure that no bias is introduced, and that representation of GC-rich and AT-rich regions is not skewed in the final library. Uniformity of GC representation can be more challenging to maintain with lower input amounts, as is demonstrated in the figure to the right.

FIGURE 8: **NEBNext Ultra II Directional RNA libraries provide uniform GC content distribution, at a broad range of input amounts**

Poly(A)-containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000), and libraries were made using the NEBNext Ultra II Directional RNA Kit (plus the NEBNext Poly(A) mRNA Magnetic Isolation Module), Illumina TruSeq Stranded mRNA Kit, Kapa Stranded mRNA-Seq Kit and Kapa mRNA HyperPrep Kit. Libraries were sequenced on an Illumina NextSeq 500 using paired-end mode (2x76 bp). Reads were mapped to the hg19 reference genome. GC content distribution for each library was calculated using mapped reads. Ultra II Directional RNA libraries had uniform GC content distribution across a range of input amounts, whereas for other kits the GC content distribution changed with different input amounts, indicating the introduction of input-dependent sequence bias.

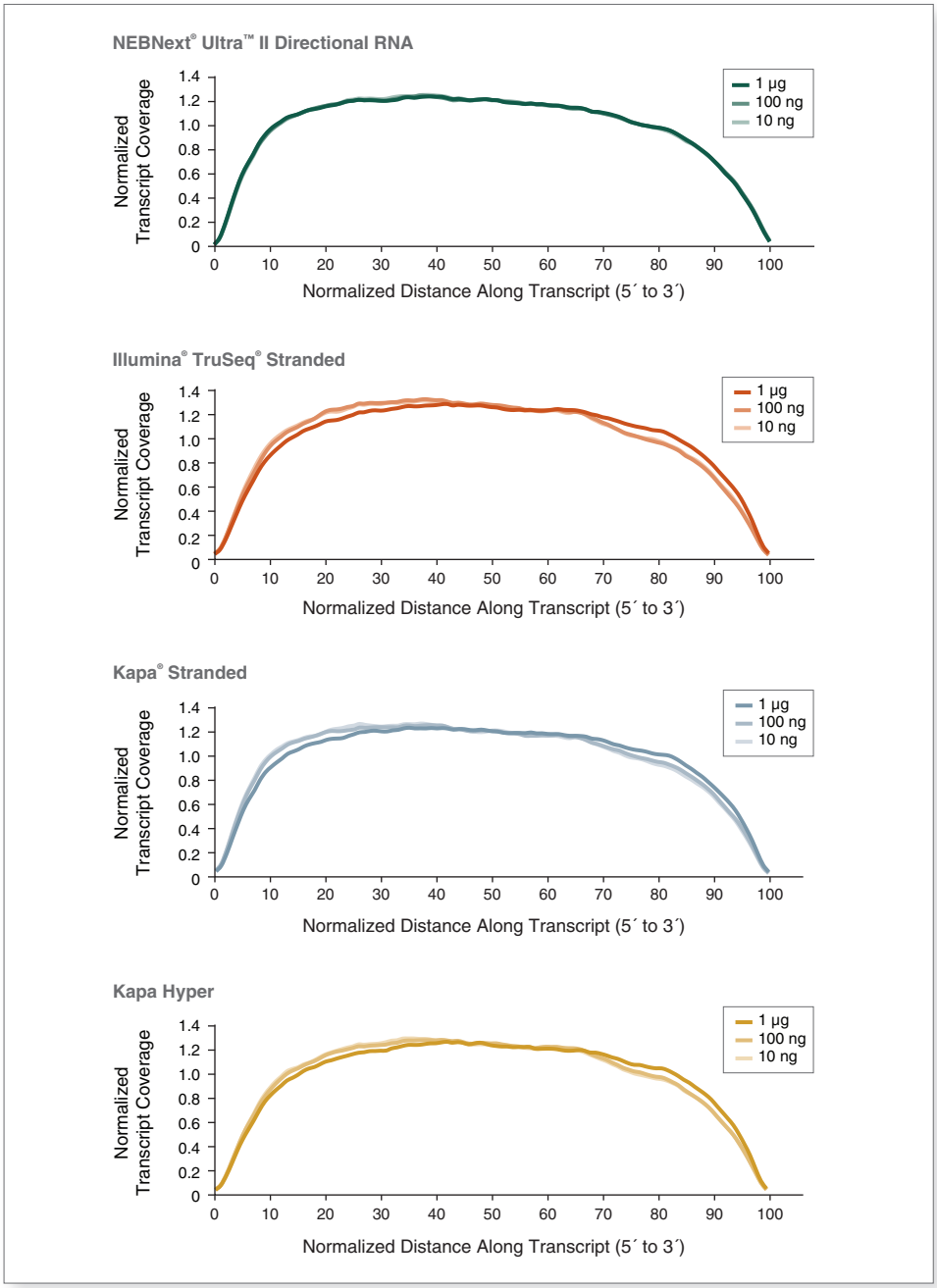


## Uniformity of Transcript Coverage

A high-quality library will not only include all transcripts from the original sample, but also cover those transcripts completely from 5' to 3'. Transcript coverage can be examined on a global basis (gene body), and by looking at individual transcripts. This can highlight differences between transcript coverage at different input amounts, and between different library kits. The use of ERCC standards, a set of RNA controls developed by the External RNA Controls Consortium (ERCC) and consisting of known, unlabeled, polyadenylated transcripts, is another useful tool in this type of experiment.

FIGURE 9: **NEBNext Ultra II Directional RNA libraries provide uniform coverage across the gene body of transcripts**

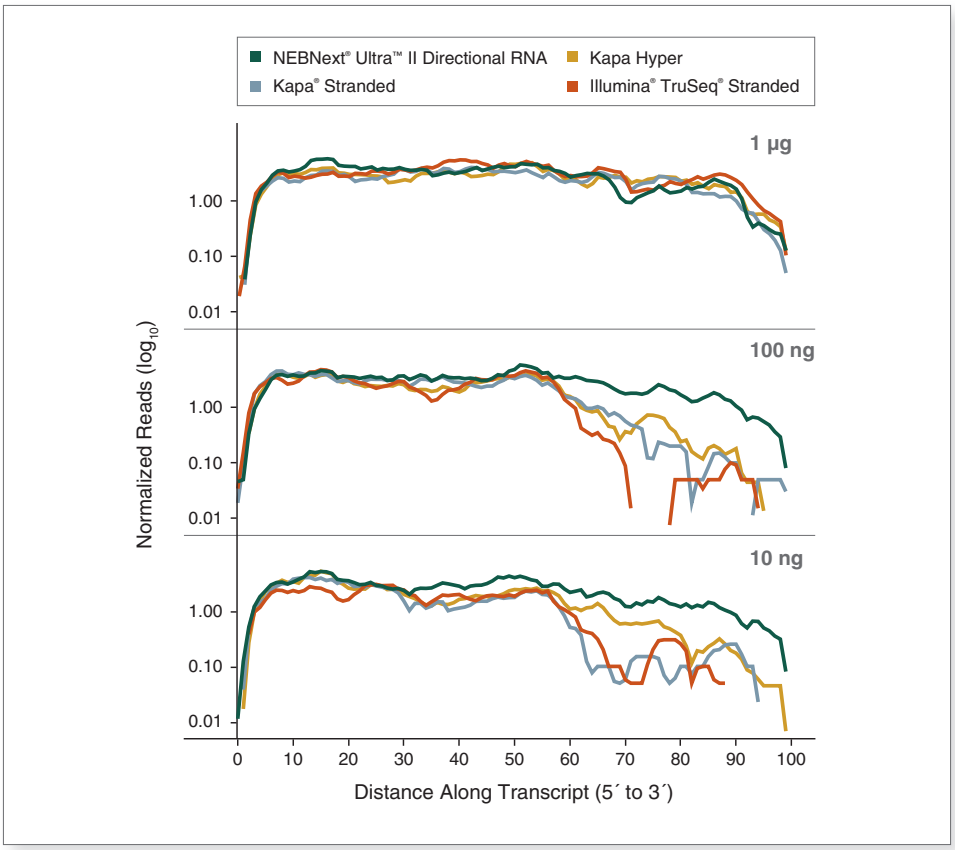
Poly(A)-containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000), and libraries were made using the NEBNext Ultra II Directional RNA Kit (plus the NEBNext Poly(A) mRNA Magnetic Isolation Module), Illumina TruSeq Stranded mRNA Kit, Kapa Stranded mRNA-Seq Kit and Kapa mRNA HyperPrep Kit. Libraries were sequenced on an Illumina NextSeq 500 using paired-end mode (2x76 bp). This view of the 5´ to 3´coverage of RefSeq transcripts (3) reveals consistent coverage for Ultra II Directional RNA libraries as input RNA is decreased from 1 µg to 10 ng. The changes apparent in other kits result from loss of coverage at the 3´ end of some transcripts.




3. O'Leary,N. A., et al. (2016) *Nucleic Acids Res.* 44 (D1), D733–D745.

FIGURE 10: **Uniformity of Coverage across the DAM1 transcript**

Poly(A)-containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000), and libraries were made using the NEBNext Ultra II Directional RNA Kit (plus the NEBNext Poly(A) mRNA Magnetic Isolation Module), Kapa Stranded mRNA-Seq Kit, Kapa mRNA HyperPrep Kit and Illumina TruSeq Stranded mRNA Kit. Libraries were sequenced on an Illumina NextSeq 500 using paired-end mode (2x76 bp). Coverage across the length of this individual transcript (ENST00000369541.3; DAM1) was assessed by mapping reads directly to the GENCODE v25 transcripts and examining 100 bins along the transcript length. NEBNext Ultra II Directional RNA libraries provided coverage across the entire length of the transcript, even as input was decreased from 1 µg to 10 ng.



## FIGURE 11: **Uniformity of Coverage across ERCC spike-ins**

Poly(A)-containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000) with recommended amounts of ERCC RNA Spike-In Mix I (Thermo Fisher Scientific #4456740), and libraries were made using the NEBNext Ultra II Directional RNA Kit (plus the NEBNext Poly(A) mRNA Magnetic Isolation Module), Illumina TruSeq Stranded mRNA Kit, Kapa Stranded mRNA-Seq Kit and Kapa mRNA HyperPrep Kit. Libraries were sequenced on an Illumina NextSeq 500 using paired-end mode (2x76 bp). A global view of the transcript coverage of spiked-in ERCCs was determined by mapping reads directly to the ERCC sequences and assessing coverage using bedtools in 100 bins along the transcript length. Superior consistency of coverage between the range of library preparation input amounts (10 ng, 100 ng, 1 µg) was seen with the Ultra II Directional RNA Kit.

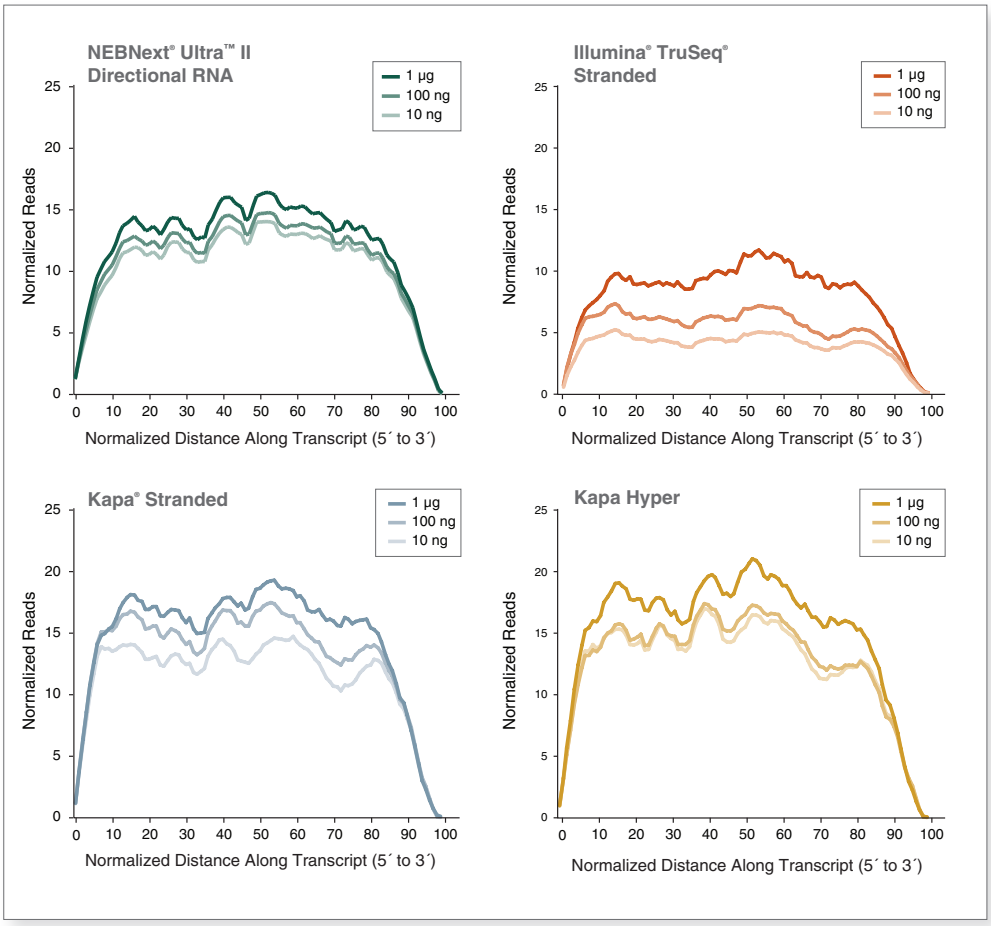
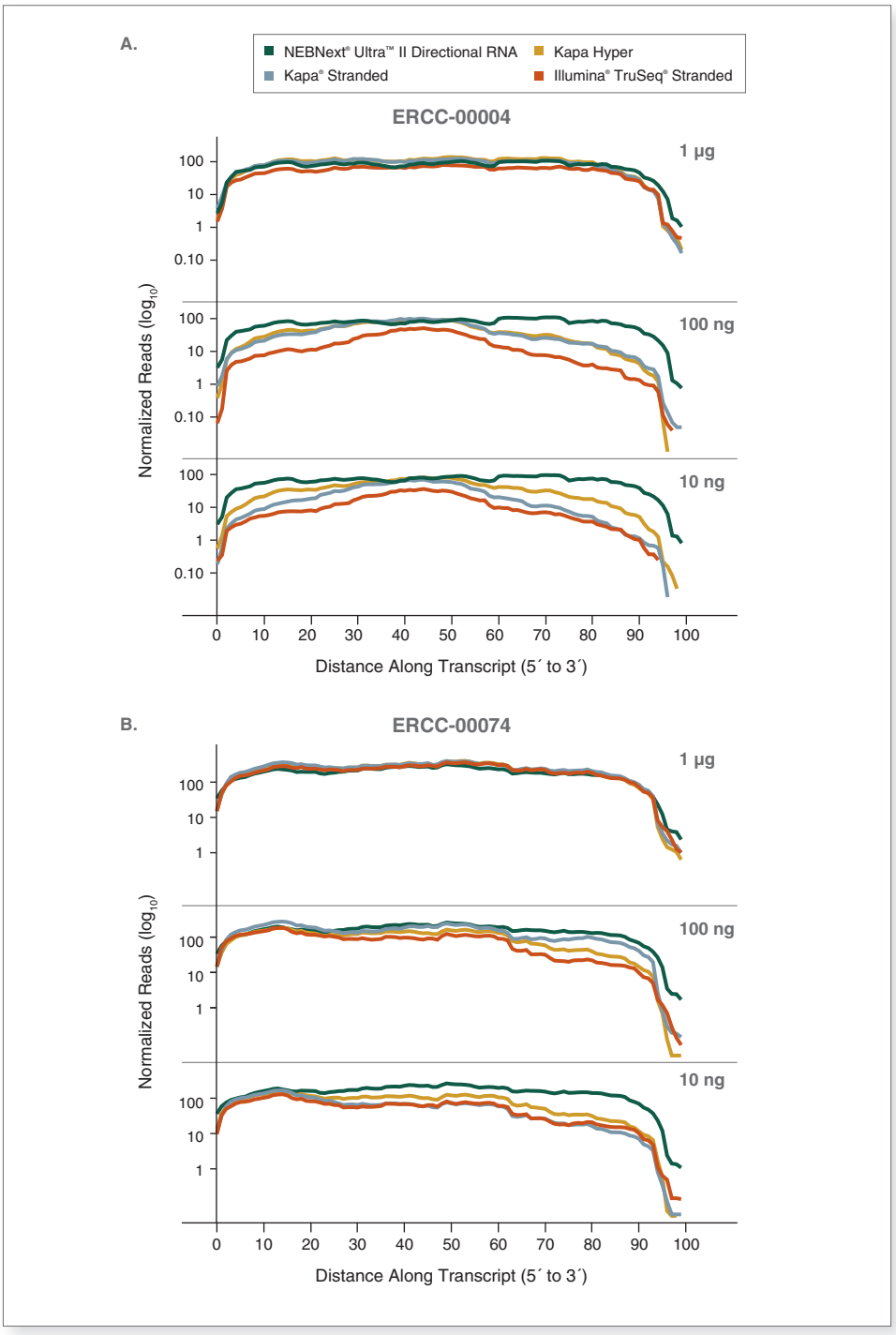


FIGURE 12: **Uniformity of Coverage across ERCC-00004 and ERCC-00074**

Coverage across transcripts ERCC-00004 (A) and ERCC-00074 (B) was assessed by mapping reads directly to the ERCC sequences and assessing coverage using bedtools cov in 100 bins along the transcript length. Libraries prepared using the NEBNext Ultra II Directional RNA Kit provided superior coverage across the transcripts.

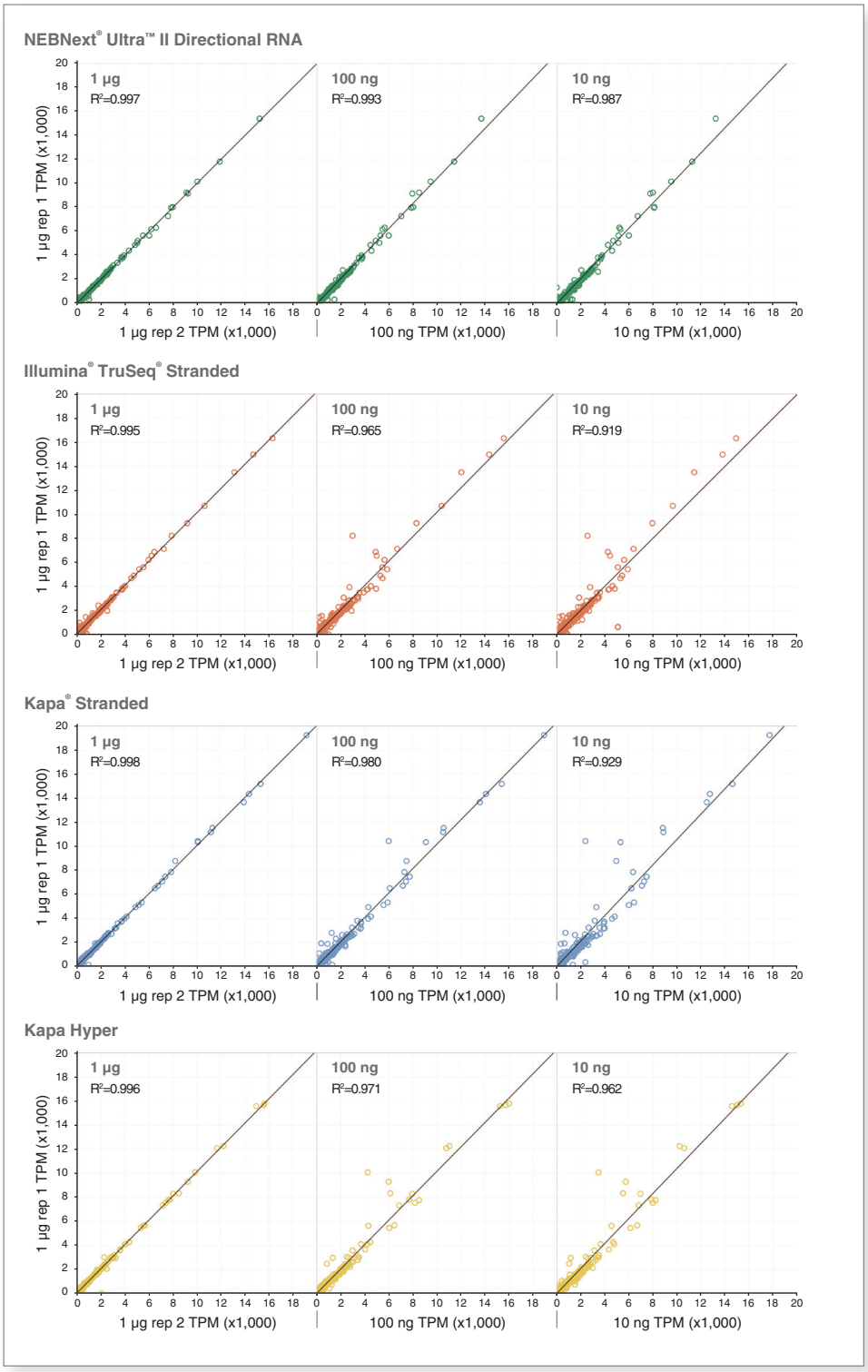


## Superior Library Complexity

As described above, an ideal library will represent completely and proportionally the sequence of the input RNA. When library preparation is inefficient or when input amounts for a library are very low, there is a risk that the resulting library will lack this diversity, and that some sequences will be over- or under-represented. Comparison of transcript abundance with libraries constructed from different input amounts of RNA is a useful measure to determine the effect of input amounts on coverage. The increased efficiency of each step in the NEBNext Ultra II Directional RNA library workflow improves the consistency of the composition of a library as input amounts are decreased from 1 µg to 100 ng and 10 ng, for Universal Human Reference RNA (Figure 13) and ERCC standards (Figure 14).

FIGURE 13: **Low input NEBNext Ultra II Directional RNA libraries retain superior complexity**

Poly(A)-containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000), and libraries were made using the NEBNext Ultra II Directional RNA Kit (plus the NEBNext Poly(A) mRNA Magnetic Isolation Module), Illumina TruSeq Stranded mRNA Kit, Kapa Stranded mRNA-Seq Kit and Kapa mRNA HyperPrep Kit. Libraries were sequenced on an Illumina NextSeq 500 using paired-end mode (2x76 bp). Salmon 0.4.0 was used for read mapping and quantification of all GENCODE v25 transcripts. TPM = Transcripts Per Kilobase Million. $R^2$ values for the linear fit are shown. Correlation analysis of the transcripts indicates superior transcript expression correlation between the different inputs for Ultra II Directional RNA libraries.



## Conclusion

The NEBNext Ultra II Directional RNA Library Prep Kit for Illumina represents a substantial advance in strand-specific library preparation for RNA sequencing in conjunction with poly(A) mRNA enrichment. Improved reagents and workflow steps increase the efficiencies of each step, and enable users to overcome many of the challenges previously associated with successful library preparation, such as:

- The use of input amounts of total RNA from low nanograms to 1 microgram
- Generation of higher yields, with the use of fewer PCR cycles
- Uniformity of transcript coverage, and high library complexity, even at very low input amounts
- Uniform GC coverage of the sample
- Fast, streamlined library preparation that is automation-friendly

For performance data and other information on the NEBNext Ultra II Directional RNA Library Prep Kit in rRNA depletion workflows, see the separate application note.

FIGURE 14: **ERCC transcript correlation**

Poly(A)-containing mRNA was isolated from Universal Human Reference RNA (Agilent #740000) with recommended amounts of ERCC RNA Spike-In Mix I (Thermo Fisher Scientific #4456740), and libraries were made using the NEBNext Ultra II Directional RNA Kit (plus the NEBNext Poly(A) mRNA Magnetic Isolation Module), Illumina TruSeq Stranded mRNA Kit, Kapa Stranded mRNA-Seq Kit and Kapa mRNA HyperPrep Kit. Libraries were sequenced on an Illumina NextSeq 500 using paired-end mode (2x76 bp). Salmon 0.4.0 was used for read mapping and quantification of all GENCODE v25 transcripts. TPM = Transcripts Per Kilobase Million. $R^2$ values for the linear fit are shown. Correlation analysis of the transcripts indicates superior transcript expression correlation between the different inputs for ERCC spike-ins. The curved fit is a consequence of log transformation.