Research &
Pharma Solutions

Guideline for Transcriptome Sequencing

Tech Note

Guideline for Transcriptome Sequencing

Next-generation sequencing allows the detection and relative quantification of RNA molecules. Transcriptome sequencing is a powerful tool to analyze RNAs in a specific tissue and at a specific time point. It can be used for various purposes.

Transcriptome Workflow

Our transcriptome workflow enables reliable and accurate insights into each step of our process, from RNA extraction to data analysis. Due to permanent and strict quality control as well as validated processing procedures, our clients benefit from high-quality results. Our pipeline is subjected to permanent improvement to offer the best solution possible for all kinds of samples.

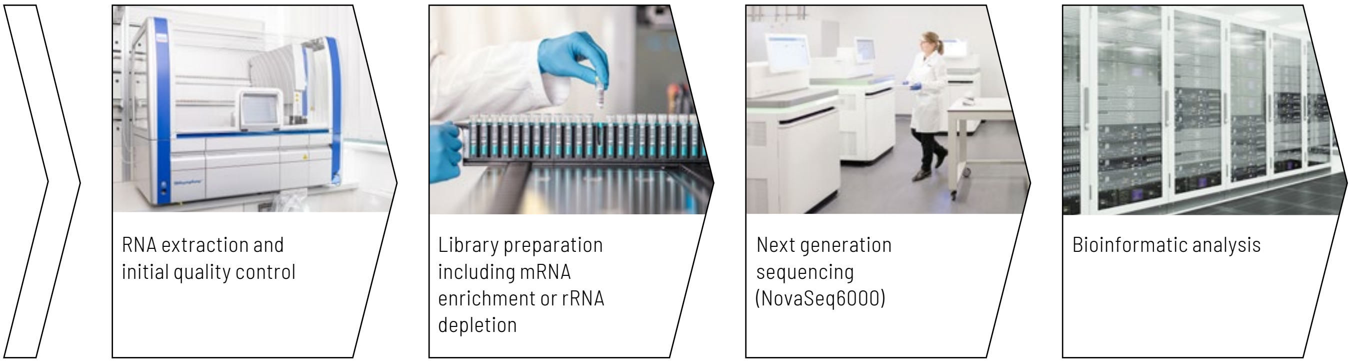

RNA extraction and initial quality control
Library preparation including mRNA enrichment or rRNA depletion
Next generation sequencing (NovaSeq6000)
Bioinformatic analysis

Figure 1: Description of the transcriptome workflow: *Either the RNA is isolated from primary samples, or directly enters our process. After the RNA passed our internal quality control pipeline, the library preparation with one selected kit for mRNA or total RNA can be started. The resulting libraries are sequenced on Illumina's NovaSeq6000 and the raw data are processed.*

Do not hesitate to contact us if you are looking for other analyses.

Please note that in order to obtain a statistically significant result, at least three replicates per experimental group are necessary.
We are looking forward to planning your project with you!

Selection of the Appropriate Library Preparation Kit

Transcriptome sequencing protocols are continuously improved at CeGaT. Our goal is to provide our clients with optimal data quality for all applications. Transcriptome sequencing is highly dependent on the quantity and quality of the RNA, and very sensitive to even minute changes in the experimental setup. Considering these, we choose the most suitable protocol for library preparation to achieve the best results possible. Reproducible, comparable and unbiased data are crucial for the comparison within a client´s project. Therefore, we always use the same library workflow and have automated processes wherever possible to avoid batch effects.

Our library preparation protocols and wet-lab processes are optimized to generate sequencing libraries using a sequencing read length of 100bp in a paired-end mode (PE100). The sequencing depth is determined by the research objective. Considering the guidelines from the ENCODE (ENCyclopedia of DNA Elements) consortium, we recommend a sequencing depth of 30 M (Million) clusters for mRNA sequencing of human samples. However, this recommendation can differ depending on the transcriptome size of the analyzed organism. Bacteria, for example, require much fewer reads than a hexaploidic plant. To discover low copy number transcripts in samples of particularly low quality and low input, we recommend sequencing at least 50 M clusters.

Coding Transcriptome Sequencing versus Whole Transcriptome Sequencing

Most applications for RNA sequencing can be accomplished either by coding transcriptome sequencing (mRNA sequencing) or whole transcriptome (total RNA) sequencing. The starting material for both analyses is single-stranded RNA. We are happy to offer the best method available for the respective research objective.

Coding transcriptome sequencing – mRNA analysis

Sequencing of coding RNA is the most common transcriptome sequencing approach. Its main objective is the quantification of gene expression and differential gene expression analysis, for example the analysis of coding regions (Figure 2).

About 1 - 5% of the transcriptome corresponds to mRNA (messenger RNA). This type of RNA is mostly poly-adenylated. By targeting all poly-adenylated transcripts, most protein-coding RNA (mRNA) can be enriched. Poly-A enrichment increases the mRNA content and decreases other unwanted RNA molecules present in a total RNA sample. Depending on the protocol, this procedure can result in a coverage of over 85% of bases aligning to coding sequences and UTRs (untranslated regions). This results in an increased sequencing depth on mRNA, including low-level expressed mRNA transcripts, and enables researchers to identify rare transcripts.

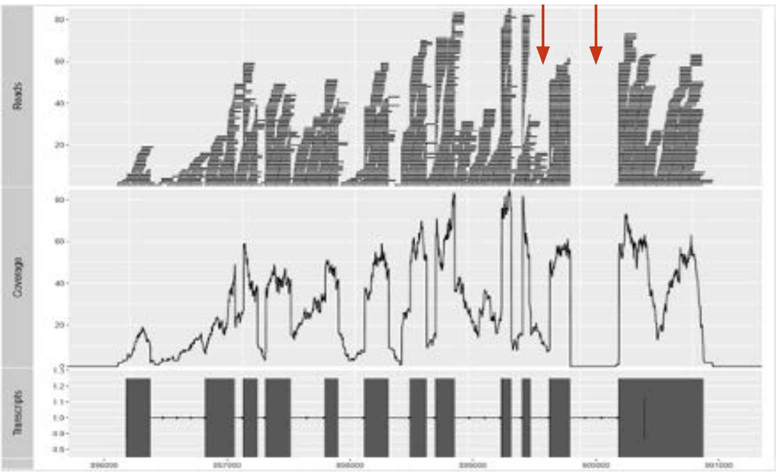


Figure 2: *Representative mapping of a representative gene. The transcripts are shown in the bottom graph and the corresponding coverage is shown in the middle. The read stacks are depicted in the uppermost graph. The coverage is decreased or absent in the intergenic regions (arrows).*

Whole transcriptome sequencing – total RNA analysis

Whole transcriptome sequencing includes the analysis of mRNA, splicing patterns, regulatory regions as well as many forms of non-codingRNAs such as lncRNA and miRNA. It allows a comprehensive view of all transcripts. Ribosomal RNA (rRNA), is usually not the research focus. For better analysis results, rRNA depletion is recommended. Since the efficiency of rRNA depletion has a major impact on the analysis results, CeGaT offers library preparation methods with the most effective rRNA depletion.

Strand specificity

A strand-specific protocol allows unambiguous assignment of reads to genes that overlap but are located on opposite strands (Figure 3). It is impossible to determine the origin of the DNA strand for a certain RNA transcript, if a non-strand-specific protocol has been used.

For coding transcriptome sequencing and whole transcriptome sequencing we offer strand-specific protocols. Identification of the DNA strand, from which the analyzed RNA transcript originated, allows more accurate transcript annotation and the detection of antisense transcript expression.

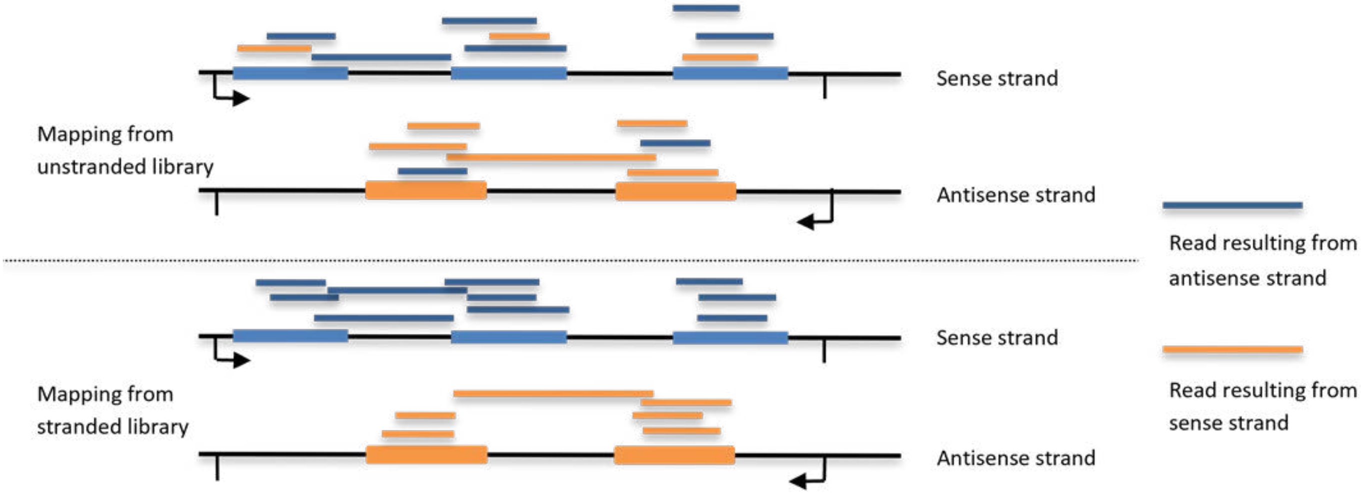


***Figure 3:** Comparison of mapping results of sequences obtained from stranded and non-strand-specific libraries. The stranded library enables the assignment of reads to sense and antisense strand.*

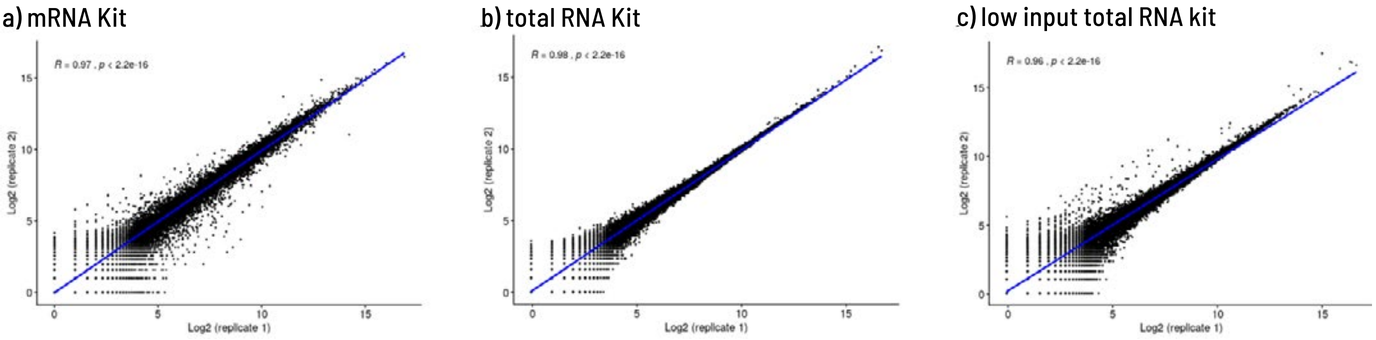
Transcriptome Sequencing Products at CeGaT

Coding transcriptome sequencing (CTS Classic)

CeGaT offers a highly reproducible protocols for a wide range of species for coding transcriptome sequencing. Using our specific mRNA library preparation kit, we offer a consistent workflow with a minimal input of total RNA as starting material (Figure 4a). Since mRNA is targeted via its poly-A tail, it is suggested to use good quality RNA with a RIN (RNA Integrity Number) > 8 to ensure unbiased sequencing data. We offer mRNA sequencing for eukaryotes with a recommended output of 30 M clusters per sample using PE100 sequencing.

Whole transcriptome sequencing (WTS Classic, WTS Deep, TS Flex)

Our total RNA sequencing service provides a reliable and reproducible workflow for analyzing mRNA, as well as many types of non-coding RNA and with the option to deplete ribosomal RNA (Figure 4b). Even challenging samples with, e.g., moderate or slightly fragmented RNA, perform well with our whole transcriptome sequencing protocol. Total RNA of various quality can be used to produce a usable sequencing library. We offer whole transcriptome sequencing for e.g. human, mouse or rat samples. Furthermore, we established workflows to process total RNA samples isolated from human blood, tumors, plants and bacteria. Depending on the sample, RIN values lower than 8 can be accepted. We recommend a sequencing output of 30 M or 50 M clusters, but customization is possible. For more information please get in touch with us.


a) mRNA Kit

b) total RNA Kit

c) low input total RNA kit

***Figure 4:** High correlation of gene expression levels between technical replicates using a) mRNA kit, b) total RNA kit and c) low input total RNA kit. Normalized counts were used to analyze the reproducibility of gene expression data from UHRR replicate samples.*

A considerable part of RNA samples is rather limited and sometimes moderately to highly degraded. We offer adequate solutions of library preparation even for very small amounts of RNA. In case the RNA derived from FFPE tissue or a small number of cells, low input total RNA sequencing is a sensitive solution to accurately detect coding and non-coding RNA transcripts. Our protocol tolerates a DV200 of >30%. Using the low input total RNA kit, CeGaT achieves reproducible results with only 1 ng total RNA of mammalian samples (Figure 4c).

In-House Data Analysis Services

Always included:

- Standard data quality control
- Detailed project report
- Delivery of raw data (untrimmed or adapter trimmed FASTQ files)
- Provision of data via a secure server

Optional:

- Alignment of trimmed sequences for a wide range of eukaryotes, prokaryotes, and viruses (depending on a published reference transcriptome) using STAR (BAM files)
- Determination of raw counts per transcript (TSV files)
- Normalization of raw counts using DESeq2 (TSV files; volcano plot, MA plot, PCA plot)
- Statistical comparison of grouped samples to identify differentially expressed genes between different conditions (at least three samples per group; TSV files, heat map)
- Gene ontology (GO) annotation of differentially expressed genes (TSV files)
- GO term enrichment analysis (TSV files)

By analyzing the normalized read count, differential gene expression analysis can be used to detect quantitative changes in gene expression between experimental groups (Figure 5).

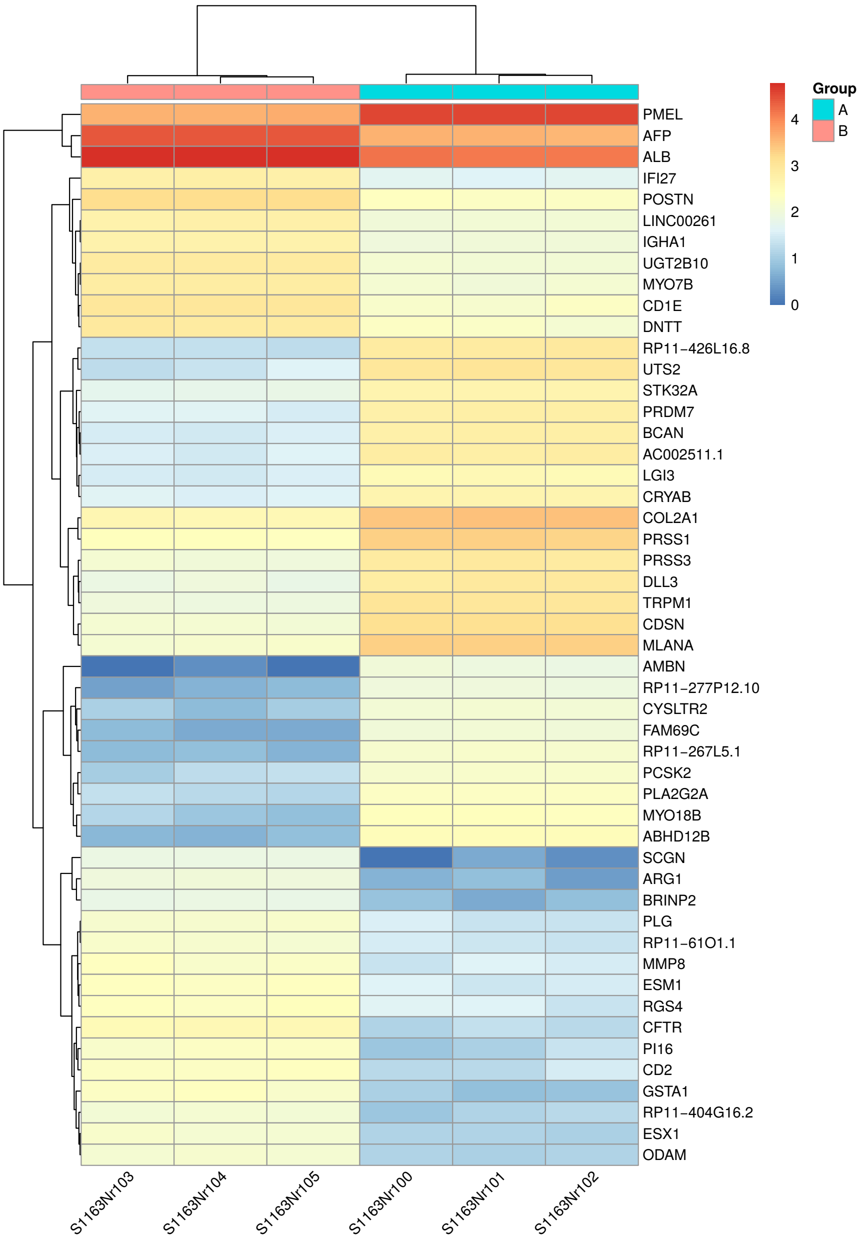


Figure 5: *Heatmap showing the gene expression and group comparison of the top 50 differentially expressed genes. Highly upregulated genes are depicted in red, downregulated genes in blue.*

Data Security

We operate according to the German Genetic Diagnostics Act. All our data are stored on our servers in-house. We offer end-to-end secured data transfer to our customers and can also provide other individual solutions.

CeGaT guarantees that the data generated within a research project remain the exclusive property of the customer. Samples, sequencing data and analysis results are only used for the purposes specified in the project agreement. We carry out all project steps in-house to ensure no third parties have access to any data. Neither data nor sample material will be sold to others.

CeGaT

About Us

CeGaT GmbH is a leading global provider of genetic diagnostics and mutation-related disease analyses. The company combines its next-generation sequencing (NGS) process and analysis pipelines with its medical expertise – dedicated to identifying the genetic cause of disease and supporting patient management.

Genetic mutations can trigger a wide range of diseases, from epilepsy to Parkinson's. Through the use of NGS, it is possible to analyze all genes associated with a disease phenotype simultaneously – both fast and effectively. An interdisciplinary team of scientists and physicians evaluates the data and summarizes the findings in a comprehensive medical report. All services are performed in-house.

CeGaT, founded in 2009 and based in Tübingen, Germany, is accredited according to CAP, CLIA and DIN EN ISO 15189:2014.

CeGaT GmbH
Research & Pharma Solutions
Paul-Ehrlich-Str. 23
72076 Tübingen
Germany

Phone: +49 7071 56544-333
Fax: +49 7071 56544-56
Email: rps@cegat.com
Web: www.cegat.com

DAkkS
Deutsche Akkreditierungsstelle
D-PL-13206-01-00

Accredited by DAkkS according to DIN EN ISO/IEC 17025:2018

CAP ACCREDITED
COLLEGE of AMERICAN PATHOLOGISTS

CLIA CERTIFIED ID: 99D2130225

2022/01