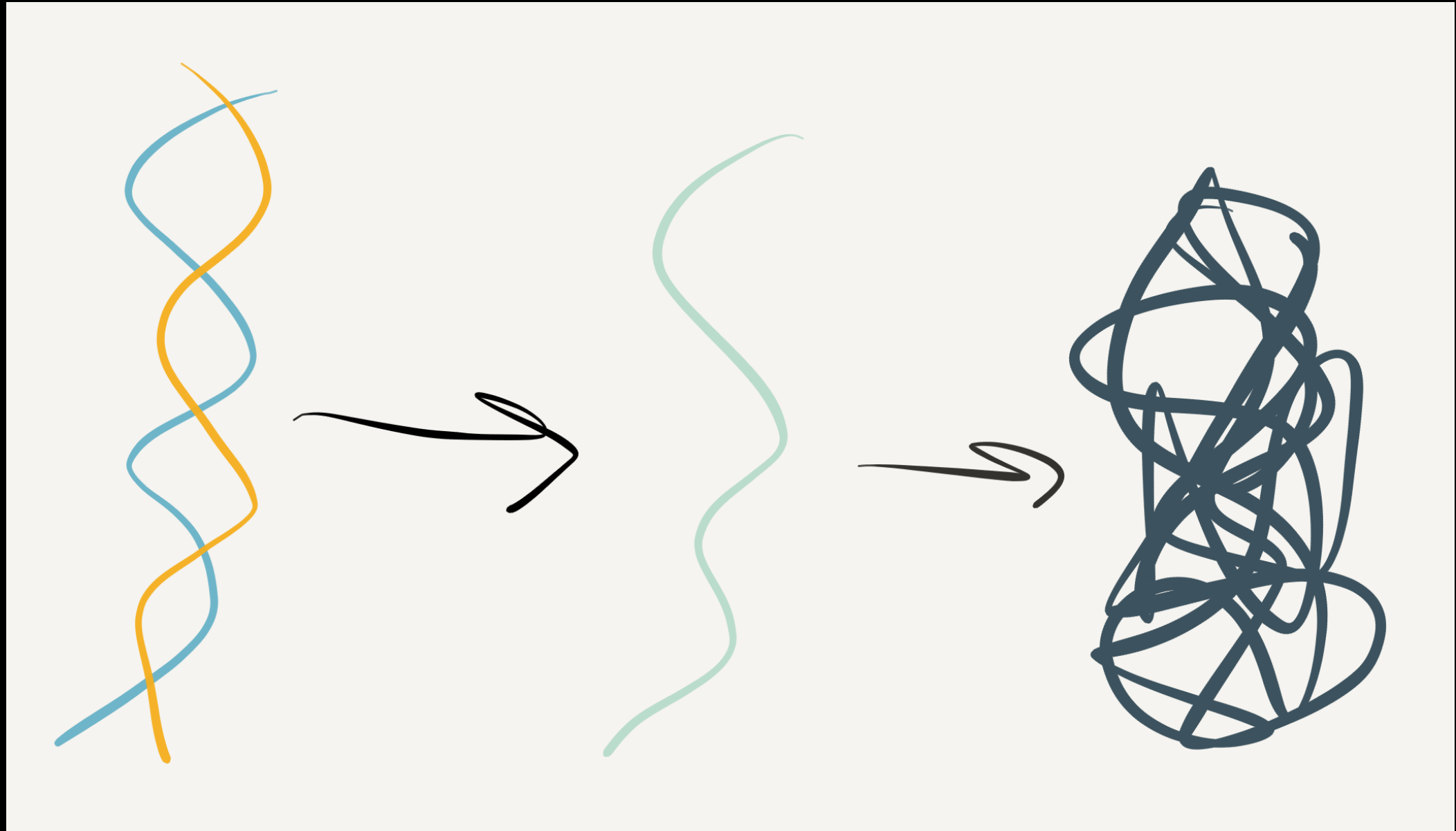
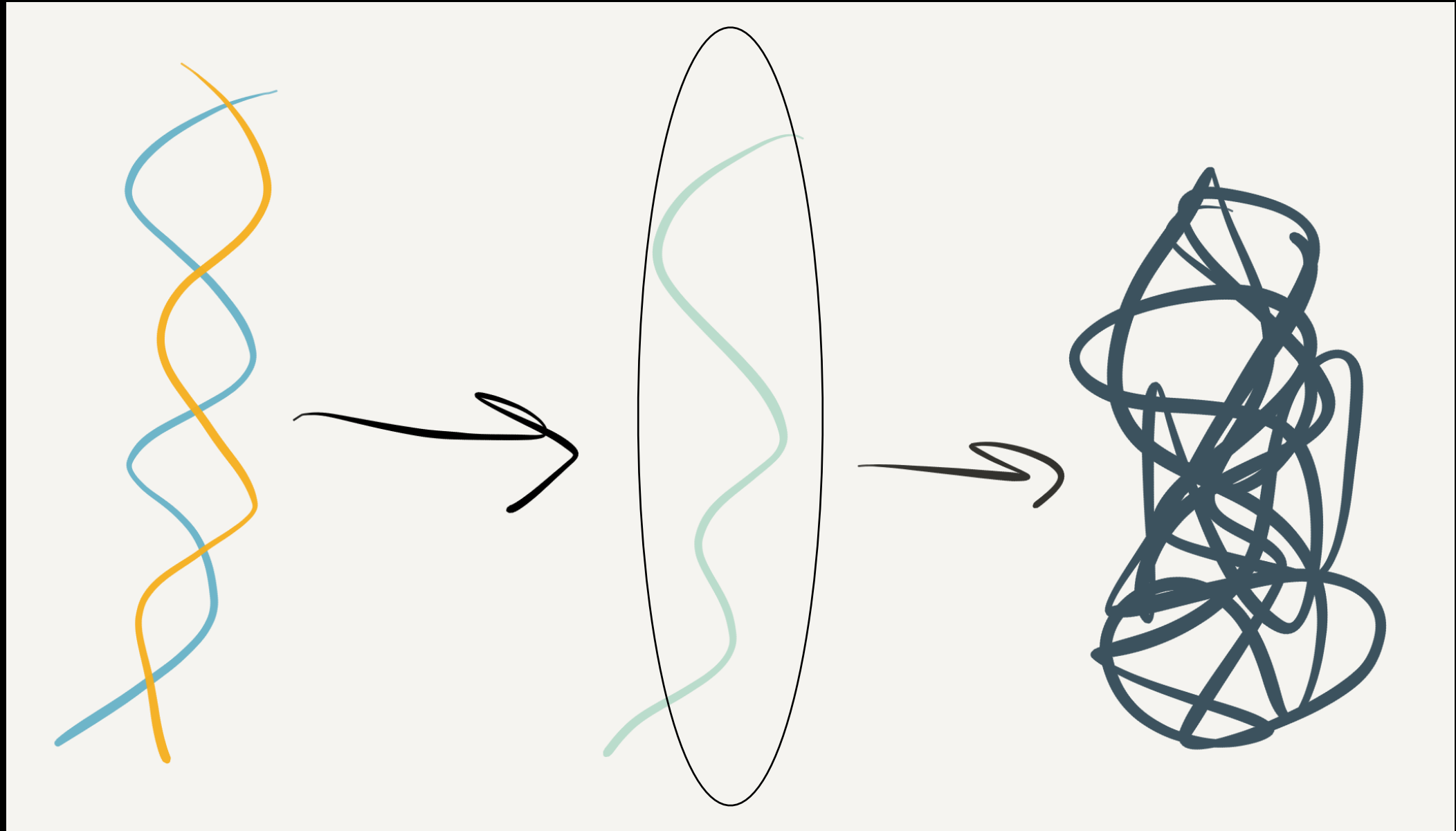

INTRO TO RNASEQ AND QUALITY CONTROL/TRIMMING

RNASEQ

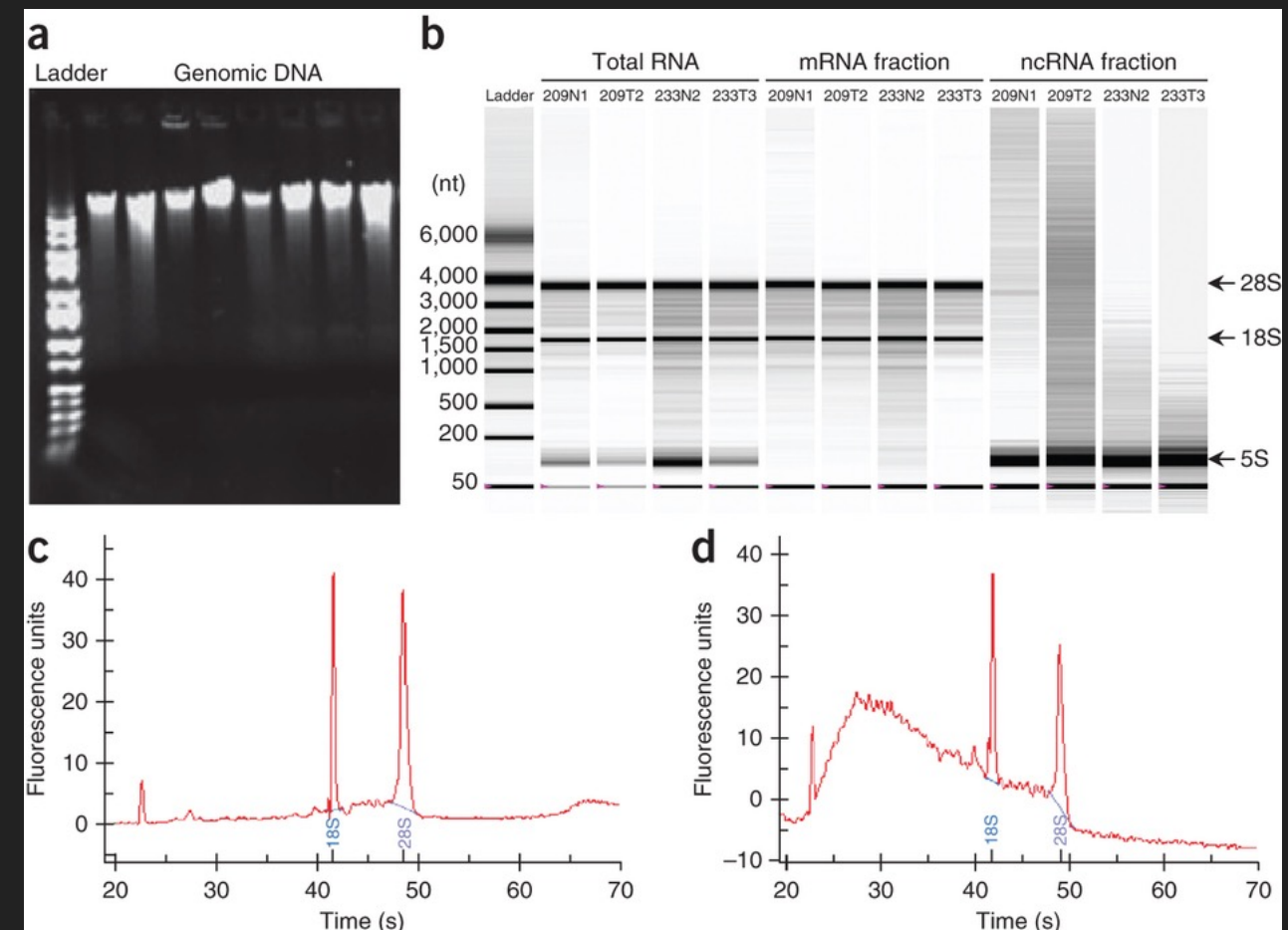


RNASEQ



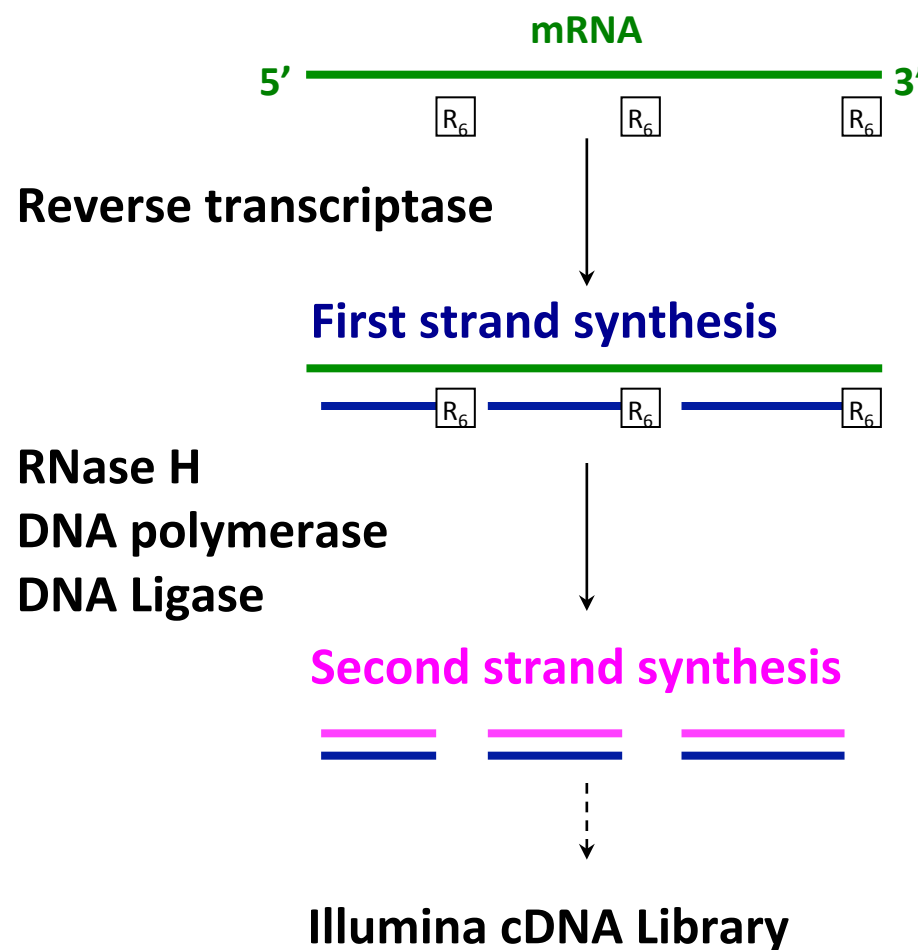
RNA EXTRACTION METHODS:

- ▶ Dependent on starting material:
Tissue type, Amount
- ▶ End product: mRNA or total RNA?
 - ▶ Total RNA: Trizol vs. Columns (Promega, RNeasy, ect.)
 - ▶ mRNA: direct extraction or isolation from total RNA



RNA-Seq: How do we make cDNA?

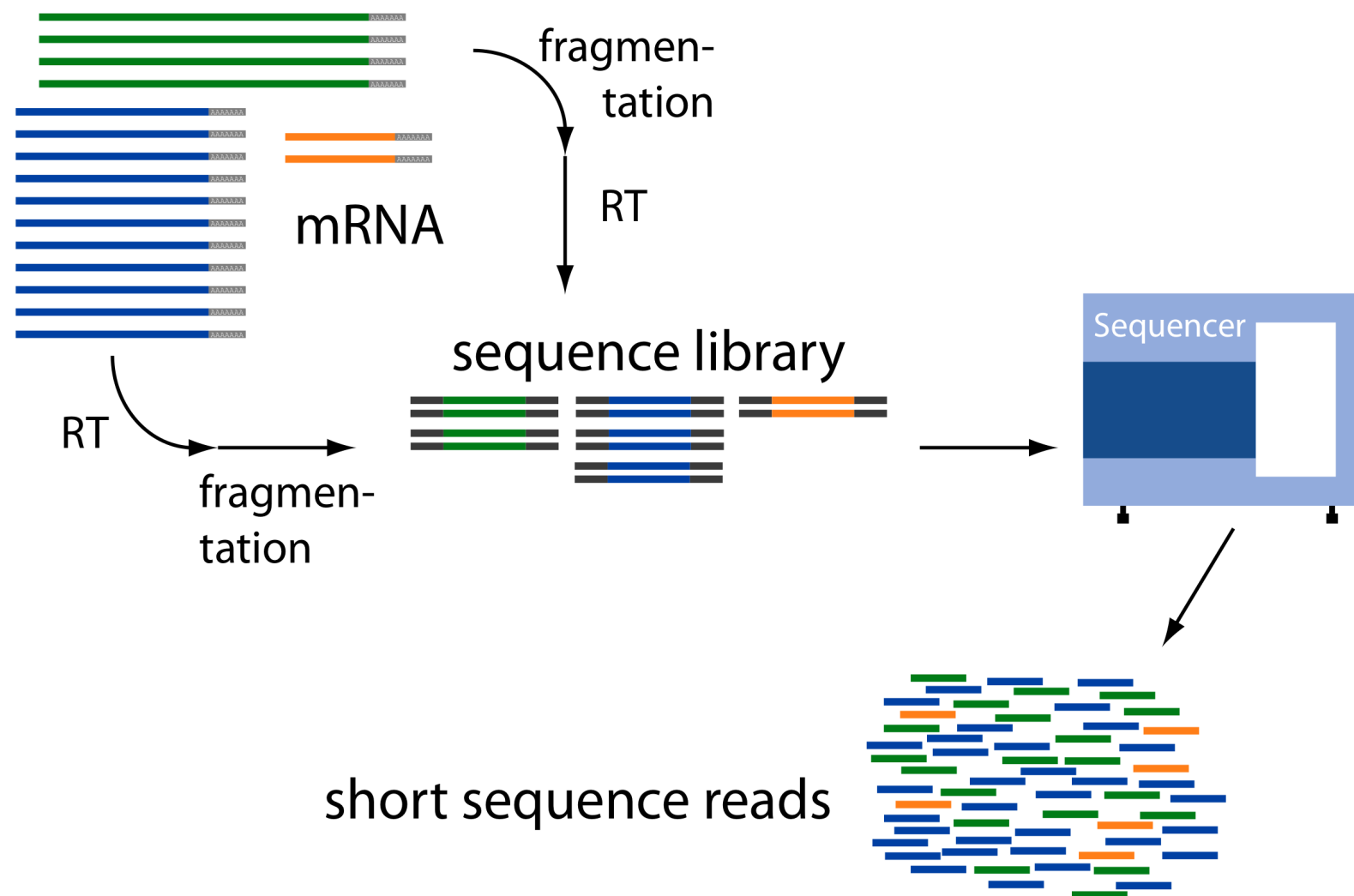
Prime with Random Hexamers (R6)



Slide courtesy of Joshua Levin, Broad Institute.

Slide from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

Overview of RNA-Seq



From: <http://www2.fml.tuebingen.mpg.de/raetsch/members/research/transcriptomics.html>

Slide from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

Common Data Formats for RNA-Seq

FASTA format:

```
>61DFRAAXX100204:1:100:10494:3070/1
AAACAACAGGGCACATTGTCACTCTTGTATTTGAAAAACACTTTCCGGCCAT
```

FASTQ format:

```
@61DFRAAXX100204:1:100:10494:3070/1
AAACAACAGGGCACATTGTCACTCTTGTATTTGAAAAACACTTTCCGGCCAT
+
ACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCBC?CCCCCCCC@@@CACCCCCA
```

Read

Quality values

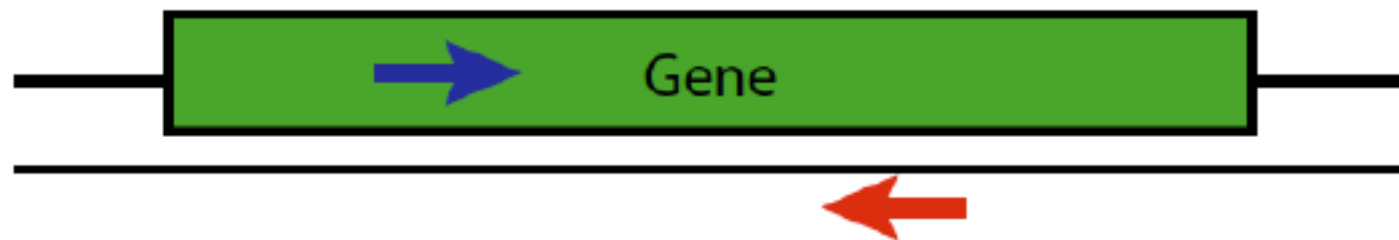
$\text{AsciiEncodedQual}(x) = -10 * \log_{10}(\text{Pwrong}(x)) + 33$



$\text{AsciiEncodedQual}('C') = 64$

So, $\text{Pwrong}('C') = 10^{(64-33/(-10))} = 10^{-3.4} = \mathbf{0.0004}$

Paired-end Sequences

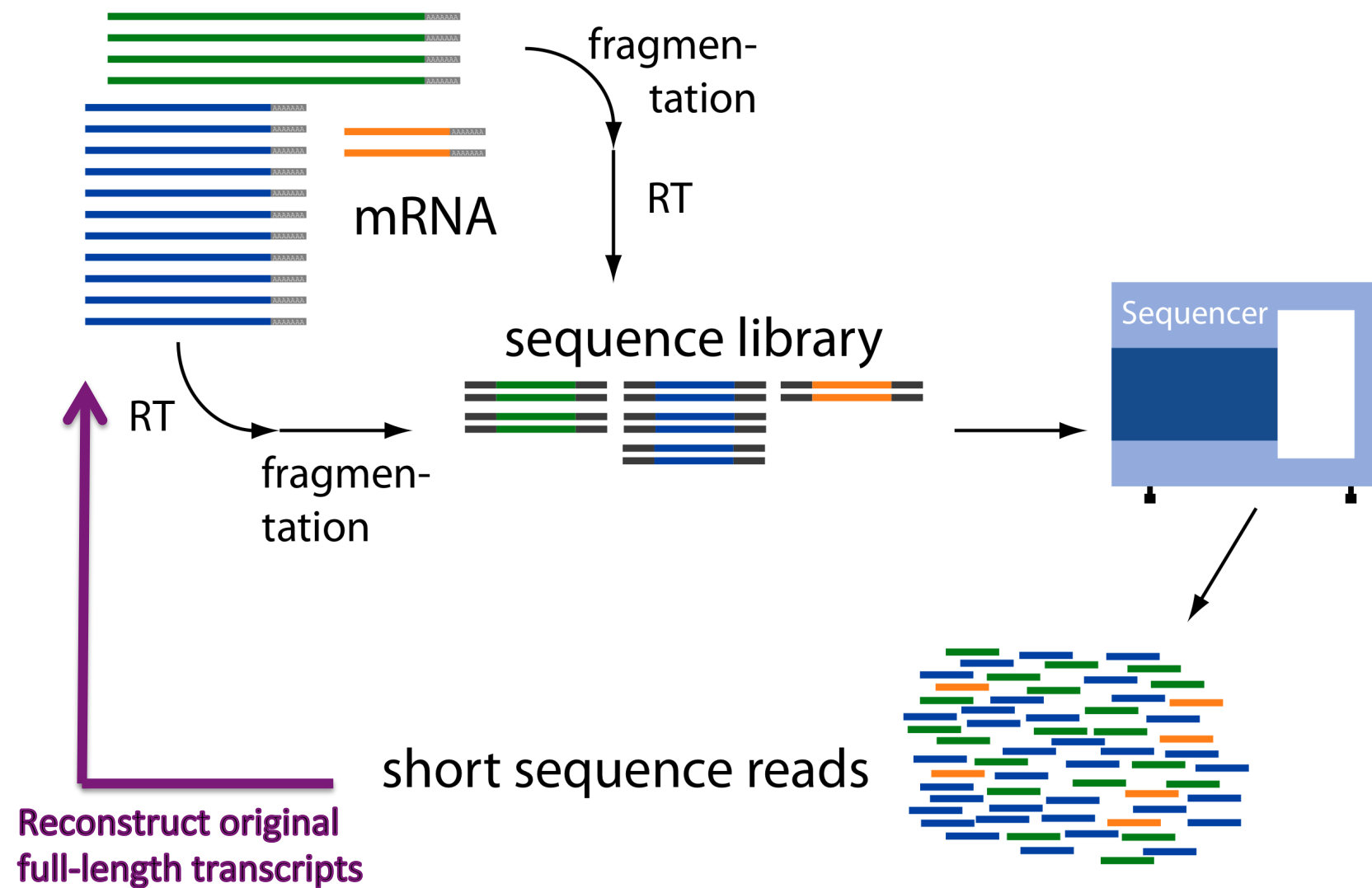


Two FastQ files, read name indicates
left (/1) or right (/2) read of paired-end

```
@61DFRAAXX100204:1:100:10494:3070/1
AAACAACAGGGCACATTGTCACCTCTTGTATTTGAAAAACACTTTCCGGCCAT
+
ACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCBC?CCCCCCCC@@@CACCCCCA
```

```
@61DFRAAXX100204:1:100:10494:3070/2
CTCAAATGGTTAATTCTCAGGCTGCAAATATTCGTTTCAGGATGGAAGAACA
+
C<CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCBCCCC
```


Overview of RNA-Seq



From: <http://www2.fml.tuebingen.mpg.de/raetsch/members/research/transcriptomics.html>

Slide from: https://github.com/trinityrnaseq/RNASeq_Trinity_Tuxedo_Workshop/blob/master/docs/rnaseq_workshop_slides.pdf

COMMON RNASEQ APPLICATIONS

- ▶ Raw genetic data
- ▶ Population genomics
- ▶ Phylogenomics
- ▶ Experimental biology
- ▶ Differential expression analysis

Journal of Ornithology (2018) 159:599–629
<https://doi.org/10.1007/s10336-018-1532-5>

REVIEW

Avian transcriptomics: opportunities and challenges

Elinor Jax^{1,2}  · Michael Wink³ · Robert H. S. Kraus^{1,2}

Received: 29 September 2017 / Revised: 27 December 2017 / Accepted: 15 January 2018 / Published online: 5 February 2018
 © The Author(s) 2018. This article is an open access publication

Journal of Ornithology (2018) 159:599–629

615

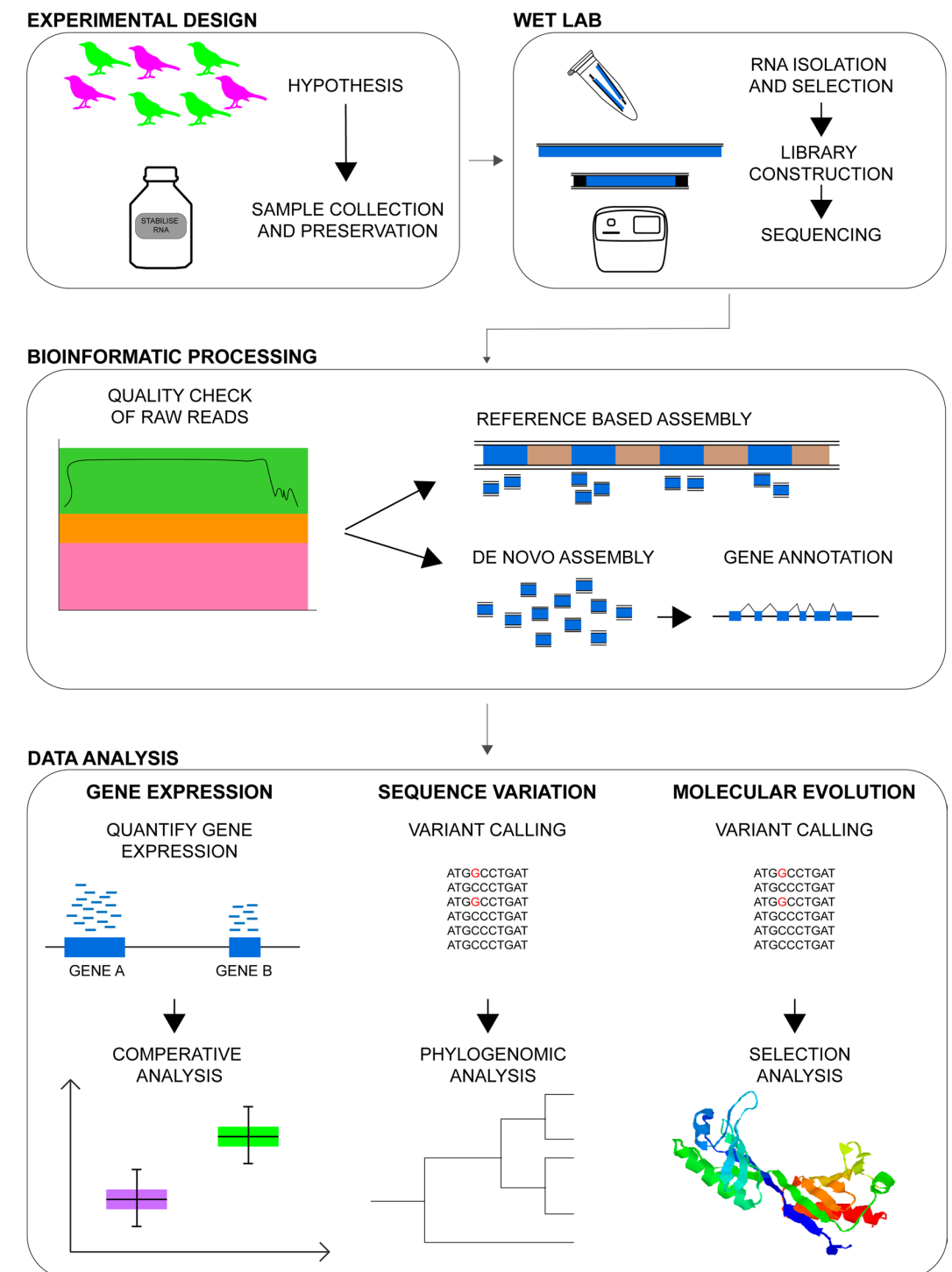
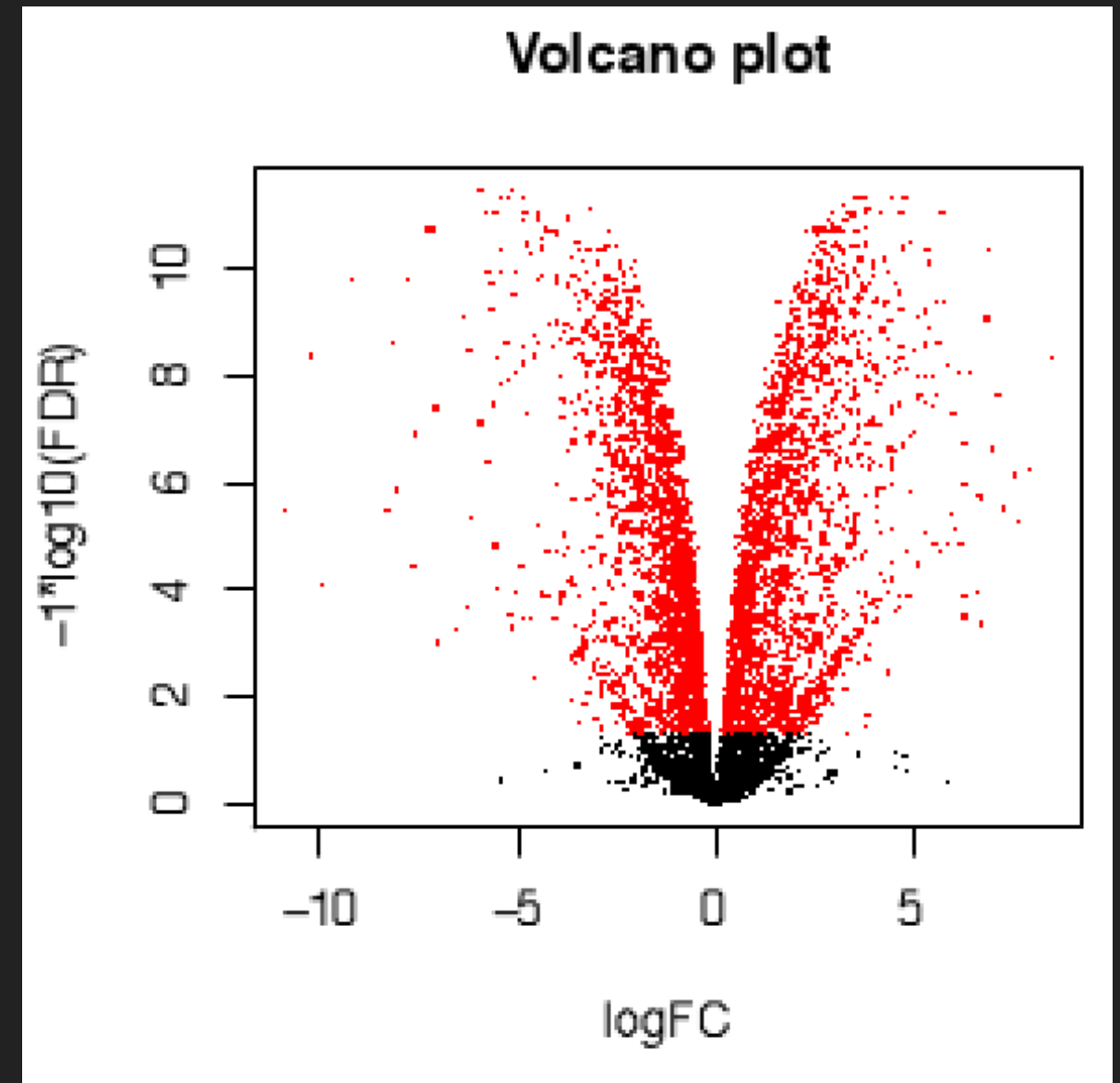


Fig. 2 Overview of a typical RNA sequencing workflow. The individual elements of the workflow are detailed in “Planning an RNA-seq study: a quick guide”. Protein structure by Richard Wheeler (Zephyris), licensed under Creative Commons 3.0, Wikimedia Commons

COMMON RNASEQ APPLICATIONS

- ▶ Raw genetic data
 - ▶ Population genomics
 - ▶ Phylogenomics
- ▶ Experimental biology
 - ▶ Differential expression analysis



QUALITY CHECK SHORT READ SEQUENCE

- ▶ FASTQC "A quality control tool for high throughput sequence data"
- ▶ <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

TRIMMING

- ▶ Many different trimmers available
- ▶ Trim adapters/low quality reads
- ▶ Trim Galore! is one that works well on Hydra

HANDS-ON

- ▶ Go to https://github.com/SmithsonianWorkshops/SMSC_2019_Conservation_Genomics/tree/master/Day8_RNAseq_assembly_and_analysis
- ▶ Click on "2a_Raw Read QA-QC.md"