

Sequencing at a New Layer of Complexity

Lumi

1 The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing

The first article is an research article by Nagalakshmi et al. The demonstrated the power of RNA-seq by sequencing the yeast transcriptome.

1.1 Former Studies

Genes are used to be defined by

1. large open reading frames (ORFs).
2. Sequence conservation through different species.
3. cDNA probing of DNA tiling microarrays.

These methods often fail to identify short exons, do not precisely reveal the boundaries of untranslated regions (UTRs), and/or have high false-positive rates

1.2 RNA-Seq workflow

1.2.1 Experimental Preperation for Sequencing

Detecting transcriptomes by sequencing.

1. Isolation of polyA RNA.
2. Generation of double-strand complementary DNA (cDNA) by reverse transcription.
3. Fragment double-strand cDNA for Illumina sequencing.

1.2.2 Analyzing Sequencing Results

1. Merging the replicates due to high resemblance. (Pearson coefficient ρ indicates the resemblance. $|\rho|$ closer to 1 means greater resemblance.)
2. Mapping the sequencing results to the reference genome. They also performed gene ontology to check the fucntions of those genes.

1.3 Results From Analysis

1. Being able to determine the 5' and 3' boundaries of a gene correctly at a lower cost. Thus being able to find the untranslated regions on both 5' and 3' ends (5' UTR and 3' UTR).
2. Determining the correct start site of a gene.
3. Finding some of the introns are actually expressed. Implying that these genes are probably not sliced at appreciable levels in vegetative cells.
4. Finding upstream ORFs (uORFs) for some of the genes. Suggesting that these genes are likely to be highly regulated.
5. Finding the transcription in intergenetic regions.
6. RNA-seq is also a quantitative method, with high correlation with QPCR. The data of QPCR and RNA-seq data have a strong correlation ($R = 0.98$). Is much better than the other methods.

1.4 Conclusions

The advantages of RNA-seq are that:

1. It allows interrogation of all unique sequences of the genome, including those that are closely related, as long as unique bases exist that can be monitored.
2. It is very sensitive and offers a large dynamic range. It can detect and quantify levels of RNAs expressed at very low levels.
3. It can accurately determine exon boundaries.

The new RNA-seq method allowed scientists to map the transcriptional landscape of yeast genomes and define UTRs and previously unknown transcribed regions.

2 RNA-seq: a revolutionary tool for transcriptomics

This article is a review article by the same group who wrote the first article. This article is more detailed in the advantages and disadvantages of RNA-seq, and also a deeper view into transcriptomics.

2.1 Introduction

The key aims of transcriptomics are:

1. to catalogue all species of transcript, including mRNAs, non-coding RNAs and small RNAs
2. to determine the transcriptional structure of genes, in terms of their start sites, 5' and 3' ends, splicing patterns and other post-transcriptional modifications
3. to quantify the changing expression levels of each transcript during development and under different conditions

Limitations of previous *microarray* methods:

- Reliance upon existing knowledge about the genome sequence.
- High background levels owing to cross-hybridization.
- Limit dynamic range of detection.
- Comparing expression levels across different experiments is often difficult and can require complicated normalization methods.

Limitations of previous *tag-based sequencing* methods:

- Sanger sequencing method is quite expensive. A significant portion of the short tags cannot be uniquely mapped to the reference genome.
- Only a portion of the transcript is analysed and isoforms are generally indistinguishable.

2.2 RNA-seq Benefits

The workflow this article describes is the same with the previous one and can be found in Section 1.2.

The benefits of RNA-seq are:

- Not limited to detecting transcripts that correspond to existing genomic sequence. So it is particularly useful for non-model organisms whose genomic sequences are not yet to be determined.
- Very low background signal because they are mapped unambiguously to the unique regions on the genome.
- Shown to be highly accurate for quantifying expression levels.
- Requires less RNA samples.

2.3 RNA-seq Challenges

Despite all the benefits of RNA-seq, it still has some drawbacks and future scientists need to overcome these challenges (some of them are already solved).

Challenges during library construction

1. Large RNA molecules need to be fragmented into smaller pieces (200-500bp) due to limitations in sequencing technologies. The fragmentation methods might be biased in different ways.
2. Large amount of short reads might result from PCR artifacts or they are a genuine reflection of abundant RNA species. One way to solve this problem is to compare the biological replicates.
3. Difficulties in building strand specific libraries.

Challenges during bioinformatics analysis

- Common informatics problems like the lacking the efficient methods to store, retrieve and process large amounts of data.
- Mapping short reads from RNA-seq data to the reference genome, or assembling them into contigs and then aligning them to the reference genome.
- Reads that span exon junctions or containing polyA tails need to be analysed separately. Exon junction spanning reads are difficult to analyze due to the presence of alternative splicing and trans-splicing.
- Short reads might match to multiple positions on the genome. Obtaining longer reads might solve this problem.
- Sequencing errors and polymorphisms are a problem while we are mapping data to the genome.

Challenges between balancing coverage and cost

A better coverage requires more sequencing depth. Which means we need more reads. Thus elevating the cost of the whole experiment.

2.4 RNA-seq Provides New Transcriptomic Insights

With the high resolution and sensitivity of RNA-seq, we are able to gain more insights into the field of transcriptomics. RNA-seq can be used for:

- Mapping gene and exon boundaries. Also discovering things like the UTR, uORFs, which are previously unknown to humans.
- Studying the splicing diversity by searching for reads that span exon junctions as well as discovering new ones.
- Finding new transcriptions with high accuracy.

2.5 Defining Transcription Level

As a quantitative method, RNA-seq can be used to determine RNA expression levels accurately.

It can be used to capture transcriptome dynamics across different types of tissues of the same individual.

2.6 The Future of RNA-seq

- Single-cell RNA-seq.
- Lowering sequencing cost.
- Determining structure and dynamics of transcriptomes.
- And much more...

Last Updated: November 30, 2022