

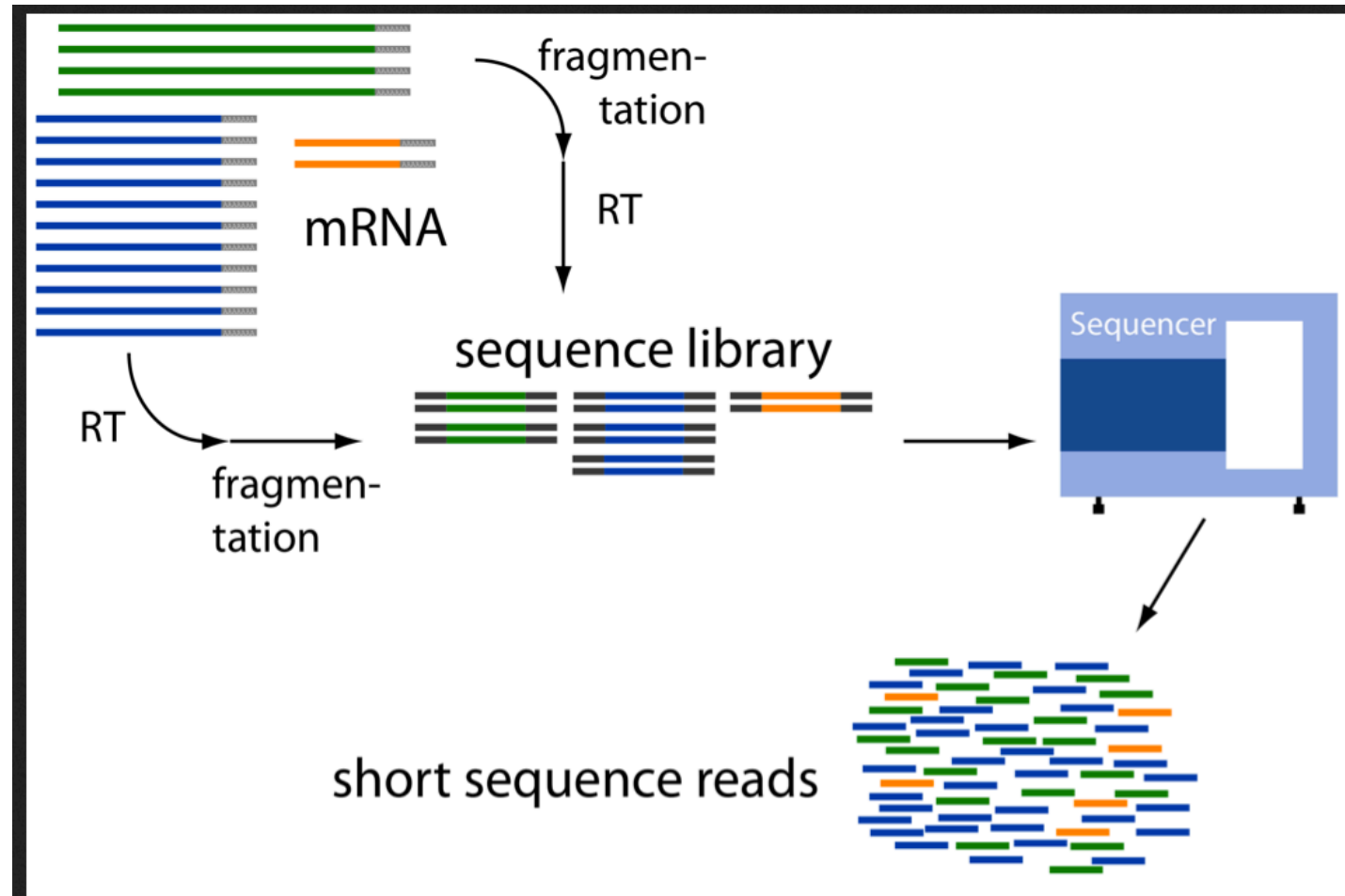
# RNA-seq introduction

---

RNA-seq data analysis

Johan Reimegård | 30-November-2020

# How are RNA-seq data generated?



Sampling process

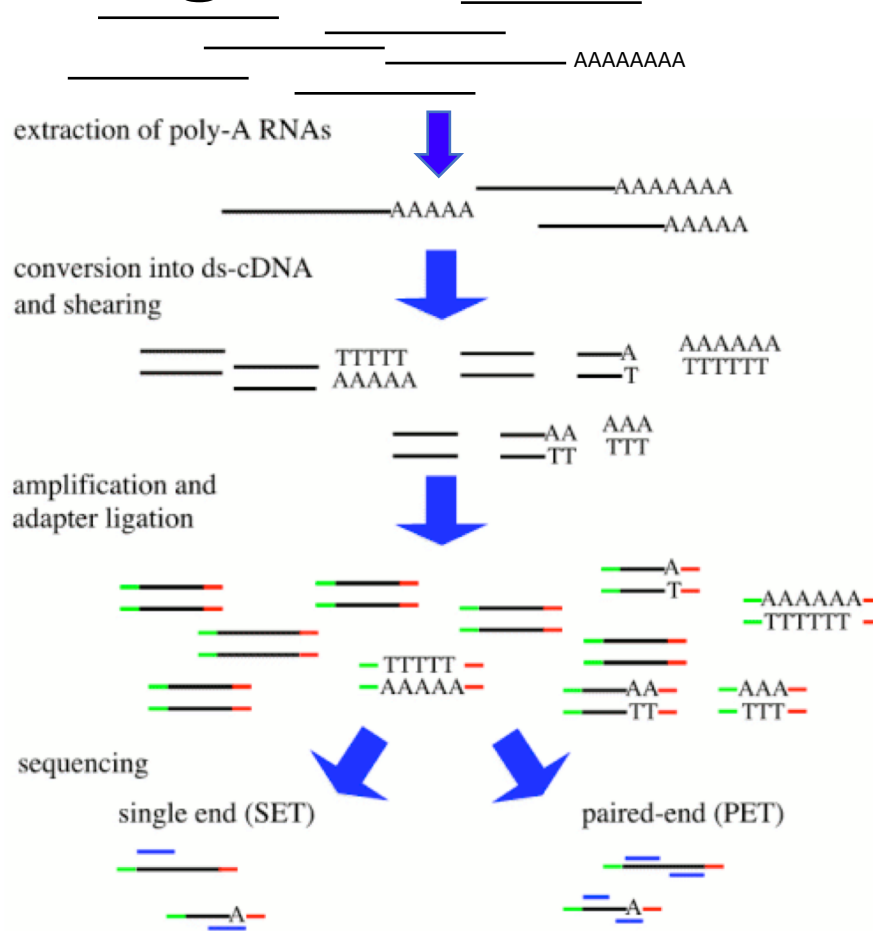
# Depending on the different steps you will get different results

RNA->

enrichments ->

library ->

reads ->



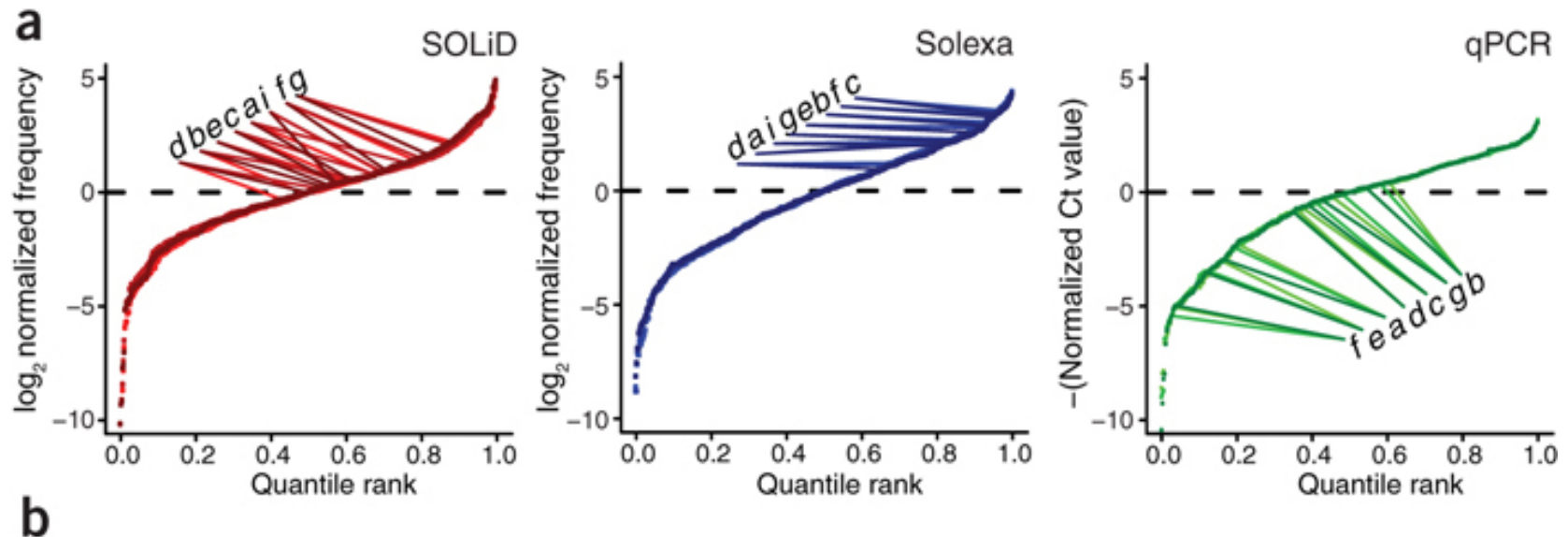
PolyA	(mRNA)
RiboMinus	(- rRNA)
Size <50 nt	(miRNA )
.....	

Size of fragment
Strand specific
5' end specific
3' end specific
.....

Single end (1 read per fragment)
Paired end (2 reads per fragment)

# Paired end read

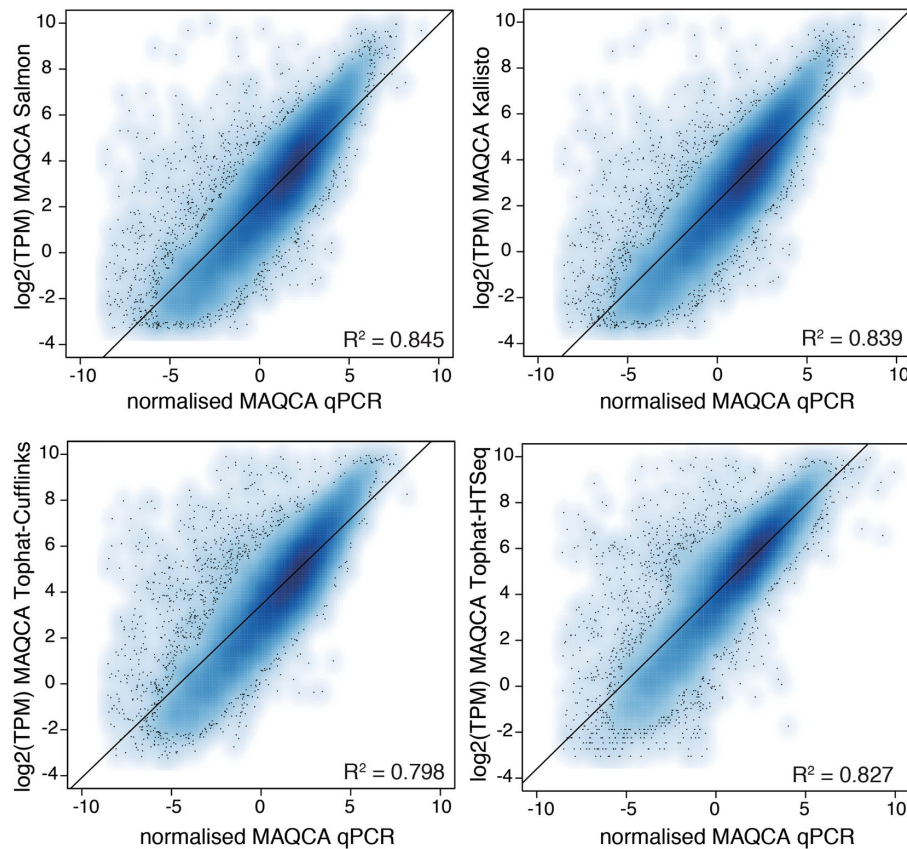
# Different sequencing techniques have different preferences



Sequencing frequency of 472 artificial miRNAs in equal abundance

(Figure from Linsen *et al.*,  
Nature Methods. 2009)

# But evens out over longer RNAs



**Figure 1.** Gene expression correlation between RT-qPCR and RNA-seq data. The Pearson correlation coefficients and linear regression line are indicated. Results are based on RNA-seq data from dataset 1.

Benchmarking of RNA-sequencing analysis workflows using whole transcriptome RT-qPCR expression data

# Fastq – read file format

The diagram shows a single line of a Fastq file. Three blue arrows point from labels to specific parts of the line: 'Unique identifier' points to '@SEQ\_ID', 'Sequence' points to the sequence string 'GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT', and 'Sequence quality' points to the quality string '!''\*(((((\*\*\*+))%%%+)) (%%%) .1\*\*\*-+\*'''))\*\*55CCF>>>>>CCCCCCC65'.

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*(((((***+))%%%+)) (%%%) .1***-+*'''))**55CCF>>>>>CCCCCCC65
```

Paired end data usually in format sampleX\_1.fastq and sampleX\_2.fastq with same SEQ\_ID for both mate pairs, followed by /1 and /2 (or \_f and \_r)

# Phret-score





**Thank you. Questions?**

---

**Johan Reimegård | 13-May-2019**