CORALL Total RNA-Seq Library Prep Kit

- Complete coverage of transcripts from start to end
- Ready-to-sequence libraries within 4.5 hours
- Unique Molecular Identifiers (UMIs) seamlessly included
- Excellent protocol-inherent strandedness (>99 %)
- 1 ng to 1 µg of total RNA input
- Easy all-in-one protocol

Introduction

CORALL is Lexogen's new stranded total RNA library prep kit with excellent whole transcriptome coverage. CORALL enables streamlined generation of Illumina-compatible libraries within 4.5 hours, featuring seamless integration of Unique Molecular Identifiers (UMIs) and exceptional protocol-inherent strand specificity (>99 %). The fragmentation-free protocol uses Lexogen's proprietary Strand Displacement Stop and Ligation technologies to deliver complete transcript representation, including start and end sites.

Workflow

CORALL libraries can be prepared from as little as 1 ng of total RNA input. Flexible input types are supported including, rRNA-depleted, poly(A)-enriched, or total RNA from a wide variety of species, as well as degraded and FFPE RNA samples.
CORALL library generation is initiated by random hybridization of Displacement Stop Primers (DSP) with partial Illumina-compatible P7 sequences, to the RNA template (Fig. 1). No prior RNA fragmentation is necessary, as the insert size is determined by the distance between hybridized DSPs. Reverse transcription extends each DSP to the next, where transcription is effectively stopped. This stop prevents spurious second strand synthesis, maintaining excellent strand specificity. Highly efficient ligation of Linker Oligos to the 3' ends of first-strand cDNA fragments then introduces partial Illumina-compatible P5 sequences and UMIs.

During PCR, second strand synthesis is performed, and the double-stranded cDNA is amplified. In doing so, i7 and (optional) i5 indices as well as complete adapter sequences required for cluster generation on Illumina instruments are added. All purification steps are based on magnetic beads, rendering the protocol highly suitable for automation.

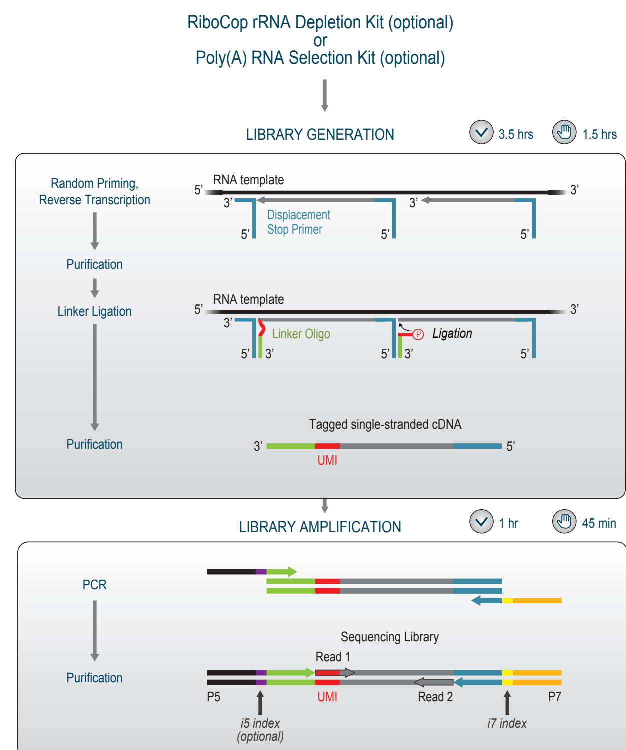

Figure 1 | Schematic overview of the CORALL library preparation workflow.

Performance

CORALL demonstrates excellent coverage and sequencing performance, with high strandedness, ERCC input-output correlations and read alignment rates (Tab. 1). CORALL has been tested with various RNA input types from human, mouse, mini-pig, hamster, plant to bacteria, as well as degraded and FFPE RNA samples.

	Value
Strandedness	>99.5 %
ERCC Pearson R	0.949
Coverage CoD [a] ERCCs	0.120
Total Aligned Reads	93.4 %
Protein Coding [b]	91.6 %
Antisense + lincRNA [b]	3.9 %

Table 1 | Paired-end mapping data for CORALL libraries. Libraries were prepared from ~2 ng of rRNA-depleted UHRR (RiboCop V1.2 Kit, Lexogen) and sequenced on a NextSeq 500 (PE150). Reads were aligned to the human reference genome GRCh38.94 and quantified with Mix² version 1.4.0.12 (Lexogen) [1, 2].
[a] Coverage CoD describes the deviation of the experimentally observed coverage from the theoretical coverage (lower value corresponds to smaller deviation). [b] The percentage of protein coding, antisense, and lincRNA was determined using featureCounts.

Transcript Coverage

CORALL generates transcriptome-wide smooth, uniform read coverage, comparable to that of competitor kits (Fig. 2).

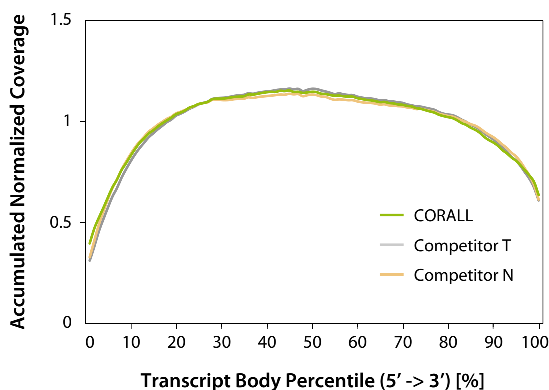

Figure 2 | Accumulated transcript body coverage (whole transcriptome). Coverage across all transcripts was generated using the gbc-tool provided by RSeQC (transcripts length normalized to 100 %).

Gene Detection

CORALL delivers excellent gene discovery rates matching conventional competitor kits (Fig. 3A). 95 % of genes are commonly detected between CORALL and each of the two competitors (at >10 counts per million (CPM), Fig. 3B). This high level of library complexity ensures faithful representation of the transcriptome, enabling sensitive expression profiling.

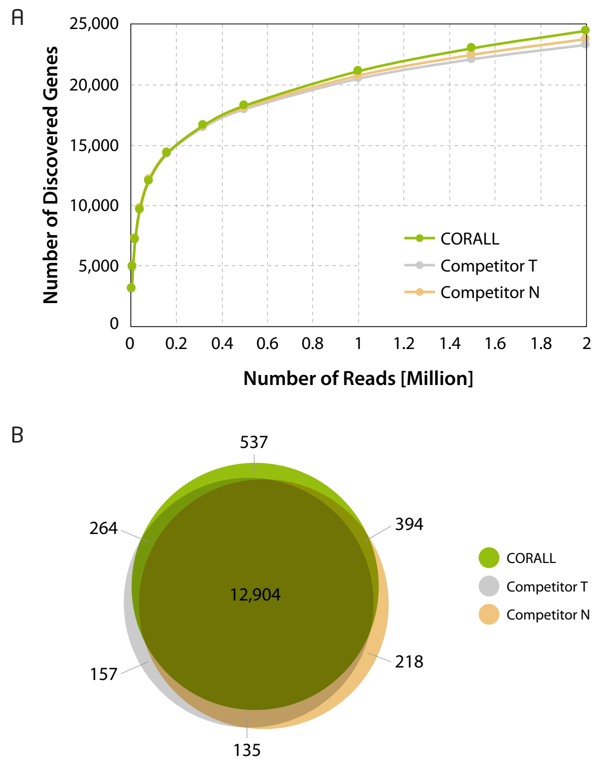

Figure 3 | Gene detection. A) Gene discovery rates. The number of detected genes is plotted against the total number of reads mapping uniquely to exons (calculated with featureCounts). **B) Overlap of detected genes.** The Venn diagram illustrates overlaps between CORALL and competitor kits for genes detected with normalized expression levels >10 CPM (for uniquely mapping reads).

Isoform Detection

CORALL enables accurate isoform detection as shown for Lexogen's Spike-In RNA Variant controls (SIRVs). SIRVs are a set of 69 defined, synthetic RNA molecules that comprehensively reflect transcriptome complexity. CORALL authentically reproduces the expected coverage (Fig. 4, gray line) and is thus ideal for isoform discovery and alternative splicing applications.

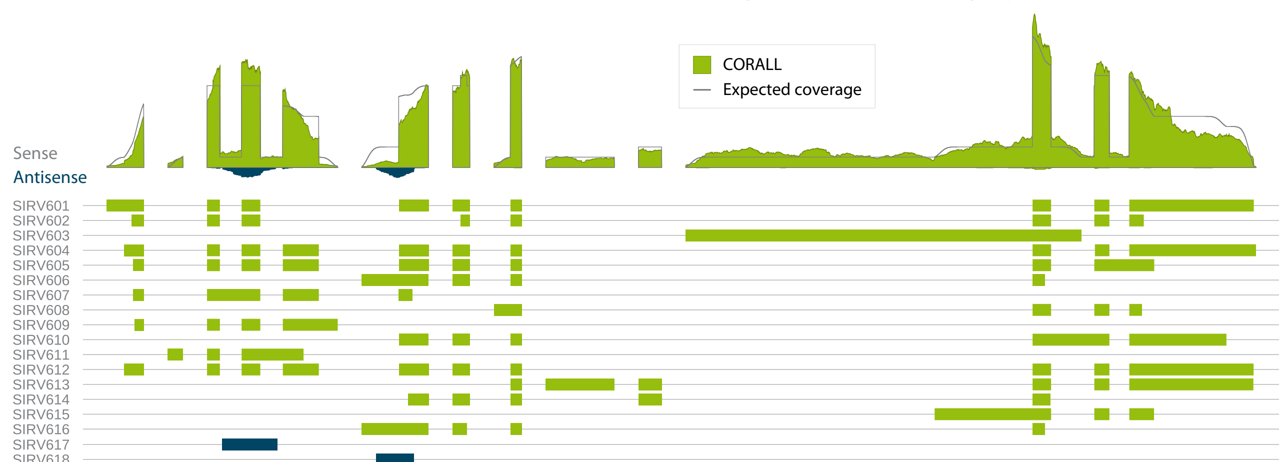

Figure 4 | Coverage of SIRV6. Lexogen's SIRV-Set 3 was spiked into the input RNA Sample before library preparation (as described in Tab. 1). Reads were mapped to the SIRV reference genome [1, 3] and visualized on gene-level (upper panel). The exon-intron structure of all 18 transcripts of the SIRV6 gene is shown in the lower panel (with antisense transcripts in blue).

Superior End-to-End Coverage

CORALL's comprehensive coverage delivers improved transcript start and end site representation. Read coverage was analyzed using the ERCC spike-in controls, which feature precise, known transcription start and end sites (TSS and TES, respectively). CORALL reads map more accurately to the exact ERCC TSS (Fig. 5A) than competitor libraries, which fail to cover the true start sites. Additionally CORALL provides elevated coverage at TES (Fig. 5B).

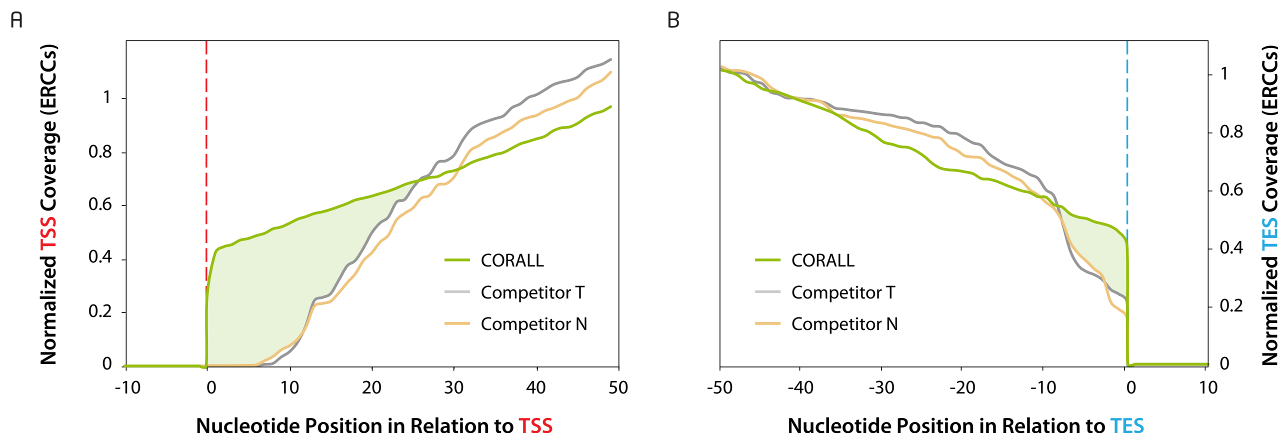

Figure 5 | Normalized ERCC coverage of A) TSS and **B) TES.** Normalized coverage of accumulated mapped reads for all detected ERCCs. The absolute nucleotide positions relative to the TSS (red dotted line, A) and TES (blue dotted line, B) are shown.

Enhanced Endogenous Transcript Coverage

Coverage profiles show consistent uniformity across the transcript body for CORALL and competitor methods alike (Fig. 6A). As for the ERCCs (Fig. 5) endogenous TSS and TES are accurately represented in CORALL libraries, with enhanced coverage at 5' and 3' ends. Coverage of the cellular GAPDH transcript exceeds that of competitors at the TSS (Fig. 6B) and is elevated at the exact TES (Fig. 6C).

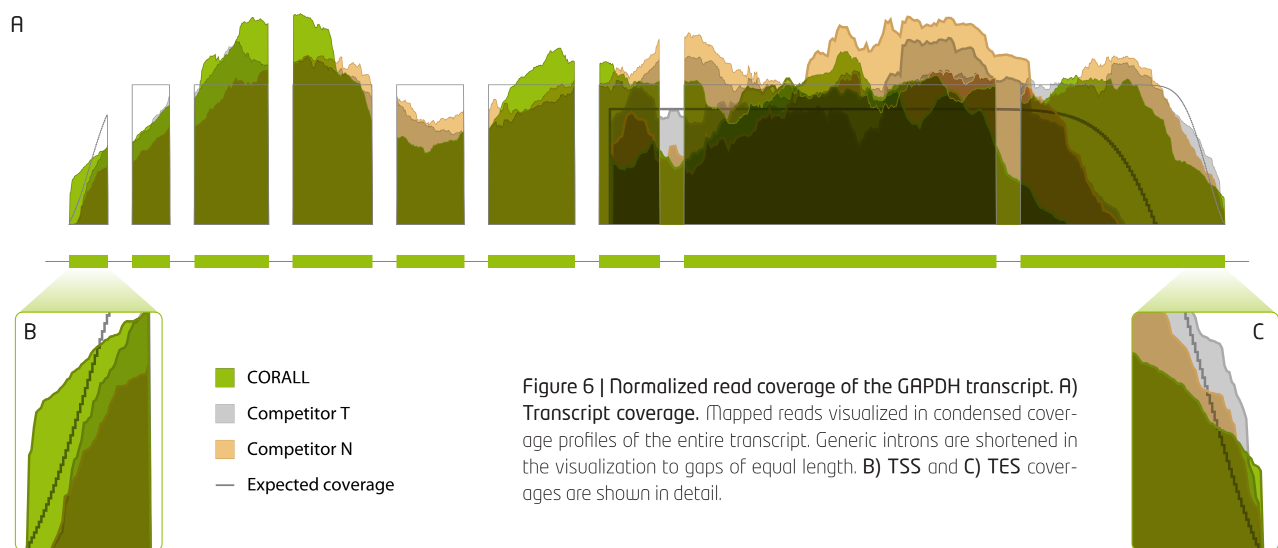

Figure 6 | Normalized read coverage of the GAPDH transcript. A) Transcript coverage. Mapped reads visualized in condensed coverage profiles of the entire transcript. Generic introns are shortened in the visualization to gaps of equal length. **B) TSS** and **C) TES** coverages are shown in detail.

RNA Sequencing

CORALL libraries can be multiplexed using up to 96 i7 indices (included in the kit). Additional i5 indices can also be introduced using the Lexogen i5 6 nt Dual Indexing Add-on Kits (Cat. No. 047). Used together, Lexogen's 96 i7 and 96 i5 indices enable up to 9,216 different index combinations for sequencing.

CORALL Total RNA-Seq libraries are compatible with single-read and paired-end sequencing on Illumina instruments. Read 1 contains the UMI and directly represents the transcript sequence. Hence the UMI information is directly accessible also in the cost-efficient single-read mode. Data can be analyzed with a number of standard bioinformatic pipelines.

"We used the new CORALL kit for performing transcriptome-analysis of CRISPR-modified cells in order to understand the consequences of deregulated epigenetic modifiers. In our hands the kit performance was highly satisfying in terms of data-quality and reproducibility across biological replicates. It furthermore convinced us with the ease of use, clarity of instructions, details in the manual and handling of reagents."

Dr. Max Koeppel, Head of the Functional Tumor Genomics Group at Leibniz-Institute DSMZ, Germany

Features, Benefits and Applications

CORALL is a universal solution for whole transcriptome RNA-Seq applications, delivering excellent performance, with unrivalled coverage of transcript start and end sites.

Features and benefits include:

- A fast and easy workflow for flexible RNA input types, with excellent protocol-inherent strandedness.
- Matched performance with conventional competitor kits.
- Integrated UMIs tag individual transcripts to eliminate amplification bias.
- Dual indexing options to minimize index hopping.
- Full compatibility with Illumina sequencing instruments for single-read or paired-end sequencing.
- Suitable for high-throughput applications and large sample numbers.
- Highly competitive pricing.

CORALL is suitable for all whole transcriptome RNA-seq applications, including:

- Gene expression profiling.
- Isoform discovery and quantification.
- Alternative splicing studies.
- Transcript (re)annotation.
- *De novo* assembly.
- SLAMseq metabolic RNA sequencing.

References

[1] Dobin A., et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, DOI: 10.1093/bioinformatics/bts635

[2] Tuerk A., et al. (2017) Mixture models reveal multiple positional bias types in RNA-Seq data and lead to accurate transcript concentration estimates. *PLoS Computational Biology*, DOI: 10.1371/journal.pcbi.1005515

[3] Paul, L. et al. (2016) SIRVs: Spike-In RNA Variants as External Isoform Controls in RNA-Sequencing. *bioRxiv*. DOI: 10.1101/080747.

Ordering Information

Catalog Numbers:
095 (CORALL Total RNA-Seq Library Prep Kit)
096 (CORALL Total RNA-Seq Library Prep Kit with RiboCop)
025, 050, 051 (SIRVs Spike-in RNA Variant Control Mixes)
037 (RiboCop rRNA Depletion Kit)
039 (Poly(A) RNA Selection Kit)
047 (Lexogen i5 6 nt Dual Indexing Add-on Kits (5001-5096))

Lexogen GmbH · Campus Vienna Biocenter 5 · 1030 Vienna · Austria

Find more about CORALL at www.lexogen.com.
Contact us at info@lexogen.com or +43 1 345 1212-41.

CORALL™
Cover all bases