LEXOGEN
The RNA Experts

143P00014V0200

LUTHOR HD – unparalleled sensitivity in subcellular sequencing

Mantas Survila, Pamela Moll, Cynthia Wollner, Musashi Tsujita, Kathrina Radl, Mira Schauper, Michael Moldaschl, Filippo Passardi 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 all RNA molecules present in each sample fraction. Here we introduce the novel LUTHOR HD 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, cytosolic cell fractions or lysed 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.** With the help of LUTHOR HD we accessed the sequencing depth detection limits for single cell experiments, and input amounts as low as 1 pg total RNA equivalents. Further, we investigated R^2 calculations which define correlations of detected but also non-detected genes in replicates taking the statistical constraints through read depth limitations into consideration. Such metrices are needed to distinguish experimental noise from expression data in cytosol sampling.

LUTHOR Innovative Workflow

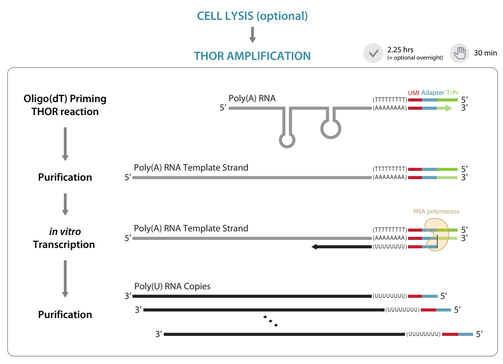


Figure 1 | LUTHOR Single Cell/Low Input RNA Kit technology principle and workflow.

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 unique dual indices are introduced. The amplified library is purified and used for sequencing.

LUTHOR HD overcomes sensitivity limitations by employing in vitro-mediated RNA amplification of poly(T)-tagged mRNA templates at the very beginning of the protocol prior to conversion into NGS libraries.

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.

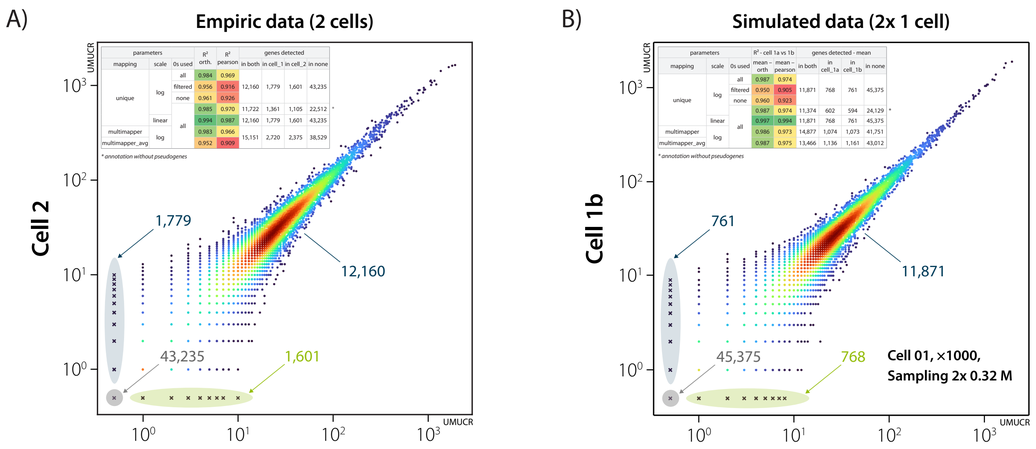
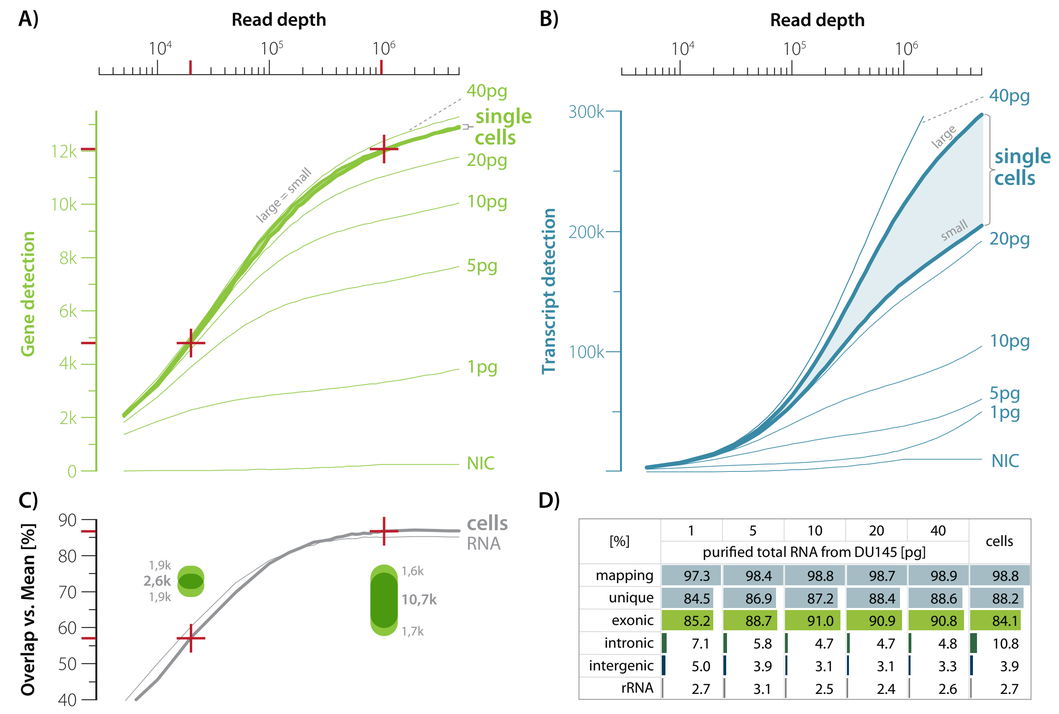


Figure 2 | Correlation of scLUTHOR 3'seq gene expression and simulation of unbiased amplification and sequencing. A) Two different HEK293T cells with an average of 14.25 pg total RNA were sequenced with 1 M raw reads which resulted in ca. 0.32 M uniquely mapping UMI collapsed reads (UMUCR). **Table Inserts**, 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.

Detection rates

The performance of LUTHOR was evaluated using 1 - 40 pg of total RNA input purified from DU-145 human prostate cancer cells and single DU-145 cells isolated by FACS. All samples were sequenced at a depth of 5 M raw reads. Gene and transcript detection rates including Venn-style diagrams for cell pairs at representative read depth of 20k and 1M raw reads, and the degree of gene detection overlap are shown in Fig. 3A-C.



[%]	purified total RNA from DU145 [pg] 1	5	10	20	40	cells
mapping	97.3	98.4	98.8	98.7	98.9	98.8
unique	84.5	86.9	87.2	88.4	88.6	88.2
exonic	85.2	88.7	91.0	90.9	90.8	84.1
intronic	7.1	5.8	4.7	4.7	4.8	10.8
intergenic	5.0	3.9	3.1	3.1	3.3	3.9
rRNA	2.7	3.1	2.5	2.4	2.6	2.7

Figure 3 | Gene and transcript detection sensitivity, and gene detection overlaps at different sequencing depths. (A) Number of detected genes and **(B)** gene based collapsed UMIs (transcripts) at different read depths are shown for single DU-145 cells (region covered by 8 cells), together with means of no input control (NIC), and of 1 - 40 pg total RNA purified from DU-145. **(C)** The gene detection overlaps between two cells, and two 20 pg RNA samples are shown together with two Venn diagrams at typical read depths of 20k reads per cell in high cell number scRNA-seq experiments, and 1M reads per cell recommended for LUTHOR HD (highlighted by the red crosses). **(D)** Table shows sequencing alignment metrics across four DU-145 cells and 1 - 40 pg of total RNA purified from DU-145 cells at one million read depth.

Typical mammalian cells contain 200k – 500k mRNA molecules[1]. The in-depth measurement of transcripts and their copy numbers requires high conversion efficiencies during library preparation and high-depth sequencing due to the random sampling of sequencing reads, even after unbiased library amplification methods[2,3]. So far, low-abundance transcripts are usually sparsely represented due to i) limited sensitivity of scRNA-seq protocols, and ii) stochastic dominance of high abundance transcripts at limited sequencing depth.

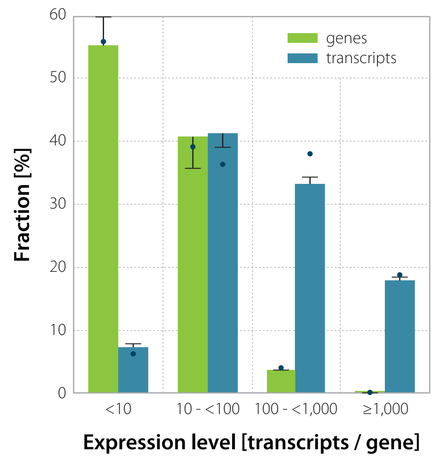


Figure 4 | Distribution of gene expression levels in single cells. The relative number of genes and their corresponding transcripts are grouped by the magnitude of the mean expression level per cell. The bars show the calculated fraction using uniquely mapping reads only, while the dashes mark the shares counting also multimapping reads averaged by the number of mapping positions. Dots depict uniquely mapping and multimapping reads summed up over all mapping positions respectively.

Due to the stochastic distribution of reads, the increase of transcript detection starts linearly (increasing slope in the logarithmic plot in Fig. 3B). UMI conversion efficiency is high and reaches values of 75 - 85 % even at 100k reads (data not shown). The transcript detection asymptotically approaches total numbers of 200k and 300k unique transcripts, which represent smaller and larger FACS sorted, non-synchronized DU-145 cells, respectively. However, the same cell type has highly similar gene detection rates independent of the cell size. Both overlaps of gene expression between cell pairs as well as ultra-low input samples of similar total RNA content of 20 pg are in the same range reaching approximately 85 % at read depths of more than 1M reads per sample. This is a result of the low read depth (random sampling) and the fact that 50 % of genes contain less than ten transcript copies (Fig. 4).

LUTHOR HD demonstrates the information depth of high-definition gene expression profiling. In contrast, less efficient methods paired with corresponding lower read depths capture only higher abundance gene expression with increased scatter at medium expression levels. At low read depths the overlap drops significantly.

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 which greatly enhances the resolution of gene expression measurements.
- High-definition RNA-seq enables analysis of complete gene expression signatures to study true heterogeneity in **cellular and subcellular systems**.
- High definition RNA-seq libraries should be sequenced at **compatible read depths**, i.e., 1 M reads per sample **for saturated gene detection** and minimum 5 M reads **for saturated transcript detection**.

References:

[1] Shapiro, E. et al. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat. Rev. Genet. 14, 618-30 (2013).
[2] Svensson, V. et al. Power analysis of single cell RNA-sequencing experiments. Nat. Methods 14, 381–7 (2017).
[3] Ziegenhein, C. et al. Comparative analysis of single-cell RNA sequencing methods. Molecular Cell 65, 631-643 (2017).

For more information please visit our website:

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