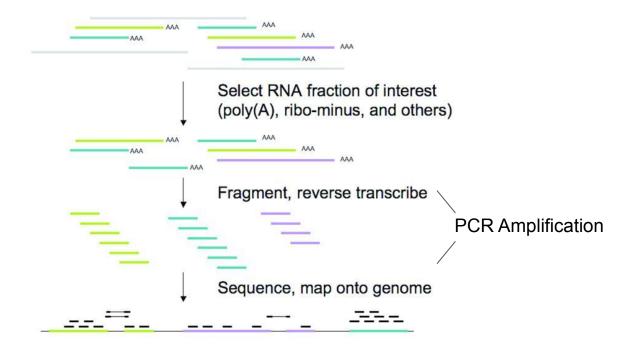
# Introduction to RNA-seq

There is no optimal pipeline for the variety of different applications and analysis scenarios in which RNA-seq can be used. Scientists plan experiments and adopt different analysis strategies depending on the organism being studied and their research goals.

- <u>Conesa et al. 2016</u>



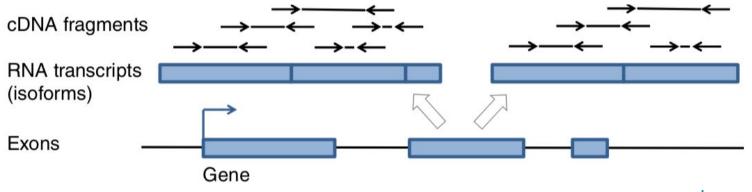


#### **Biases to be aware of**

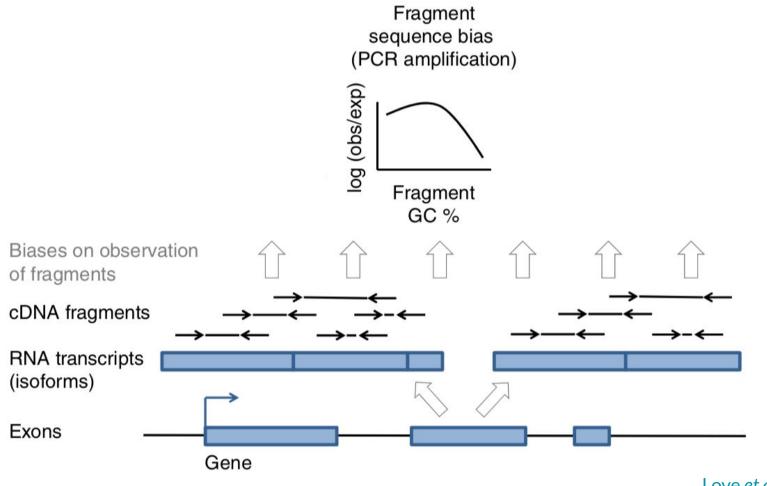
- Library size or sequencing depth the total number of reads is not always equivalent between samples
- Gene length longer genes are more likely to be observed

#### Abundance measures like TPM (Transcripts Per Million) take this into account

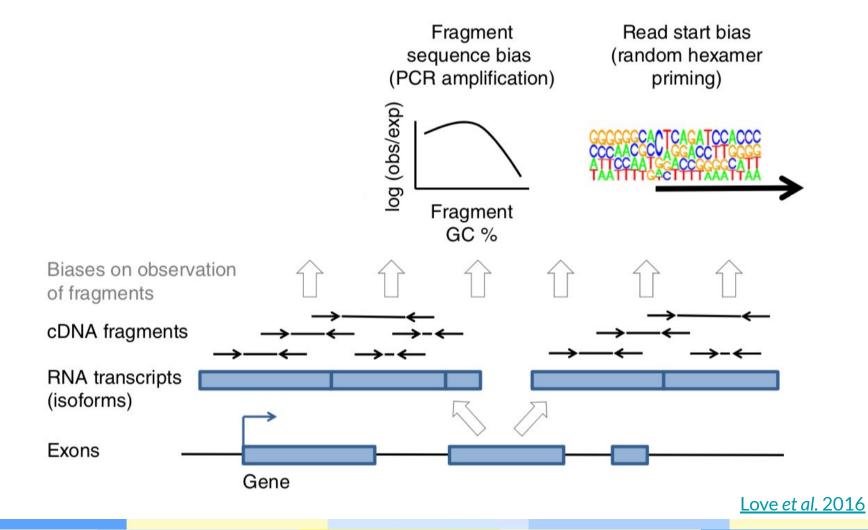
StatsQuest. RPKM, FPKM and TPM, clearly explained.

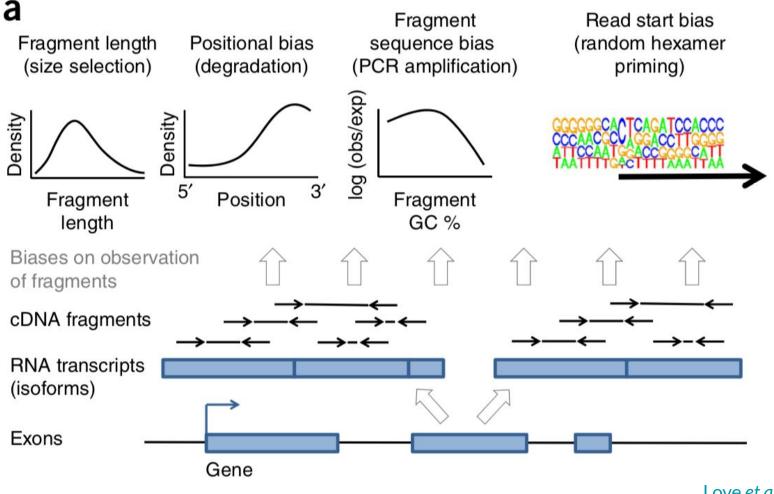


Love et al. 2016

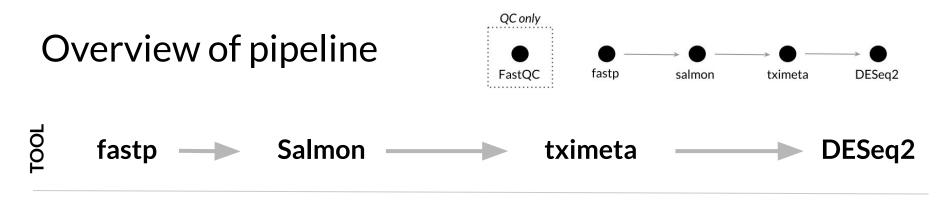


Love et al. 2016





Love et al. 2016



PURPOSE

**NPUT** FILES Adapter trimming, quality filtering, length filtering Quantification of transcripts via lightweight mapping to *transcriptome*, GC-bias correction Import of transcript (tx) abundances and counts from Salmon and summary to the gene-level for more robust statistics, accounts for gene length changes across samples due to differential isoform usage

Library size and composition normalization, transformation for visualization and clustering, testing for differential gene expression

FASTQ

FASTQ that have been preprocessed with fastp, transcriptome to map against Estimated counts and abundances from Salmon (transcript metadata is added automatically)

SummarizedExperiment R object which contains unnormalized counts and length information

## What you'll learn to do in this module

- Perform quality control checks with FastQC (<u>Andrews</u>)
- Perform FASTQ preprocessing with fastp (<u>Chen *et al.* 2018</u>)
- Quantify transcripts with Salmon (<u>Patro et al. 2017</u>)
- Import quantification estimates with tximeta and summarize to the gene level (Love et al. 2020; Soneson et al. 2015)
- Perform exploratory data analysis with DESeq2 (Love *et al.* 2014)
- Perform differential expression analysis with DESeq2
- Make fancy volcano plots and fancy heatmaps (<u>Blighe et al.</u>; <u>Gu 2016</u>)

### **Tool-specific tutorials**

**Getting Started with Salmon** 

Tximeta: transcript quantification import with automatic metadata

Note: if you are not using Salmon, you can't use tximeta (for now) so you will want to look at tximport: <u>Importing transcript abundance datasets with tximport</u>

Analyzing RNA-seq data with DESeq2

### Links to follow-up information

StatQuest Video: A Gentle Guide to RNA-seq

StatQuest Video: RPKM, FPKM, and TPM

StatQuest Video: DESeq2, part 1, Library Normalization

<u>Hansen et al. Biases in Illumina transcriptome sequencing caused by random</u> <u>hexamer priming. *Nucleic Acid Research*. 2010.</u>

Michigan State University Research Technology Support Facility "FastQC Tutorial & FAQ"