APPLICATION NOTE

Enrichment of coding transcripts in degraded samples using hybrid probe capture

Consistent and sensitive detection of RNA transcripts in FFPE samples

- Reliable detection of low-level transcripts
- Consistent, highly uniform gene coverage
- Robust performance in low-quality samples

illumina®

Introduction

RNA transcripts are dynamically expressed and highly informative about the activity and condition of tissues at the cellular level. RNA sequencing (RNA-Seq) with next-generation sequencing (NGS) is a powerful method for profiling and quantifying expressed RNA transcripts. Unlike other technologies, sequencing provides significant advantages in multiplexing and parallel integration when identifying new variants and even gene fusions. However, at times the most interesting samples are damaged, low-quality RNA isolated from challenging, archival, and formalin-fixed, paraffin-embedded (FFPE) tissue. FFPE samples are common in disease research, especially oncology where biopsies taken for histological analysis are later used for genetic studies. These samples can be difficult to process due to the small quantities involved and nucleic acid degradation that commonly results from fixation and storage conditions.[1,2]

To address this challenge and enable delivery of reproducible results, Illumina offers the Illumina RNA Prep with Enrichment and TruSeq™ RNA Exome library preparation solutions. The Illumina RNA Prep with Enrichment protocol is optimized for use with 10-100 ng of purified total RNA from high-quality samples, or 20-100 ng RNA input from degraded or FFPE (DV_{200} ≥ 36.5%) samples.[3] Illumina RNA Prep with Enrichment has fewer steps than the well-established TruSeq RNA Exome and it requires roughly 50% less sample preparation time. This technical note highlights the performance of Illumina RNA Prep with Enrichment when working with FFPE samples with DV_{200} from 32% to 66% (Table 1). Comparisons with TruSeq RNA Exome are shown to demonstrate reliability and consistency across both assays.

Materials and methods

Illumina scientists recommend the use of DV_{200} as the quality indicator when preparing RNA for RNA-Seq analysis from FFPE samples. An explanation of DV_{200} and details for easy calculation using the value using the Agilent Bioanalyzer instrument can be found in the technical note, *Expression Analysis of FFPE Samples*.[4] For this technical note, sequencing performance between samples will be discussed based on their respective DV_{200} scores. Samples analyzed are summarized in Table 1.

Table 1: FFPE RNA samples analyzed

Tissue	Sample ID, supplier[a]	DV_{200}	RIN[b]
FFPE thyroid	12889, Amsbio	32.0%	2.3
FFPE breast	3468, BioOptions	51.0%	2.2
FFPE breast	3886, BioPartners	58.0%	1.6
FFPE breast	3967, BioPartners	66.0%	1.1

a. Illumina internal sample IDs for unique human biopsy tissues from listed suppliers
b. RNA integrity scores shown for comparison to DV_{200} values

Illumina RNA Prep with Enrichment (Catalog no. 20040537) with the Illumina Exome Panel (Catalog no. 20020183) and TruSeq RNA Exome (Catalog no. 20020189) assays were carried out following manufacturer protocols.[3,5] A workflow summary for the enrichment protocols is shown in Figure 1.

Sequencing was performed on a NovaSeq™ 6000 Sequencing System. Paired-end sequencing runs were performed with 2 × 101 bp reads with a minimum of 25M clusters per sample. Transcript coverage and alignment results were generated using the Illumina RNA-Seq Alignment App v1.1.1 through BaseSpace™ Sequence Hub. Expression analysis results were generated using the DRAGEN™ RNA pipeline v3.7.5 and reported as relative abundance of assayed transcripts in units of transcripts per million (TPM). Sequencing details are summarized in Table 2.

Table 2: Sequencing parameter summary

Library preparation	Sequencing	Analysis
Illumina RNA Prep with Enrichment + Exome Panel or TruSeq RNA Exome	NovaSeq 6000 system 2 × 101 paired-end reads 25M clusters	RNA-Seq Alignment App v1.1.1 DRAGEN RNA Pipeline v3.7.5

Figure 1: TruSeq RNA Exome and Illumina RNA Prep with Enrichment workflows—On-Bead Tagmentation and a single, 90 minute hybridization step combine to deliver a faster workflow with fewer steps when Illumina RNA Prep with Enrichment in comparison with TruSeq RNA Exome

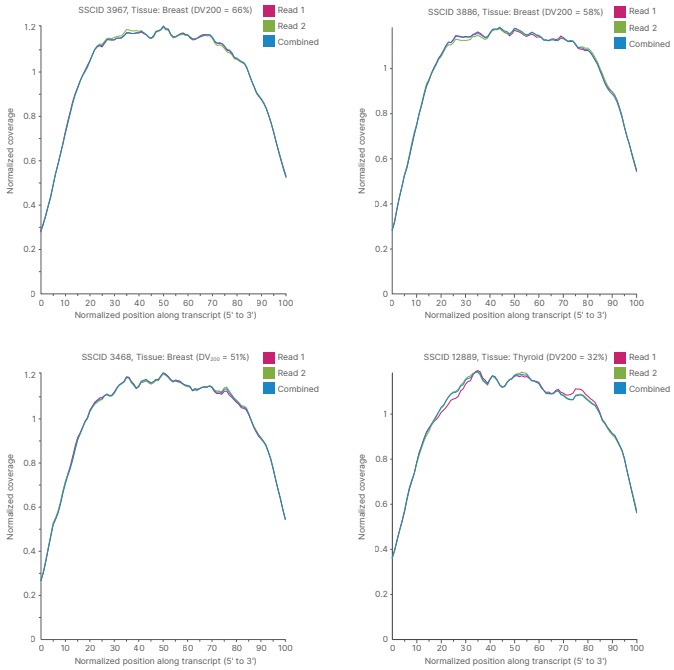

Figure 2: Highly uniform transcriptome coverage in FFPE samples—Plots demonstrate transcript coverage for the top 1000 most highly expressed genes. The X-axis shows normalized positions 1-100 for transcript lengths. The Y-axis shows normalized coverage, calculated by taking the coverage at each normalized position and dividing by the mean coverage. Perfect coverage would have a normalized coverage value of 1.0.

Results

RNA-Seq coverage and specificity

Illumina RNA Prep with Enrichment uses enrichment bead-linked transposomes optimized for RNA (eBLTL) and expertly designed probes to capture specific sequences with minimal background (Figure 1). The use of probes targeting the entire transcript eliminates the 3′ bias that is common with degraded samples using polyA selection and provides more uniform coverage of full transcripts. To evaluate the coverage performance of the Illumina RNA Prep with Enrichment, four FFPE samples with DV_{200} values ranging from 32% to 66% were analyzed. Using the minimum recommended 20 ng input, we observed uniform normalized coverage across all transcripts, regardless of sample quality (Figure 2).

A challenge with all RNA-Seq focused on coding transcripts is measuring mRNA expression while avoiding noncoding RNA species like ribosomal RNA. Ribosomal RNA makes up most of the RNA content in cells and has the potential to overwhelm sequencing reads. The exome-specific probe designs used with Illumina RNA Prep with Enrichment and TruSeq RNA Exome provide the necessary specificity to achieve a high percent of alignment to coding targets while maintaining a low percentage of ribosomal RNA sequences (Figure 3). Both assays deliver similar, robust results across FFPE samples with DV_{200} values ranging from 32% to 66% and using RNA input amounts of 20 ng, 50 ng, or 100 ng.

Gene detection and discovery

Illumina RNA Prep with Enrichment and TruSeq RNA Exome perform well for detection of both known and novel isoforms in degraded samples. Low-level transcripts were reliably detected in FFPE samples with consistent coverage for genes detected at both ≥ 1× and ≥ 10×. Results were similar at 20 ng, 50 ng, or 100 ng RNA input (Figure 4).

Illumina RNA Prep with Enrichment was further assessed for the ability to identify novel and known gene fusions in degraded samples. We were able to identify an *ARMT1-ESR1* fusion in all replicates—three each, at 20 ng, 50 ng, and 100 ng—of the SCID 3967, breast cancer RNA sample (data not shown). Similarly, we detected an *ESR1-CCDC170* fusion in all nine replicates of the SCID 3886, breast cancer RNA sample at 20 ng, 50 ng, and 100 ng (data not shown). These data demonstrate the sensitivity and versatility of Illumina RNA Prep with Enrichment for detection and discovery of transcripts.

Quantification of transcripts

Illumina RNA Prep with Enrichment libraries also maintain quantitative information for expressed transcripts when using degraded samples with DV_{200} values ranging from 32% to 66%. Gene expression values for libraries generated with either 20 ng or 100 ng of RNA input were highly correlated within each sample tested in this study (Figure 5).

Samples with DV_{200} < 36%

Illumina RNA Prep with Enrichment is optimized for RNA-Seq applications in FFPE tissues with quality scores ≥ 36%. In this application note, the data presented for thyroid tissue sample SCID 12889 is usable despite having a DV_{200} = 32%, which is modestly below the recommended lowest threshold. However, as RNA quality drops further, users should expect results to become more variable and eventually unusable for some applications.

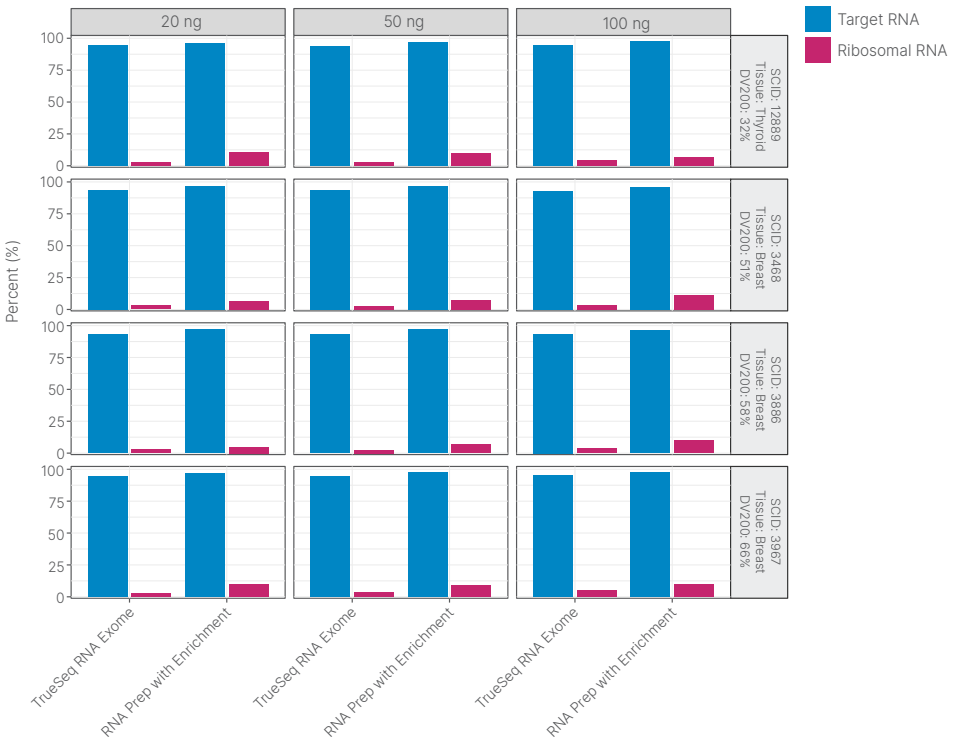

Figure 3: Consistent alignment between library preparation solutions with low ribosomal RNA abundance across FFPE sample quality and input quantity—TruSeq RNA Exome and Illumina RNA Prep with Enrichment solutions successfully enrich a high proportion of targets that align with the exome (% Target RNA) and low background from ribosomal RNA (% Ribosomal RNA). Results were consistent for the tested FFPE samples with DV200 scores ranging from 32% to 66%, and with RNA input of 20 ng, 50 ng, or 100 ng.

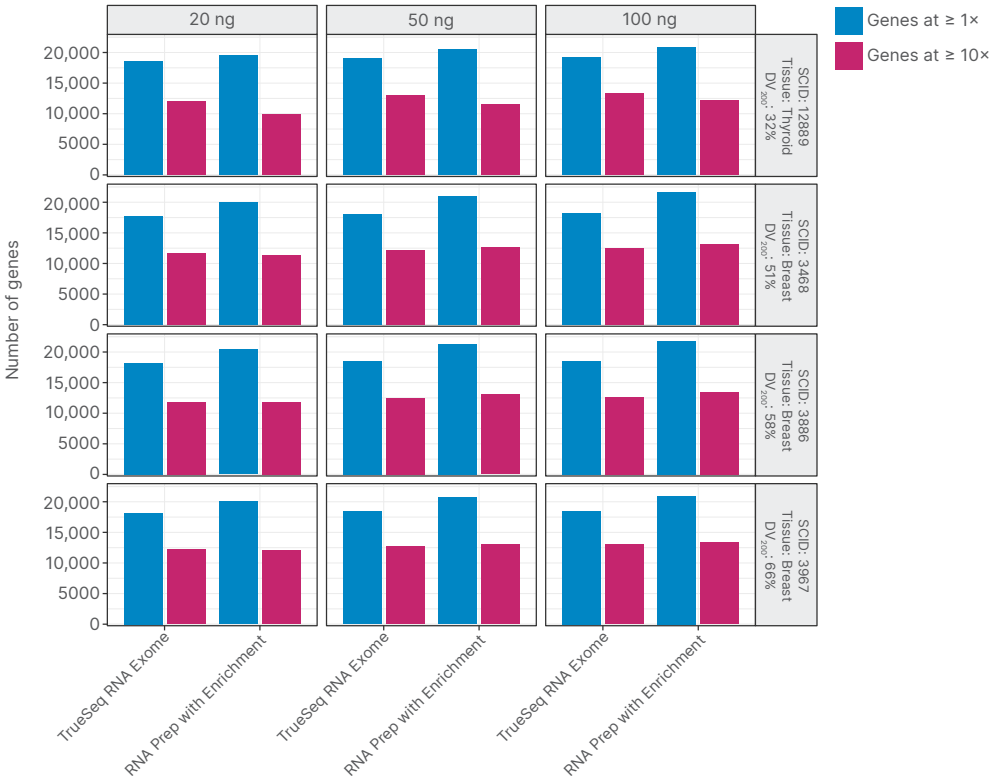

Figure 4: High confidence gene detection with greater gene discovery efficiency— TruSeq RNA Exome and Illumina RNA Prep with Enrichment can detect transcripts, even at low expression levels, in FFPE samples with DV_{200} scores from 32% to 66%, and RNA input of 20 ng, 50 ng, or 100 ng.

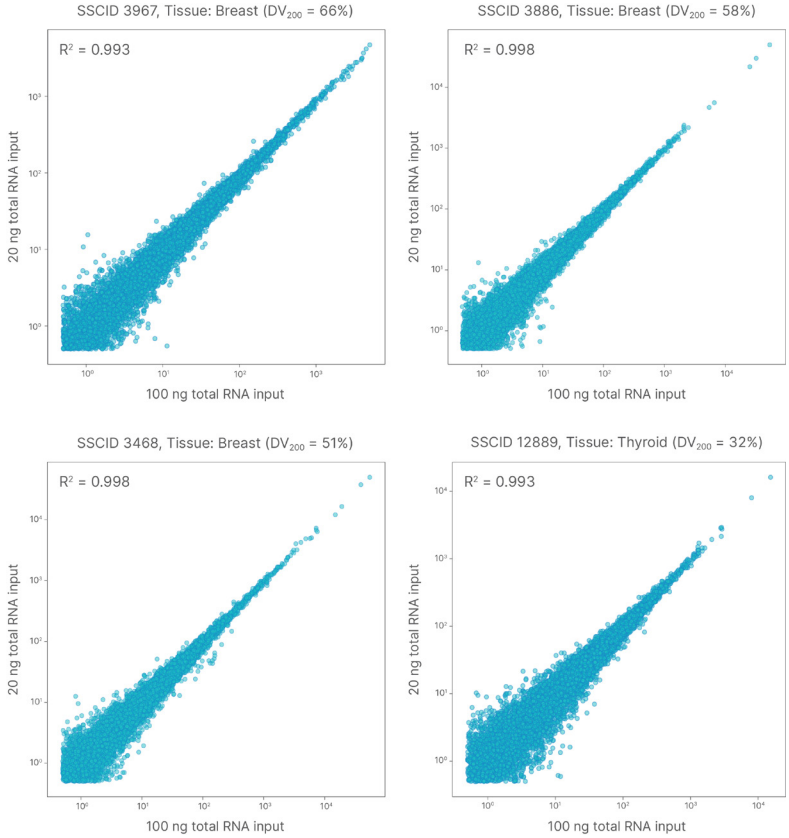

Figure 5: High concordance in gene expression at 20 ng and 100 ng inputs across varying quality FFPE samples— Gene expression quantification using Illumina RNA Prep with Enrichment is robust in FFPE samples with DV_{200} scores from 32% to 66%, down to 20 ng of RNA input. Expression data are plotted in units of TPM. R^2 for TPM values is very high when FFPE samples are sequenced with either 20 ng or 100 ng of RNA input.

Despite the challenges, researchers may be able to get usable data from valuable samples, even those with very low DV_{200} quality values. We analyzed FFPE skin tissue (DV_{200} = 9%) and FFPE thyroid tissue (DV_{200} = 27%) and found some sequencing metrics were impacted by decreasing RNA quality. However, improvement in the data was also possible by increasing sample input from 20 ng to 100 ng for measures, including gene expression (Figure 6), high confidence gene detection (Figure 7), and duplicate reads (data not shown).

The impressive performance of the Illumina RNA Prep with Enrichment methods and reagents allows better than expected data from these very low-quality samples. For experimental applications, it would be preferable to obtain samples with higher RNA quality, but it may still be possible to produce valuable information when samples cannot be easily replaced. It is up to the researcher to determine usability in these cases.

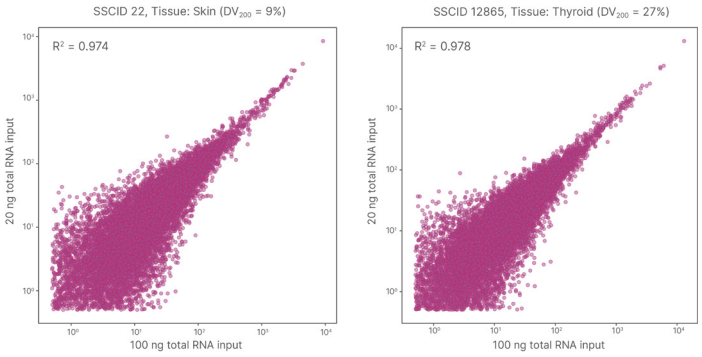

Figure 6: Sample gene expression at 20 ng and 100 ng inputs for FFPE samples with very low DV_{200} quality scores—Gene expression using Illumina RNA Prep with Enrichment on samples with DV_{200} scores as low as 9% can provide meaningful results. There is more variability (especially in low-expressed genes) in very low-quality samples compared to DV_{200} ≥ 32%, but R^2 values remain high between 20 ng and 100 ng of total RNA input.

Summary

FFPE samples are valuable sources of genetic information, particularly for cancer studies. However, their quality is highly variable due to both FFPE fixation protocols and varying storage conditions. For RNA-Seq studies, it is important to assess the sample quality and to have a sequencing solution that is validated for work with degraded samples.

Both Illumina RNA Prep with Enrichment and TruSeq RNA Exome offer reliable performance in RNA-Seq applications using FFPE degraded samples with DV_{200} scores ≥ 36%, and possibly lower as we observed with the SCID 12889 thyroid sample, DV_{200} = 32%.

While both solutions work well, Illumina RNA Prep with Enrichment has some advantages over TruSeq RNA Exome, especially with regards to workflow. Illumina RNA Prep with Enrichment uses optimized eBLTL for tagmentation, requires fewer steps, and takes approximately 50% less time than TruSeq RNA Exome (Figure 1). Tagmentation uses eBLTL to fragment and add adapter sequences in a single reaction step, avoiding ligation steps. For enrichment assays, tagmentation-based methods consistently show high library conversion with few primer dimers.[6]

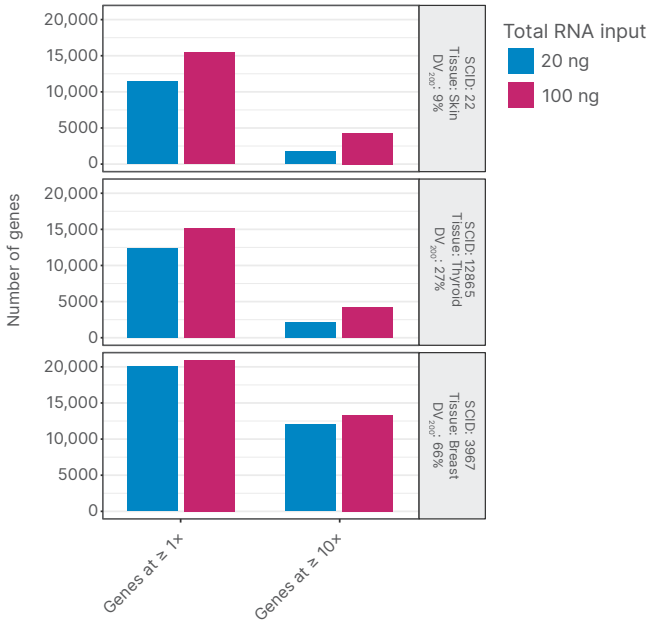

Figure 7: Sample data for high confidence gene detection at 20 ng and 100 ng inputs for FFPE samples with very low DV_{200} quality scores—When using RNA with very low DV_{200} quality scores of 9% and 27%, Illumina RNA Prep with Enrichment is able to perform well at gene detection as indicated by genes detected at ≥ 1× and ≥ 10×. Gene detection was improved when increasing from 20 ng to 100 ng of total RNA input. Higher-quality SCID 3697, DV_{200} = 66%, is shown for comparison.

References

1. von Ahlfen S, Missel A, Bendrat K, Schlumpberger M. Determinants of RNA quality from FFPE samples. *PLoS One*. 2007;2(12):e1261. Published 2007 Dec 5. doi:10.1371/journal.pone.0001261
2. Penland SK, Keku TO, Torrice C, et al. RNA expression analysis of formalin-fixed paraffin-embedded tumors. *Lab Invest*. 2007;87(4):383-391. doi:10.1038/labinvest.3700529.
3. Illumina. Illumina RNA Prep with Enrichment (L) Tagmentation Reference Guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/illumina_prep/RNA/illumina-rna-prep-reference-guide-1000000124435-03.pdf. Accessed June 23, 2021.
4. Illumina. Expression Analysis of FFPE Samples technical note. https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/technote-expression-analysis-ffpe-samples.pdf. Accessed June 23, 2021.
5. Illumina. TruSeq RNA Exome Reference Guide. https://support.illumina.com/downloads/truseq-rna-exome-reference-guide-1000000039582.html. Accessed June 23, 2021.
6. Illumina. Improved detection of circulating transcripts https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-rna-enrichment-crna-app-note-470-2020-009/illumina-rna-enrichment-crna-app-note-470-2020-009.pdf. Accessed July 15, 2021.

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-00214 v1.0