LEXOGEN
The RNA Experts

LUTHOR HD: High-definition single-cell gene expression analyses

LUTHOR HD 3′mRNA-seq libraries display comprehensive transcription profiles, enabling high-definition gene expression characterization from single eukaryotic cells, cell biopsies, or purified total RNA down to 1 pg. The method uses the proprietary THOR reaction to initiate direct mRNA amplification. The data complexity generated by LUTHOR HD covers the entire dynamic range of gene expression, identifying true, detailed transcriptional signatures.

Introduction

The sensitivity of single-cell RNA-seq methods depends on multiple library preparation steps that may contribute to dropouts during reverse transcription, template-switching, tailing, or ligation. Any losses during these steps will result in reproducible data for highly expressed genes only. To compensate for low mRNA conversion rates, current high-throughput methods rely on sequencing large numbers of cells at shallow read depth, which provides an incomplete picture of gene expression. Although shallow sequencing is sufficient for cell type identification based on few highly expressed marker genes, it lacks the transcriptional complexity to define cell-specific heterogeneity[1-3].

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.

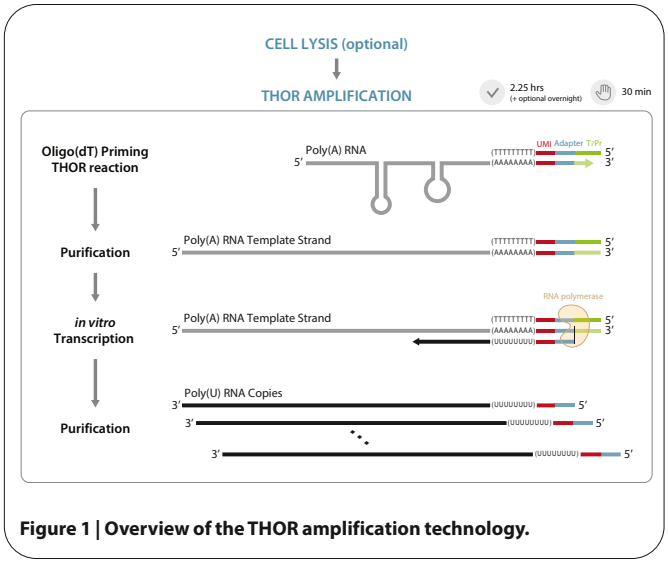


Figure 1 | Overview of the THOR amplification technology.

The THOR technology

THOR linearly amplifies original mRNA molecules fused to library-specific tags and a T7 promoter required for RNA-templated *in vitro* transcription. Generation of a double-stranded T7 promoter region on the 3′ end of single-stranded mRNAs enables synthesis of antisense RNA copies directly from mRNA templates. Because only the endogenous poly(A)-tailed mRNA molecules are primed, no RNA extraction and no prior poly(A) enrichment or ribosomal RNA depletion are required (Fig. 1). THOR is inherently inert to genomic DNA background.

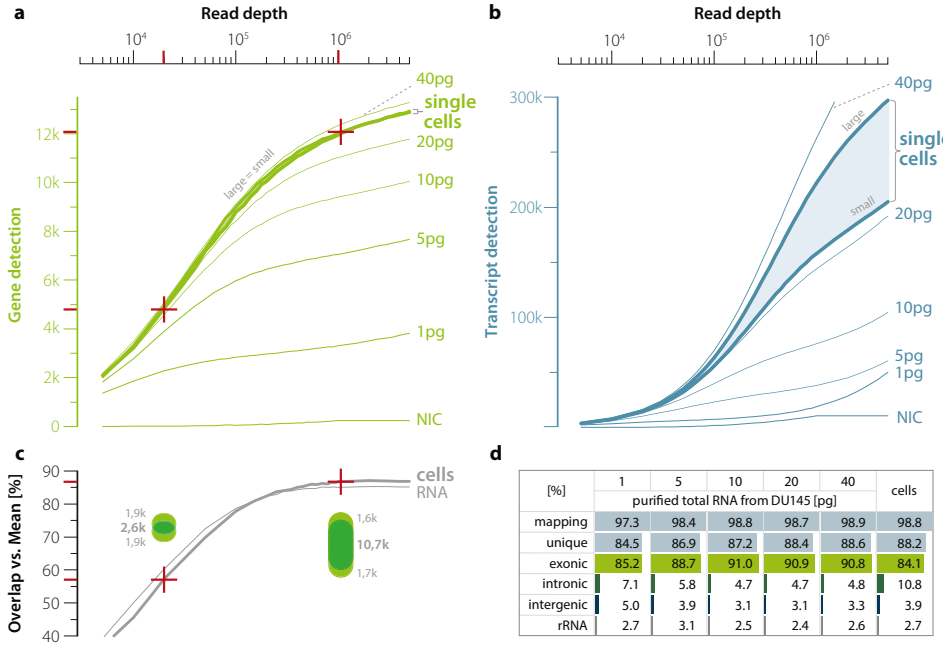
Detection rates

Typical mammalian cells contain 200k – 500k mRNA molecules[4]. 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[5,6]. 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.

In this context, the performance of LUTHOR was evaluated using 1 to 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 5M 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. 2a-c.

Mantas Survila, Pamela Moll, Michael Moldaschl, and Torsten Reda

Lexogen GmbH, Campus Vienna Biocenter 5, 1030 Wien, Austria
Email: mantas.survila@lexogen.com



[%]	1	5	10	20	40	cells
	purified total RNA from DU145 [pg]					
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 2 | 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 to 40 pg total RNA purified from DU-145. **c**, The gene detection overlaps between two cells (bold line and Venn diagrams), and two 20 pg RNA samples (fine line) are shown 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 to 40 pg of total RNA purified from DU-145 cells at one million read depth

Due to the stochastic distribution of reads, the increase of transcript detection starts linearly (increasing slope in the logarithmic plot in Fig. 2b). 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. 3).

Conclusions

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. Not only less genes but also fewer common genes can be used for bioinformatic characterization, e.g., PCA, t-SNE, or UMAP clustering. The direct mRNA amplification in LUTHOR HD greatly enhances the resolution of gene expression measurements. High-definition single cell RNA-seq enables

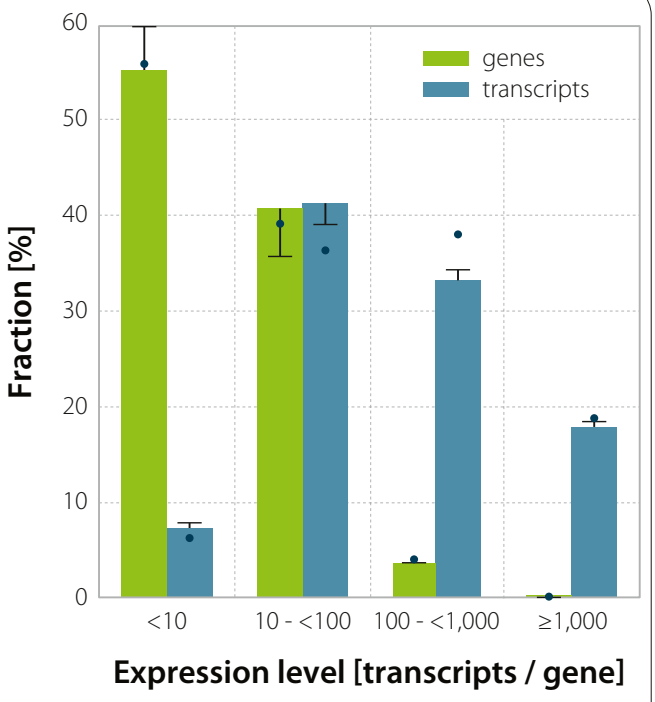


Figure 3 | Distribution of gene expression levels in single cells. The relative number of genes and their corresponding transcripts are grouped by the magnitude of the mean expression level per cell. The bars show the calculated fraction using uniquely mapping reads only, while the dashes mark the shares 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.

analysis of complete gene expression signatures to study true heterogeneity in cellular systems. Applications comprise improved biomarker identification in clinical settings using microbiopsies or singularized circulating tumor cells. LUTHOR HD enables precise characterization of compound-target response in drug discovery and development using precious samples derived from, e.g., differentiated cells, or primary cells from tissue samples. Highest resolutions for gene expression profiling can be achieved in single-cell CRISPR-based screenings.

REFERENCES

1. Kharchenko, P.V. *et al*. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11**, 740-2 (2014).
2. Haque, A. *et al*. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine* **9**, (2017).
3. Zhang, X. *et al*. Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. *Molecular Cell* **73**, 130-142 (2019).
4. Shapiro, E. *et al*. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet*. **14**, 618-30 (2013).
5. Svensson, V. *et al*. Power analysis of single cell RNA-sequencing experiments. *Nat. Methods* **14**, 381–7 (2017).
6. Ziegenhain, C. *et al*. Comparative analysis of single-cell RNA sequencing methods. *Molecular Cell* **65**, 631-643 (2017).