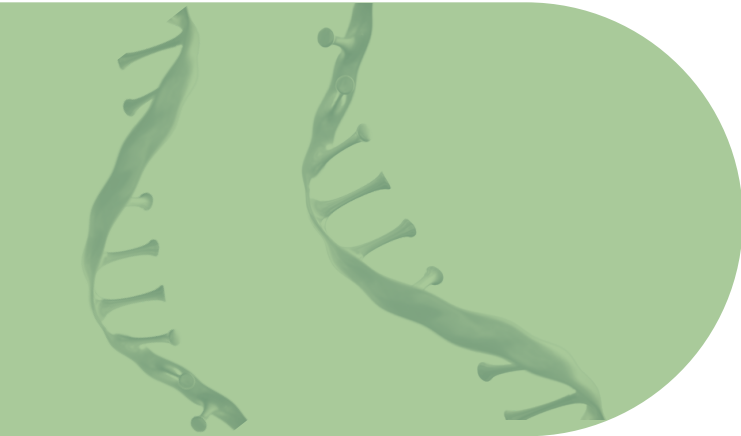




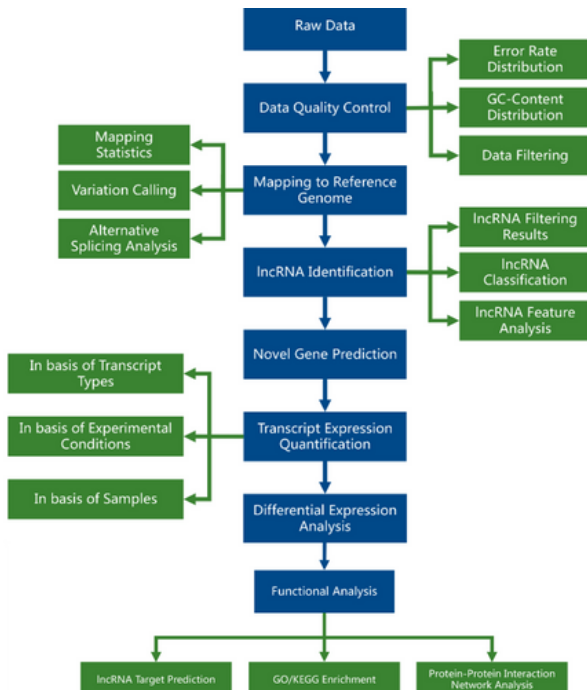
Non-coding RNA-seq Results Explained:
**What You Can Explore With
Non-coding RNA Data**



ADVANCING GENOMICS
IMPROVING LIFE

The bioinformatics analysis of ncRNA sequencing involves the following steps.

- Quality control
- Mapping
- Annotation
- Quantitative analysis (expression)
- Functional analysis
- Qualitative analysis (characterization)



- **Quality control**

Here, first, Raw data is generated and stored in FASTQ format (a text-based format for storing a nucleotide sequence). The FASTQ format has four different rows for “Sequence ID”, Read bases, separator, and quality score provider of the FASTQ format. Moving on to the data filtering step, which uses the fast program to introduce raw reads and produce clean reads. In the end, three things; error rate, base content, and the portion of raw reads transformed into clean reads, are attained.

- **Mapping**

On the clean reads, mapping is performed using a program called Hierarchical Indexing for Spliced Alignment of Transcripts (HISAT2). The reference genome is indexed using a graph-based method, and the Bowtie2 algorithm is used for alignment. This method yields more accurate results with quick and sensitive alignment. Here, the output is a binary form of a SAM (Sequence Alignment Map) file called a BAM file. These BAM files may now be seen in the Integrative Genomics Viewer and compared to the reference genome to determine their differences.

- **Annotation**

After getting BAM files, annotation is done through a software named StringTie. Annotation means identifying functional elements along the sequence of a genome. It uses a network for algorithm as well as an optional de novo assembly step to assemble steps into known or novel gene models based on known gene annotations. In this case, BAM files plus reference annotation files are introduced (input) and a GTF is obtained (output) through transcript annotation of the assembled and aligned reads. Now the assembled transcripts are merged to remove duplicate or redundant transcripts. After this, different filters (Exon number filter, transcript length filter, coding potential filter, etc.) are used to identify and predict the ncRNA types.

- **Quantitative analysis**

The simplest approach to quantify the ncRNA and coding gene is to count the number of reads that map to each transcript. However, two factors need to be taken into consideration. First, the estimated expression level depends on the read counts and total reads sequenced for each sample and the second is that read counts also depend on total gene/transcript length.



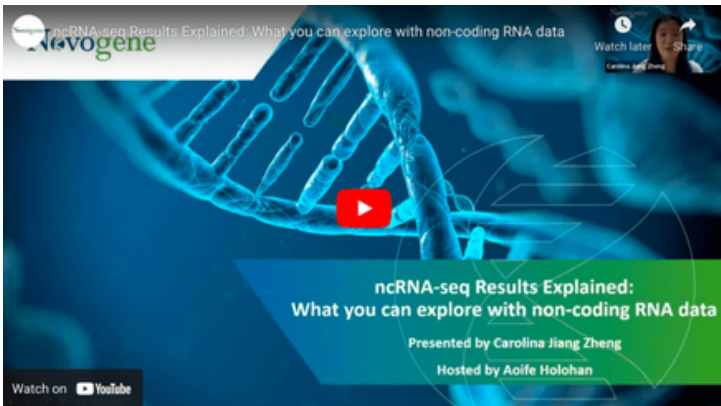
This means it is essential to perform a normalization step to make the data comparable between and within samples.

- **Functional analysis**

In functional analysis, biological reference is assigned to a set of genes. It is determined whether there is the enrichment of any known biological functions, interactions, or pathways. So, a software called ClusterProfiler is used which implements methods to analyze and visualize functional profiles of genomic coordinates, gene and gene clusters and enrich the data. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) are the most frequently used databases for functional analysis. Aside from data enrichment, ncRNA target prediction can also be done.

- **Qualitative analysis**

In this step, variant discovery, and alternative splicing (AS) is done through software; GATK and rMATS respectively. GATK is used to identify where the aligned reads differ from the reference genome and write to a variant call format (VCF) file. BAM files are introduced and VCF files are obtained. On the other hand, rMATS is designed for detecting differential AS in replicated RNA-seq data.



[Click here to watch our related webinar](#)

ncRNA Interaction and Regulatory Networks

Competing Endogenous RNAs theory says that lncRNAs, mRNAs and circRNAs have a competition relationship, trying to interact with the miRNAs and depending on who interacts with which type of microRNA is going to change the gene expression and the transcription regulation. lncRNA and mRNA co-expression analysis can be done via Pearson analysis. lncRNA-mRNA interaction can be visualized through a network software "Cytoscape".

How Novogene Can Help

The company's mission is to advance genomics and improve life. The services offered by Novogene are whole genome and transcriptome analysis, epigenome, human, microbial, plant, and animal genome sequencing, etc. The services cover a huge range of real-life applications from complex disease research, oncology, forensics, food, and water safety to biopharma and industrial diagnosis.

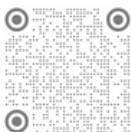


[Click to learn more services Novogene provided](#)

	Library preparation	Strand specific/ directional	Sequencing strategy	Recommended data output	Optional features
lncRNA-seq (mRNA+lncRNA)	rRNA removal	Yes	PE150	12 – 15 Gb	<ul style="list-style-type: none"> • Globin mRNA removal • Exosome RNA • Dual rRNA depletion
circRNA-seq	rRNA and linear RNA removal	Yes	PE150	12 – 15 Gb	
sRNA-seq	Size selection	NA	SE50	10-20M reads (0.5-1Gb)	<ul style="list-style-type: none"> • Exosome RNA

The state-of-the-art lab, latest Novaseq 600 sequencing machines, Falcon II sequencing delivery platform and supercomputing capacity of Novogene provide faster results with reduced human error and time requirements. In the case of ncRNA sequencing, the Novogene can prepare rRNA removal libraries, with a sequencing strategy of PE150, keeping strand-specific directional library by default for lncRNA-seq and circRNA-seq. Optional features of Globin mRNA removal and exosome RNA are also provided.

To avoid bacterial contamination dual rRNA depletion strategy is adopted. For sRNAs removal and directional libraries are not needed due to their small sizes and the SE50 strategy is adopted. Novogene also provides only sequencing services for pre-made libraries as well. Novogene will always provide more efficient & higher quality services in order to be your trusted partner in genomics.



Novogene Co., Ltd

 en.novogene.com

 [Novogene Global](#)  [Novogene Global](#)

Follow us on LinkedIn

 inquiry_us@novogene.com | marketing@novogene-europe.com | marketing_amea@novogeneait.sg