Introduction to RNAseq Methods

June 2021

HTS Applications - Overview

DNA Sequencing

- Genome Assembly
- SNPs/SVs/CNVs
- DNA methylation
- DNA-protein interactions (ChIPseq)
- Chromatin Modification (ATAC-seq/ChIPseq)

RNA Sequencing

- Transcriptome Assembly
- **Differential Gene Expression**
- Fusion Genes
- Splice variants

Single-Cell

RNAseq Workflow

Experimental Design

Library Preparation

Sequencing

Bioinformatics Analysis

Image adapted from: Wang, Z., et al. (2009), Nature Reviews Genetics, 10, 57–63.

Designing the right experiment

A good experiment should:

- Have clear objectives
- Have sufficient power
- Be amenable to statisical analysis
- Be reproducible
- More on experimental design later

Designing the right experiment

Practical considerations for RNAseq

- Coverage: how many reads?
- Read length & structure: Long or short reads? Paired or Single end?
- Controlling for batch effects
- Library preparation method: Poly-A, Ribominus, other?

Designing the right experiment - How many reads do we need?

The coverage is defined as:

$$\frac{Read\ Length \times Number\ of\ Reads}{Length\ of\ Target\ Sequence}$$

The amount of sequencing needed for a given sample is determined by the goals of the experiment and the nature of the RNA sample.

- For a general view of differential expression: 5–25 million reads per sample
- For alternative splicing and lowly expressed genes: 30–60 million reads per sample.
- In-depth view of the transcriptome/assemble new transcripts: 100–200 million reads
- Targeted RNA expression requires fewer reads.
- miRNA-Seq or Small RNA Analysis require even fewer reads.

Designing the right experiment - Read length

Long or short reads? Paired or Single end?

The answer depends on the experiment:

- Gene expression – typically just a short read e.g. 50/75 bp; SE or PE.
- kmer-based quantification of Gene Expression (Salmon etc.) - benefits from PE.
- Transcriptome Analysis – longer paired-end reads (such as 2 x 75 bp).
- Small RNA Analysis – short single read, e.f. SE50 - will need trimming.

Designing the right experiment - Replication

Biological Replication

- Measures the biological variations between individuals
- Accounts for sampling bias

Technical Replication

- Measures the variation in response quantification due to imprecision in the technique
- Accounts for technical noise

Designing the right experiment - Replication

Biological Replication

Each replicate is from an indepent biological individual

- *In Vivo*:
  - Patients
  - Mice
- *In Vitro*:
  - Different cell lines
  - Different passages

Designing the right experiment - Replication

Technical Replication

Replicates are from the same individual but processed separately

- Experimental protocol
- Measurement platform

Designing the right experiment - Batch effects

Designing the right experiment - Batch effects

Designing the right experiment - Batch effects

Designing the right experiment - Batch effects

- Batch effects are sub-groups of measurements that have qualitatively different behavior across conditions and are unrelated to the biological or scientific variables in a study.
- Batch effects are problematic if they are confounded with the experimental variable.
- Batch effects that are randomly distributed across experimental variables can be controlled for.

Designing the right experiment - Batch effects

- Batch effects are sub-groups of measurements that have qualitatively different behavior across conditions and are unrelated to the biological or scientific variables in a study.
- Batch effects are problematic if they are confounded with the experimental variable.
- Batch effects that are randomly distributed across experimental variables can be controlled for.
- Randomise all technical steps in data generation in order to avoid batch effects

Designing the right experiment - Batch effects

- Batch effects are sub-groups of measurements that have qualitatively different behavior across conditions and are unrelated to the biological or scientific variables in a study.
- Batch effects are problematic if they are confounded with the experimental variable.
- Batch effects that are randomly distributed across experimental variables can be controlled for.
- Randomise all technical steps in data generation in order to avoid batch effects

Designing the right experiment - Batch effects

- Batch effects are sub-groups of measurements that have qualitatively different behavior across conditions and are unrelated to the biological or scientific variables in a study.
- Batch effects are problematic if they are confounded with the experimental variable.
- Batch effects that are randomly distributed across experimental variables can be controlled for.
- Randomise all technical steps in data generation in order to avoid batch effects

Designing the right experiment - Batch effects

- Batch effects are sub-groups of measurements that have qualitatively different behavior across conditions and are unrelated to the biological or scientific variables in a study.
- Batch effects are problematic if they are confounded with the experimental variable.
- Batch effects that are randomly distributed across experimental variables can be controlled for.
- Randomise all technical steps in data generation in order to avoid batch effects

Designing the right experiment - Batch effects

- Batch effects are sub-groups of measurements that have qualitatively different behavior across conditions and are unrelated to the biological or scientific variables in a study.
- Batch effects are problematic if they are confounded with the experimental variable.
- Batch effects that are randomly distributed across experimental variables can be controlled for.
- Randomise all technical steps in data generation in order to avoid batch effects

Designing the right experiment - Batch effects

- Batch effects are sub-groups of measurements that have qualitatively different behavior across conditions and are unrelated to the biological or scientific variables in a study.
- Batch effects are problematic if they are confounded with the experimental variable.
- Batch effects that are randomly distributed across experimental variables can be controlled for.
- Randomise all technical steps in data generation in order to avoid batch effects

Designing the right experiment - Batch effects

Multiplexing

Designing the right experiment - Hidden Confounding variables

- Think deeply about the samples you are collecting
- This will be covered in more detail tomorrow
- Age, sex, litter, cell passage ..
- Record everything

RNAseq Workflow

Experimental Design

Library Preparation

Sequencing

Bioinformatics Analysis

Image adapted from: Wang, Z., et al. (2009), Nature Reviews Genetics, 10, 57–63.

Library preparation

- Ribosomal RNA
- Poly-A transcripts
- Other RNAs e.g. tRNA, miRNA etc.

Total RNA extraction

Library preparation

Poly-A Selection

Poly-A transcripts e.g.:

- mRNAs
- immature miRNAs
- snoRNA

Ribominus selection

Poly-A transcripts + Other mRNAs e.g.:

- tRNAs
- mature miRNAs
- piRNAs

Library preparation

Library preparation

RNAseq Workflow

Experimental Design

Library Preparation

Sequencing

Bioinformatics Analysis

Image adapted from: Wang, Z., et al. (2009), Nature Reviews Genetics, 10, 57–63.

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

Sequencing by synthesis

RNAseq Workflow

Experimental Design

Library Preparation

Sequencing

Bioinformatics Analysis

Image adapted from: Wang, Z., et al. (2009), Nature Reviews Genetics, 10, 57–63.

Case Study

->