LEXOGEN
The RNA Experts

143P00014V0101

LUTHOR – unparalleled sensitivity in single cell sequencing

Mantas Survila, Pamela Moll, Cynthia Wollner, Musashi Tsujita, Kathrina Radl, Mira Schauper, Michael Moldaschl and Torsten Reda

Lexogen GmbH, Campus Vienna Biocenter 5, 1030 Wien, Austria

ABSTRACT

A key quality metric in single-cell RNA-seq is the ability to capture ideally all RNA molecules present in each cell. Here we introduce the novel LUTHOR 3' mRNA-Seq protocol based on THOR amplification that holds the power to capture mRNA transcripts close to the estimated maxima.

THOR technology can be used to generate **individual and pooled 3' mRNA-seq, or full-length individual libraries** from ultra-low input RNA or single cells. The technology is template-switch-, and ligation-free, and employs a unique THOR reaction. The THOR reaction is initiated at oligo(dT) primed poly(A) tails introducing a T7 promoter sequence to all 3' ends of transcripts. The resulting structure allows swift **amplification of antisense RNA directly from mRNA templates.**

To demonstrate the power of LUTHOR, the quantitative performance was also compared in a high-throughput compatible setup. LUTHOR libraries were generated from HEK293T cells captured in 1 nL size droplets using Onyx instrument (DropletGenomics) and compared to 10X Single Cell v3.1 data with a limited number of reads. The gene and UMI detection rates were minimum 2-fold higher when compared to 10X at all initial read depths.

LUTHOR Innovative Workflow

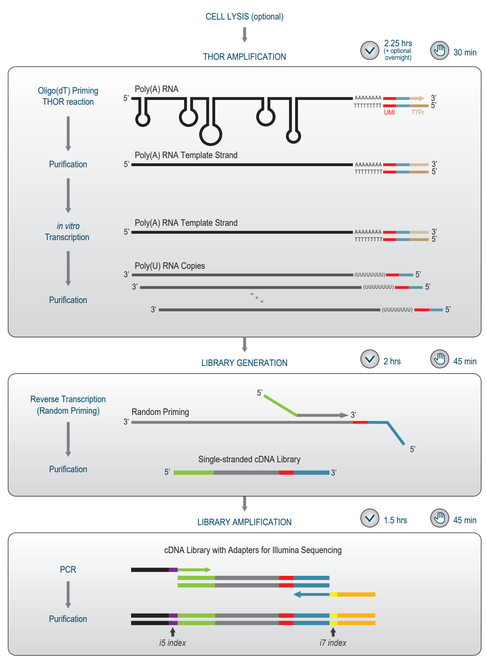


Figure 1 | LUTHOR Single Cell/Low Input RNA Kit technology principle and workflow.

Procedure

THOR amplification is initiated by hybridizing an oligo(dT) primer containing a 12 nt long Unique Molecular Index (UMI), a partial Illumina-compatible P7 linker and a T7 promoter sequence. The proprietary THOR reaction removes the single-stranded 3' poly(A) overhang and generates a double stranded T7 promoter sequence for RNA amplification. Antisense-RNA copies are generated from original mRNA templates. The Illumina P5 adapter is added by random priming. The resulting cDNA contains both partial adapters and UMIs. During library amplification, the second strand is generated, adapter sequences for cluster generation and indices are introduced. The amplified cDNA is purified and used for sequencing.

Performance

Sequencing depth detection limit

The combination of amplification and subsequent random sampling during sequencing determines the detection limits as function of sequencing depth. Highly efficient library preps require compatible high sequencing depths to resolve complete quantitative single cell gene expression signatures.

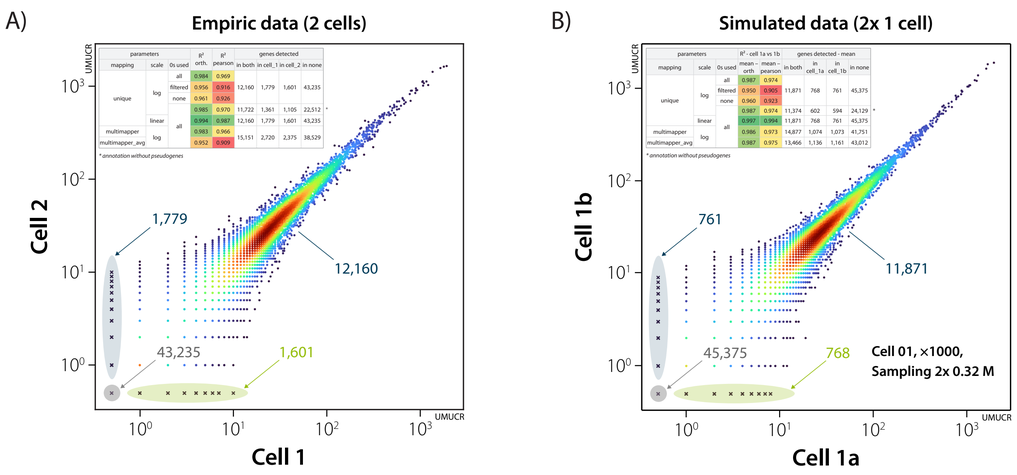
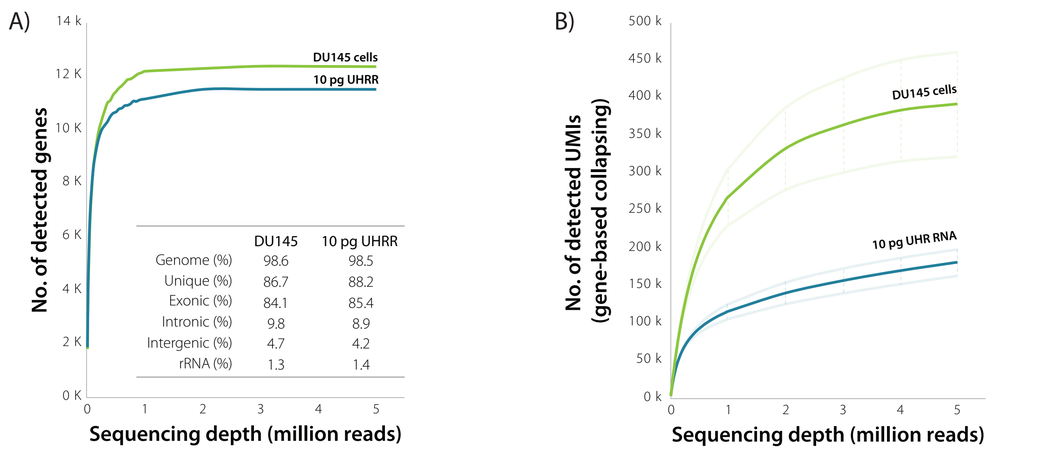


Figure 2 | Correlation of scLUTHOR 3'seq gene expression and simulation of unbiased amplification and sequencing. **A)** Two different HEK293T cells with an average of 14.25 pg total RNA were sequenced with 1 M raw reads which resulted in ca. 0.32 M uniquely mapping UMI collapsed reads (UMUCR). **Table Inserts**, R2 values were calculated based on orthogonal or Pearson correlations including all 0 values (all), filtering out just 0 values from both cells (filtered), or omitting all 0 values (none) using linear or logarithmic scaling. Mapping was done against hsa_GRCh38.94 or a *) curated annotation without pseudogenes used standardly in single cell sequencing experiments ("evidence-based annotation of the human genome (GRCh38), version 32 (Ensembl 98)"). **B)** the 0.32 M reads (UMUCR) of cell 1 were bias-free amplified 1,000 times in silico and random sampled twice to a depth of 0.32 M reads which results in a close to identical correlation compared to A) illustrating the noise caused by limited read depths. As for the comparison of two different cells, different R2 values are highly similar.

Based on the complexity of mammalian cells one can estimate to require ca. 1 M raw reads to detect about 95 % of the expressed **genes**, ca. 2 M raw reads for 95 % of the **transcripts** and up to 10 M raw reads to detect all transcripts simply due to the random sampling of otherwise perfectly captured and amplified transcriptome libraries.

LUTHOR captures theoretical max number of unique mRNA molecules



	DU145	10 pg UHRR
Genome (%)	98.6	98.5
Unique (%)	86.7	88.2
Exonic (%)	84.1	85.4
Intronic (%)	9.8	8.9
Intergenic (%)	4.7	4.2
rRNA (%)	1.3	1.4

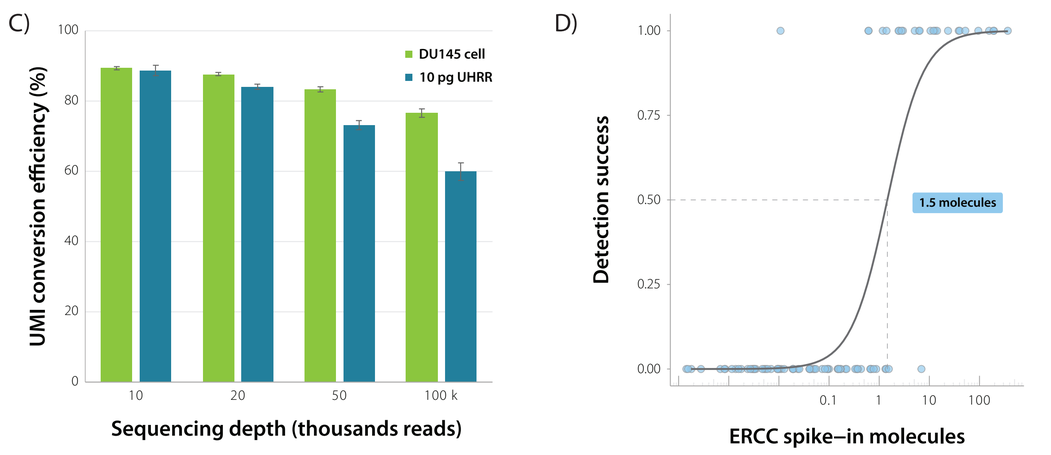


Figure 3 | Sensitivity of LUTHOR. **A)** Scatter plots of the average number of genes detected per DU145 human cell (contains 18.3 ± 1.5 pg of total RNA) and 10 pg Universal Human Reference RNA (UHRR) inferred across four replicates at stepwise-reduced read fractions (CPM > 1), plateaus above 1 million reads. Table shows sequencing alignment metrics across four DU145 and 10 pg UHRR replicates at 1 million read depth. **B)** Scatter plots of the average number of UMIs (UMI = 12 bp) detected per DU145 human cell and 10 pg UHRR inferred across four replicates at stepwise-reduced read fractions. UMI counts represent all unique alignments after gene-based collapsing. **C)** The rate of conversion from shallow initial sequence reads to UMI counts (UMI_{counts}/$fastq_{raw}$). **D)** Logistic regression-based detection of 92 ERCC transcripts at 1 million read depth for 10 pg UHRR samples as measured in Svensson[2]. Estimation is based on relative ERCC capture rate of 40 % compared to the endogenous mRNA.

Comparison of LUTHOR and Chromium™ Next GEM Single Cell 3' kit v3.

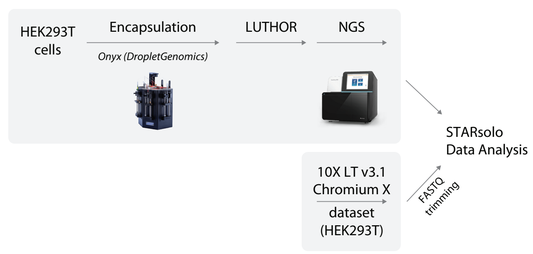


Figure 4 | Experimental design. Fresh human HEK293T cells were captured in droplets with Onyx microfluidic system. In sequential steps, cells were lysed inside droplets and mRNA processed to sequencing-ready LUTHOR library generated from approx. 1500 cells. The pooled library was sequenced on Illumina NextSeq 500 with 51 bp for Read 1 (UMI = 6 bp) and 32 bp for Read 2 (cDNA). 10X 3' LT v3.1(Chromium X) dataset[1] generated from approx. 1630 cells was used in a direct comparison. UMI 12 bp sequence for Read1 was trimmed down to 6 bp, and 90 bp cDNA sequence on Read2 was trimmed down to 32 bp as for LUTHOR. Both datasets were then analyzed using STARsolo 2.7.9a with intron mode enabled.

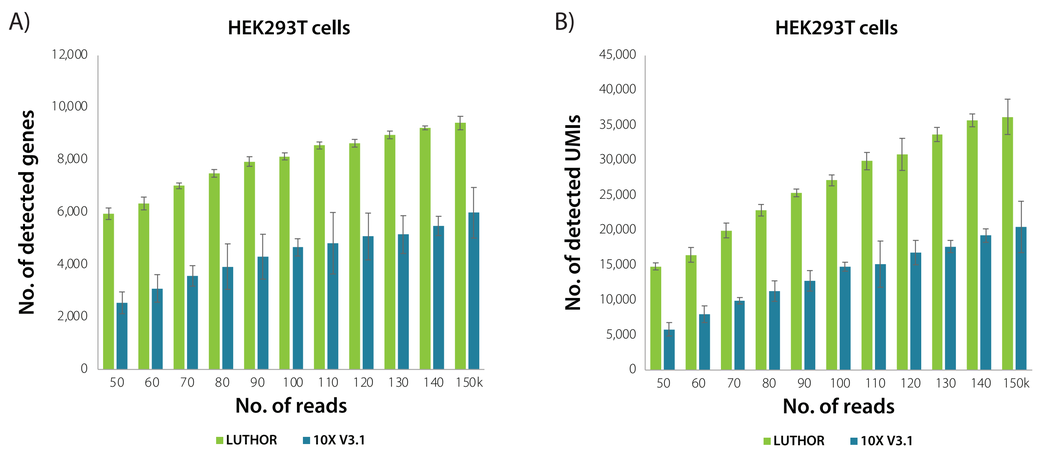


Figure 5 | Gene and UMI detection comparison between LUTHOR and 10X v3.1 **A)** Average number of detected genes per cell in relation to indicated sequencing depths inferred across min eight cells. **B)** Average number of detected UMIs per cell in relation to indicated sequencing depths inferred across min eight cells.

Conclusions

- The proprietary THOR reaction generates double-stranded T7 RNA polymerase promoter onto single-stranded poly(A) primed mRNA templates. This allows in vitro RNA-transcription-mediated amplification directly from mRNAs in the beginning of the protocol.
- Early mRNA pre-amplification allows to reach nearly 100 % mRNA molecule capture efficiency.
- Due to high sensitivity and UMI conversion efficiency, LUTHOR requires three-fold less reads to generate the same UMI counts for high-throughput droplet-based scRNAseq methods. This allows to analyse three-fold more single cells with limited sequencing.
- The majority of LUTHOR reads generated from single cells or UHRR map to exonic sequences.
- High definition scRNAseq libraries should be sequenced at 1 million reads per sample for saturated gene detection and minimum 5 million reads for saturated transcript detection.

References:

[1] Reference path: /mnt/scratch2/cellranger-6.1.0/references/refdata-gex-GRCh38-and-mm10-2020-A
[2] Svensson et al. 2017: doi:10.1038/nmeth.4220

For more information please visit our website:

LUTHOR product page:
www.lexogen.com/luthor-single-cell-low-input-3mrna-sequencing