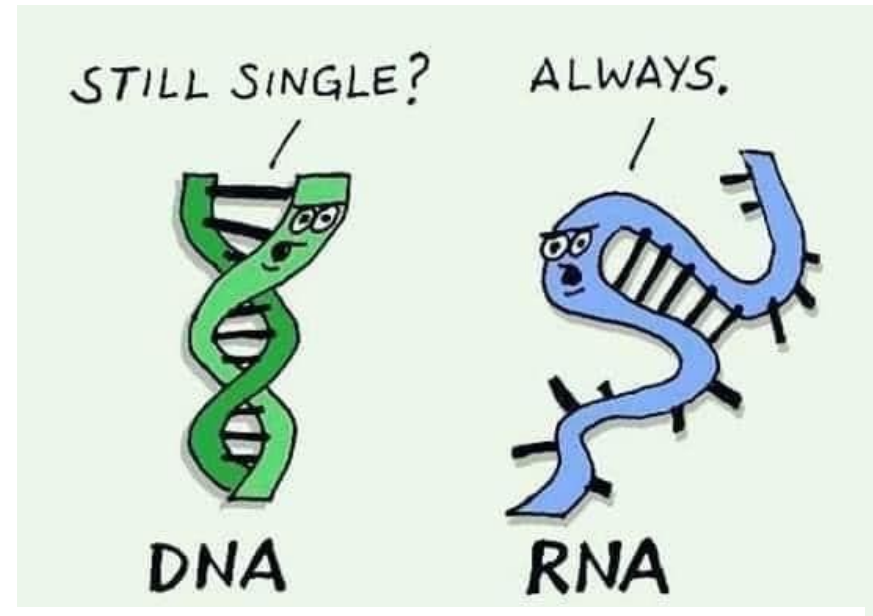




Transcriptomics

What is transcriptome?

All RNA being transcribed
at a certain stage in development
in a certain type of cells
in response to certain stimuli
...



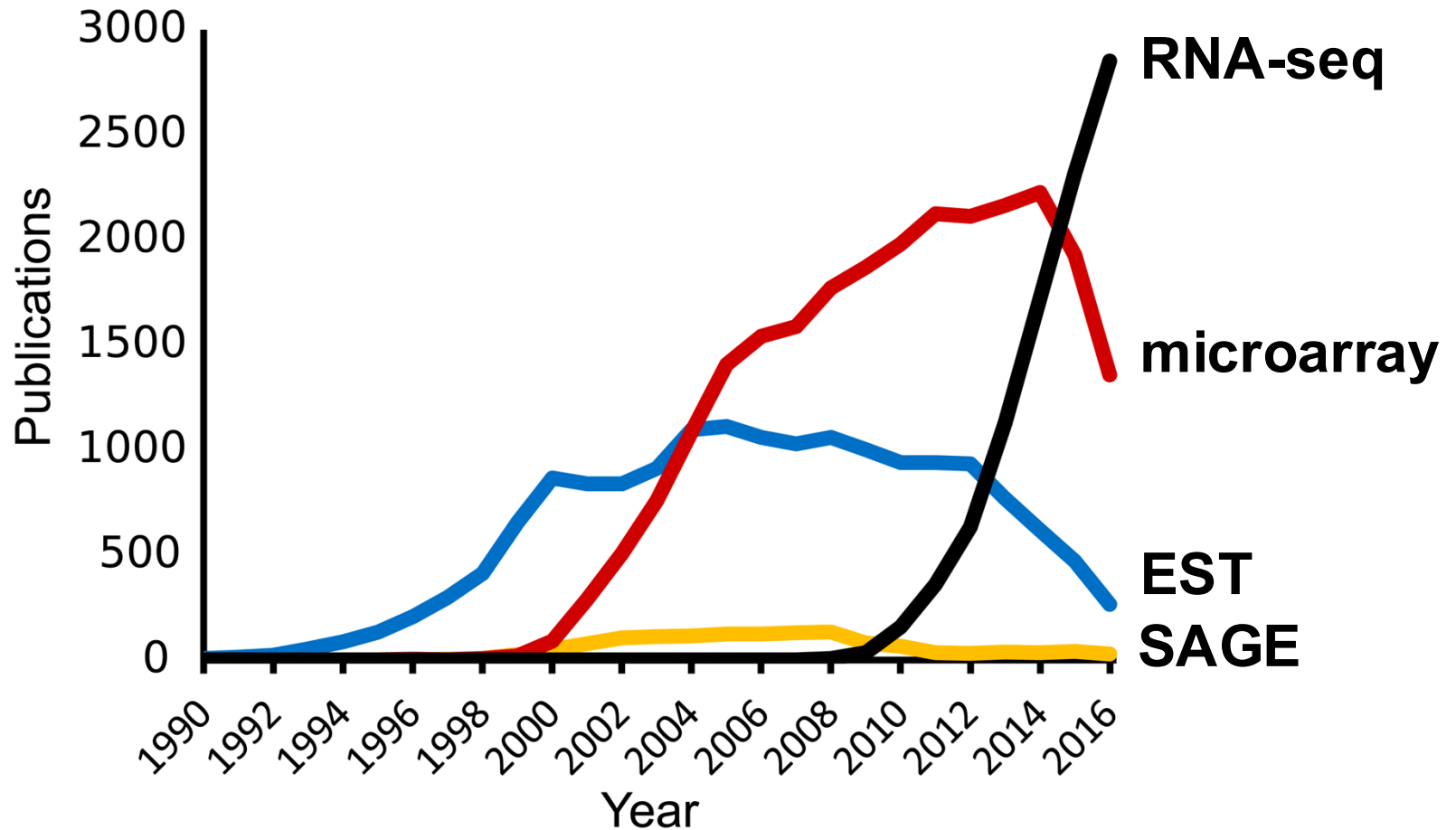
Tyler Gable siRNA, miRNA, ceRNA, piRNA, piRNA-like RNA, pesRNA, many viral RNAs ALL DISAGREE.

Like · Reply · 4d

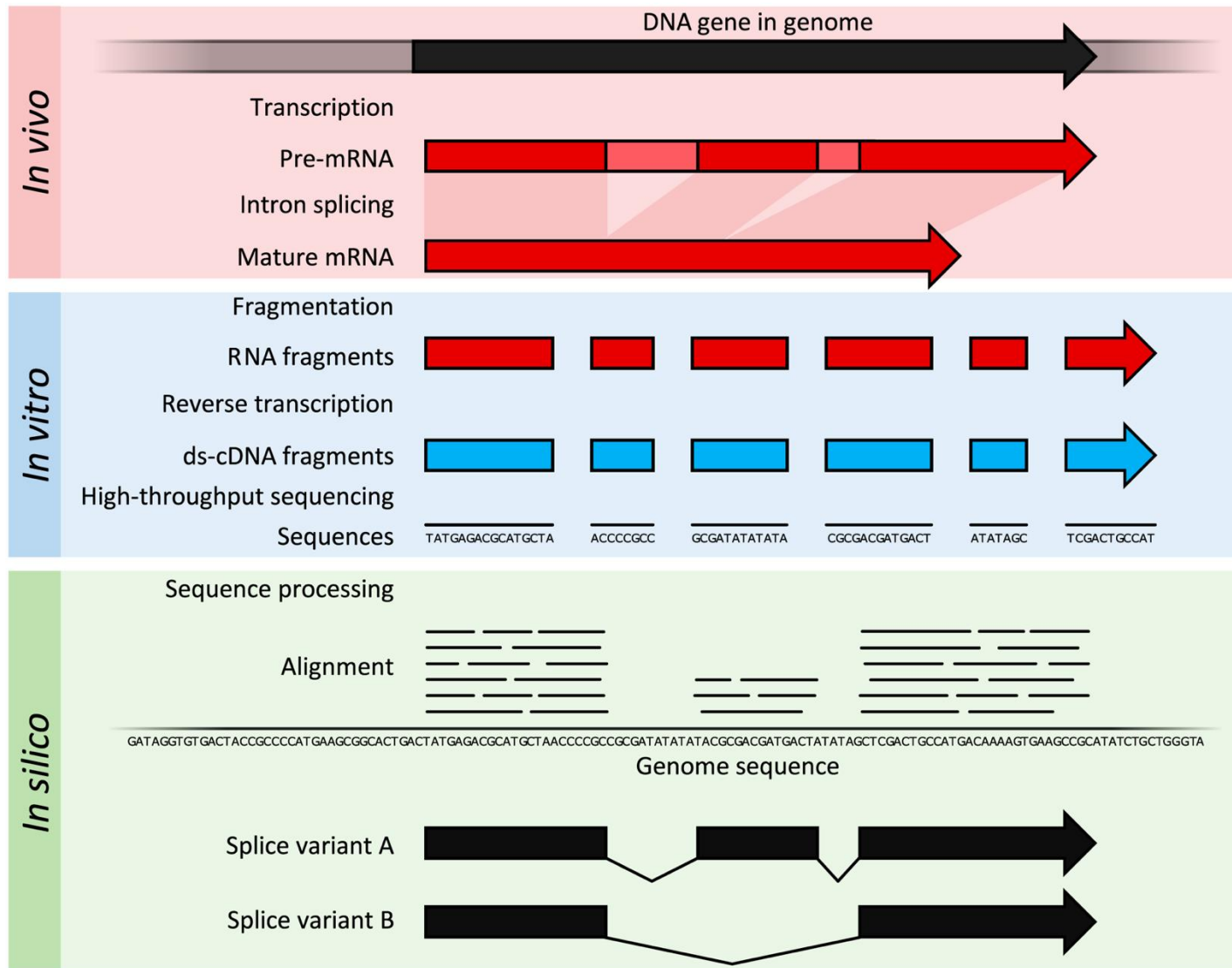


Image from: <https://www.facebook.com/trust.biologist/>

How have researchers been studying transcriptomes?



What RNA-seq sequences represent



Common usage of RNA-seq data

Gene expression study

e.g. differential expression, time course profile

Profiling total RNA (e.g. miRNA and mRNA)

e.g. in exosomes and other secretory products

Splice isoform

only useful for organism with polished reference genomes

SNP calling

use transcriptome as a reduced subset of genomic variation study

Profiling genes in an organism

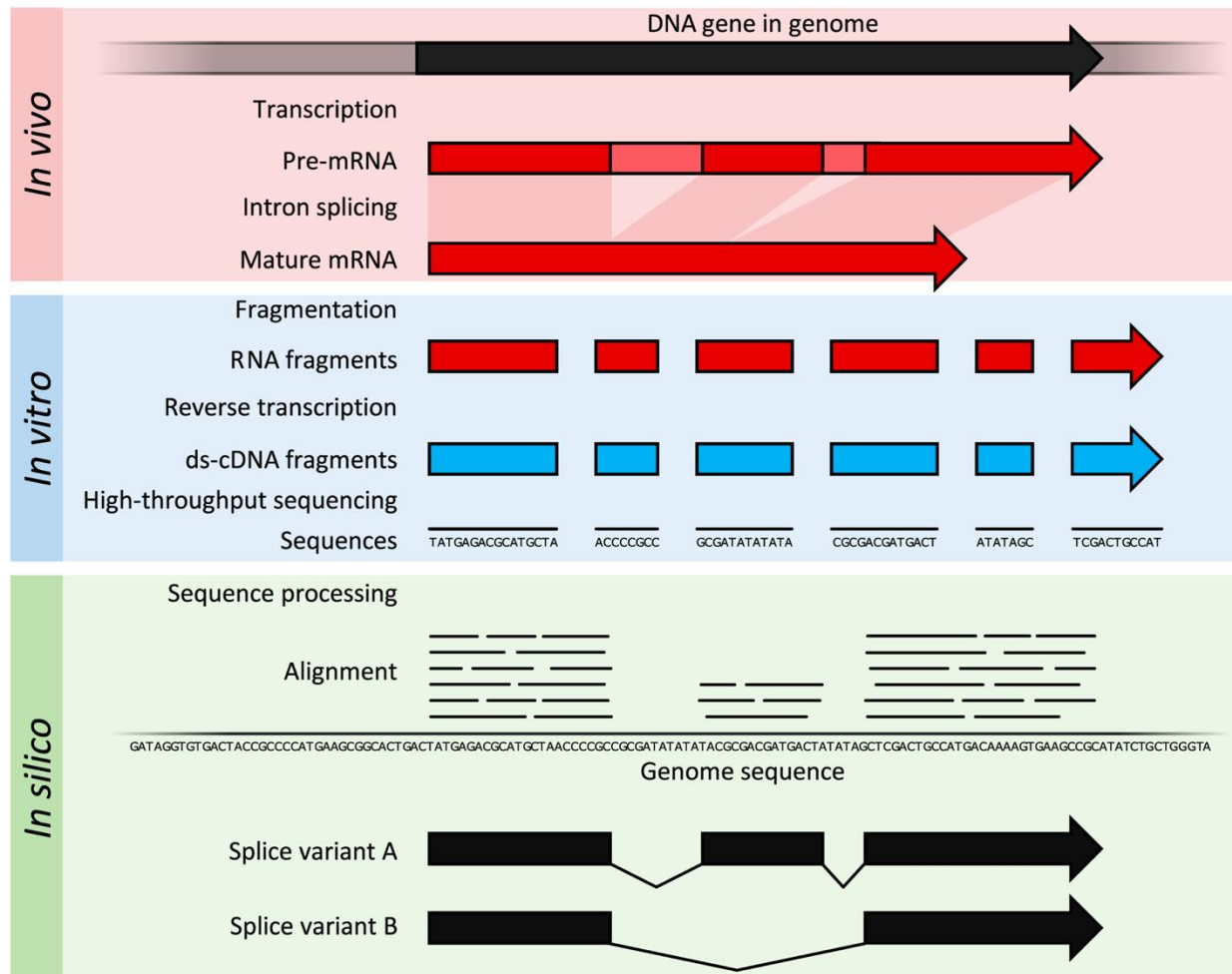
e.g. for gene annotation, refining gene model

Terms you might come across

number of reads

strand-specific

single-end/pair-end



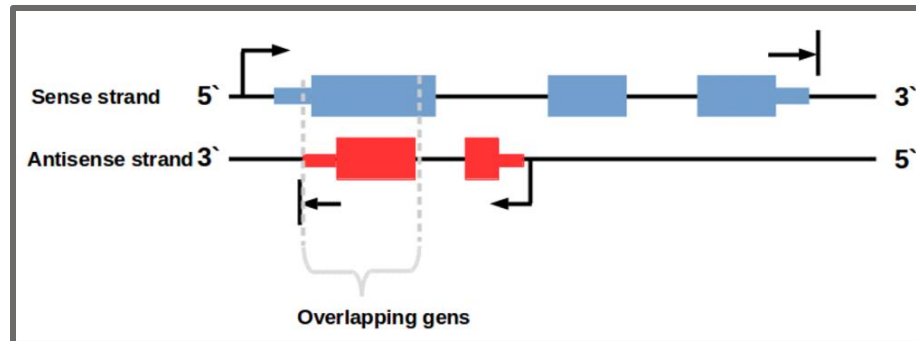
Terms you might come across

number of reads

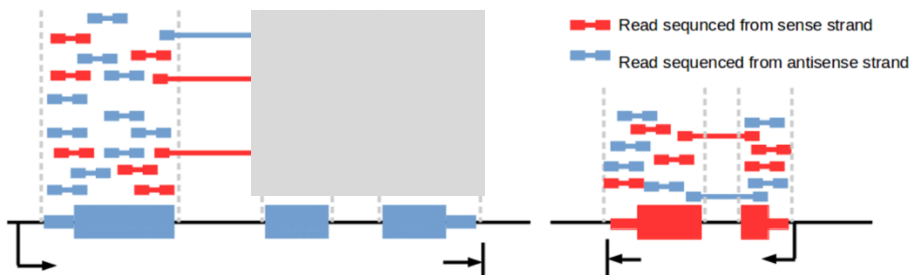
strand-specific

single-end/pair-end

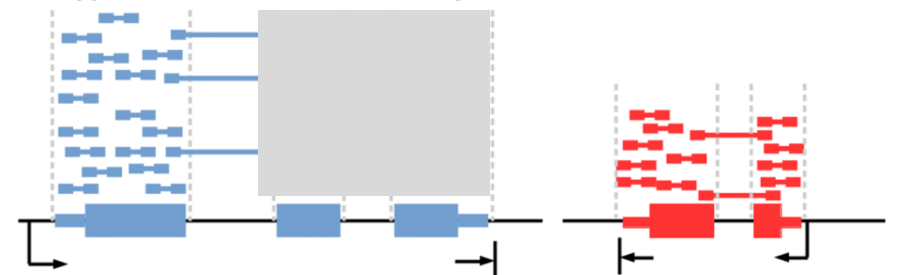
- More reliable quantification of genes on opposite strand
- Allow discovery of anti-sense transcription



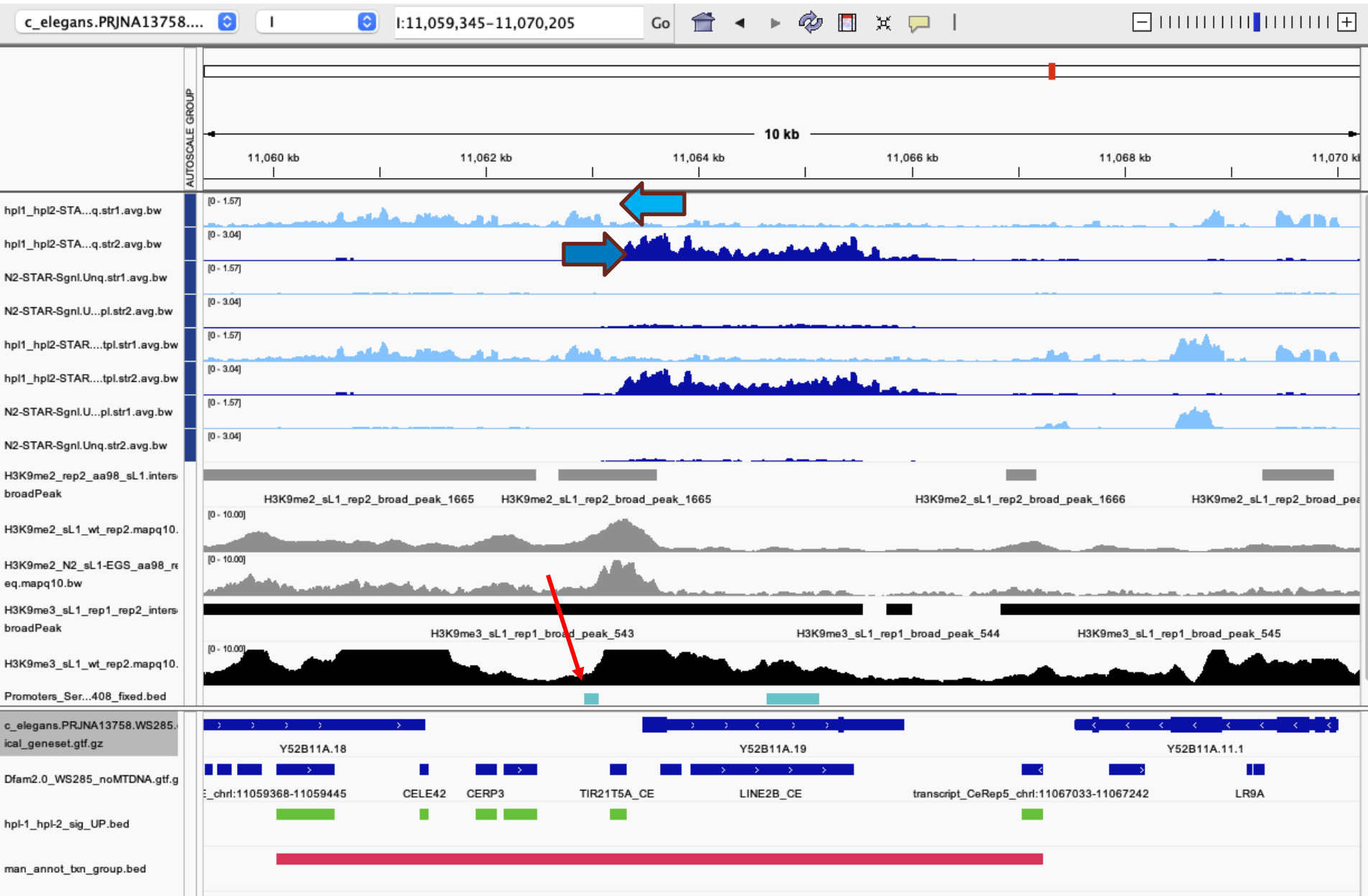
A. Mapped reads from an unstranded library



B. Mapped reads from a stranded library

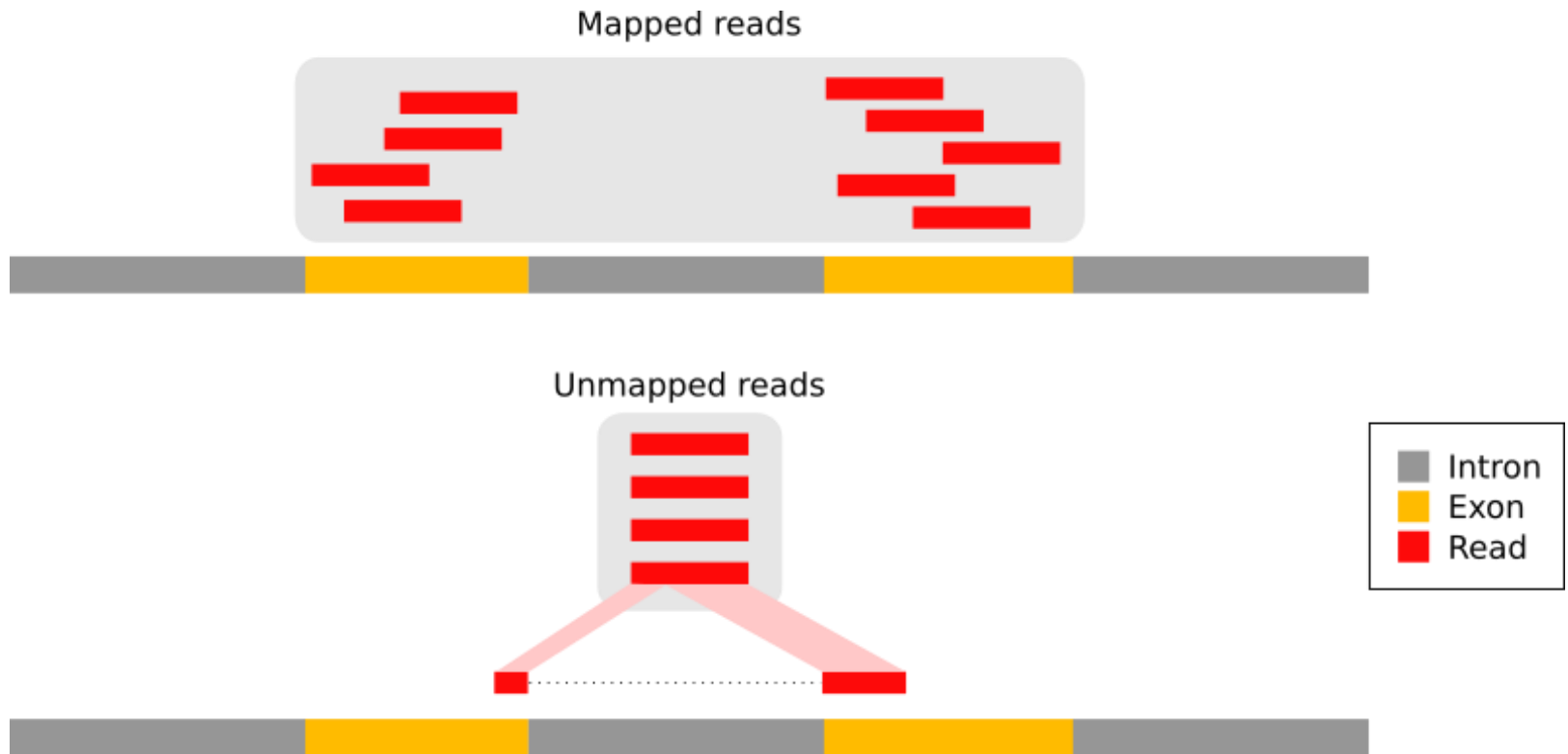


Real data example



Terms you might come across

splice-aware mapper



Terms you might come across

number of reads

strand-specific

single-end/pair-end

Single-end

Read fragment from only one end

Can be good enough for gene expression study, if there is a good reference genome

Pair-end

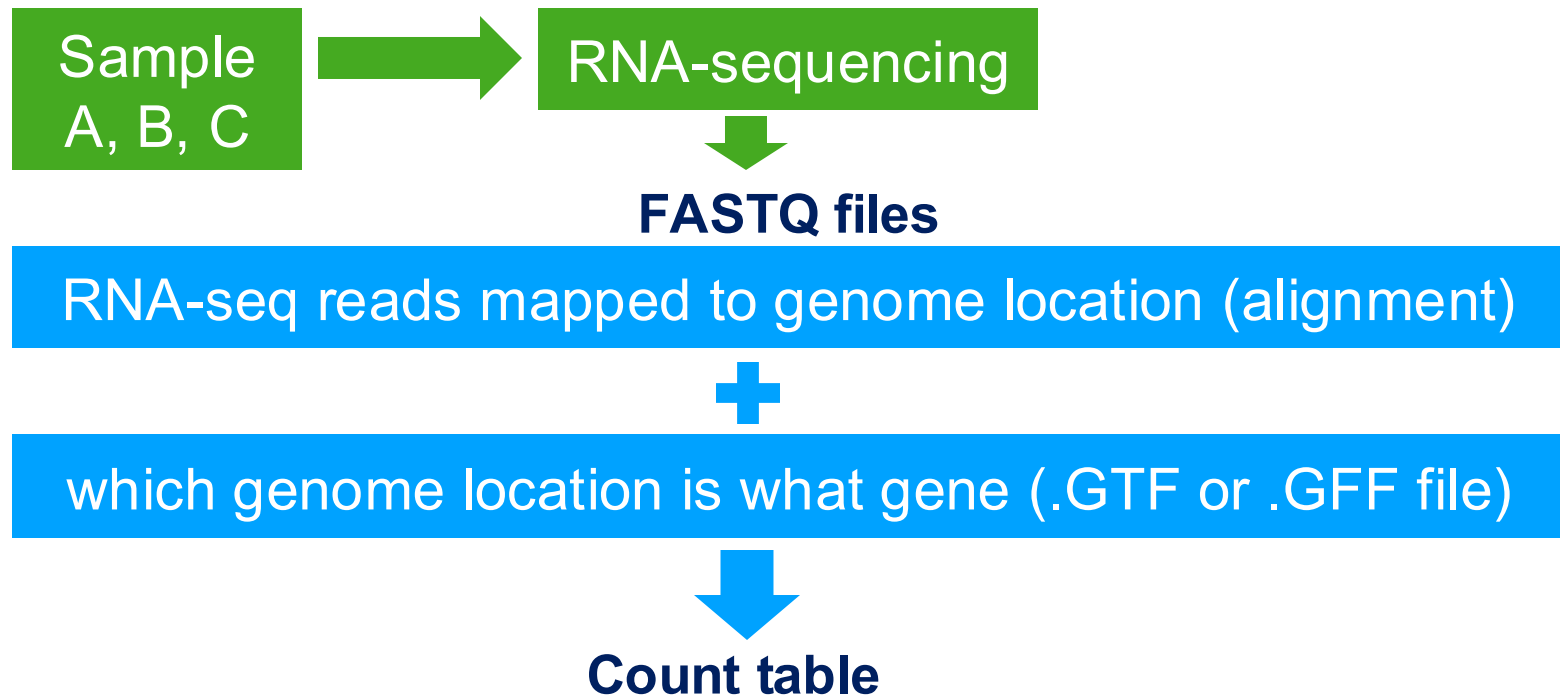
Read from both ends of the fragment

Provide more information which can help with mapping

Highly recommend for organism with only draft reference genome, or without a genome

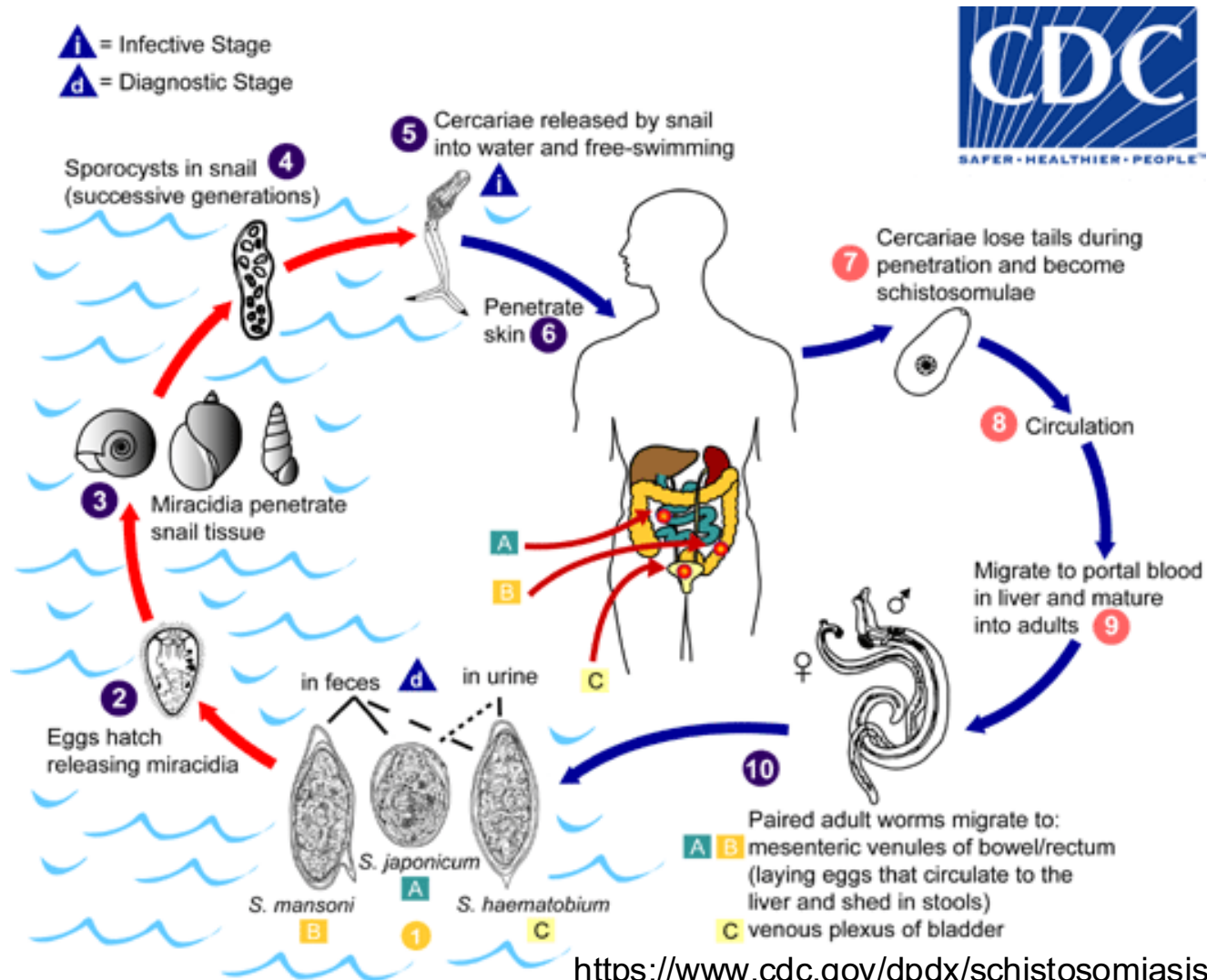


From sequencing data to read count



Gene	Count in sample A	Count in sample B	Count in sample C
gene1	4	8	20
gene2	6	3	16
gene3	5	5	15

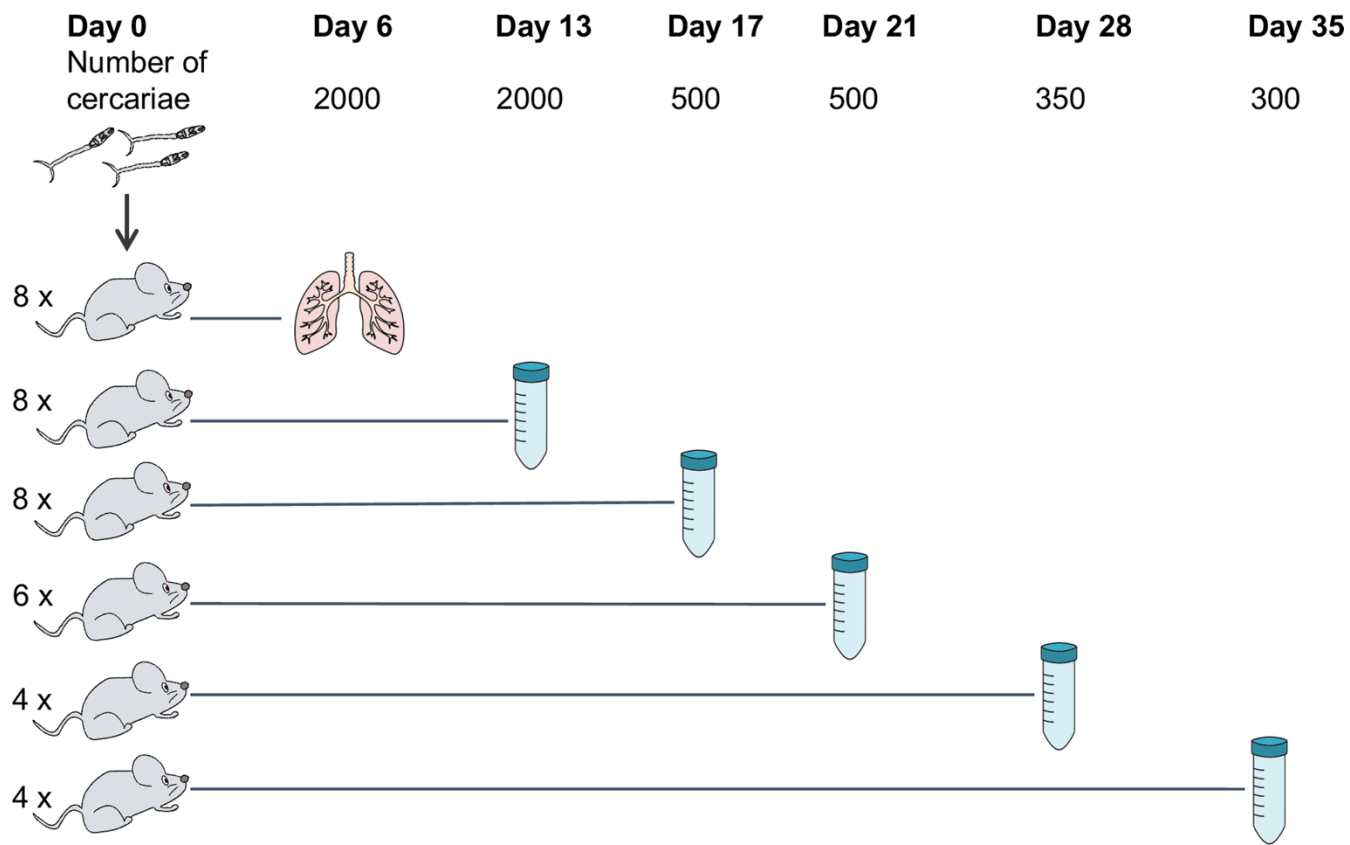
Schistosoma mansoni



Transcriptome of the parasitic flatworm *Schistosoma mansoni* during intra-mammalian development

Arporn Wangwiwatsin^{1,2}, Anna V. Protasio^{1,3}, Shona Wilson³, Christian Owusu¹, Nancy E. Holroyd¹, Mandy J. Sanders¹, Jacqueline Keane¹, Mike J. Doenhoff⁴, Gabriel Rinaldi¹, Matthew Berriman^{1*}

1 Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom, **2** Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand, **3** Department of Pathology, Tennis Court Road, University of Cambridge, Cambridge, United Kingdom, **4** School of Life Sciences, University of Nottingham, University Park, Nottingham, United Kingdom



Module practice data

FASTQ files (sequences from RNA-seq)

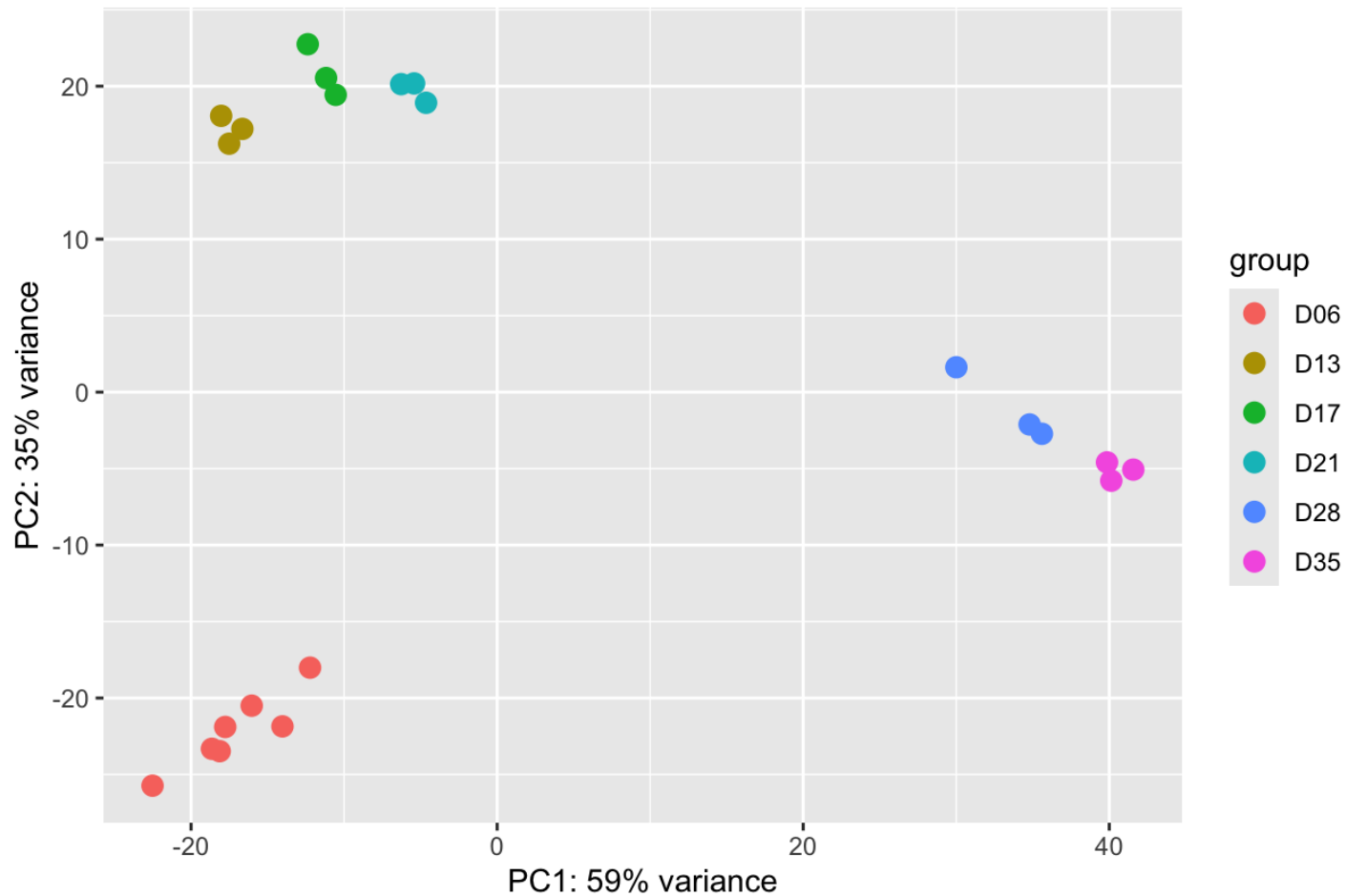
Transcriptomics/RNAseqData

ERR506076_sub_1.fastq

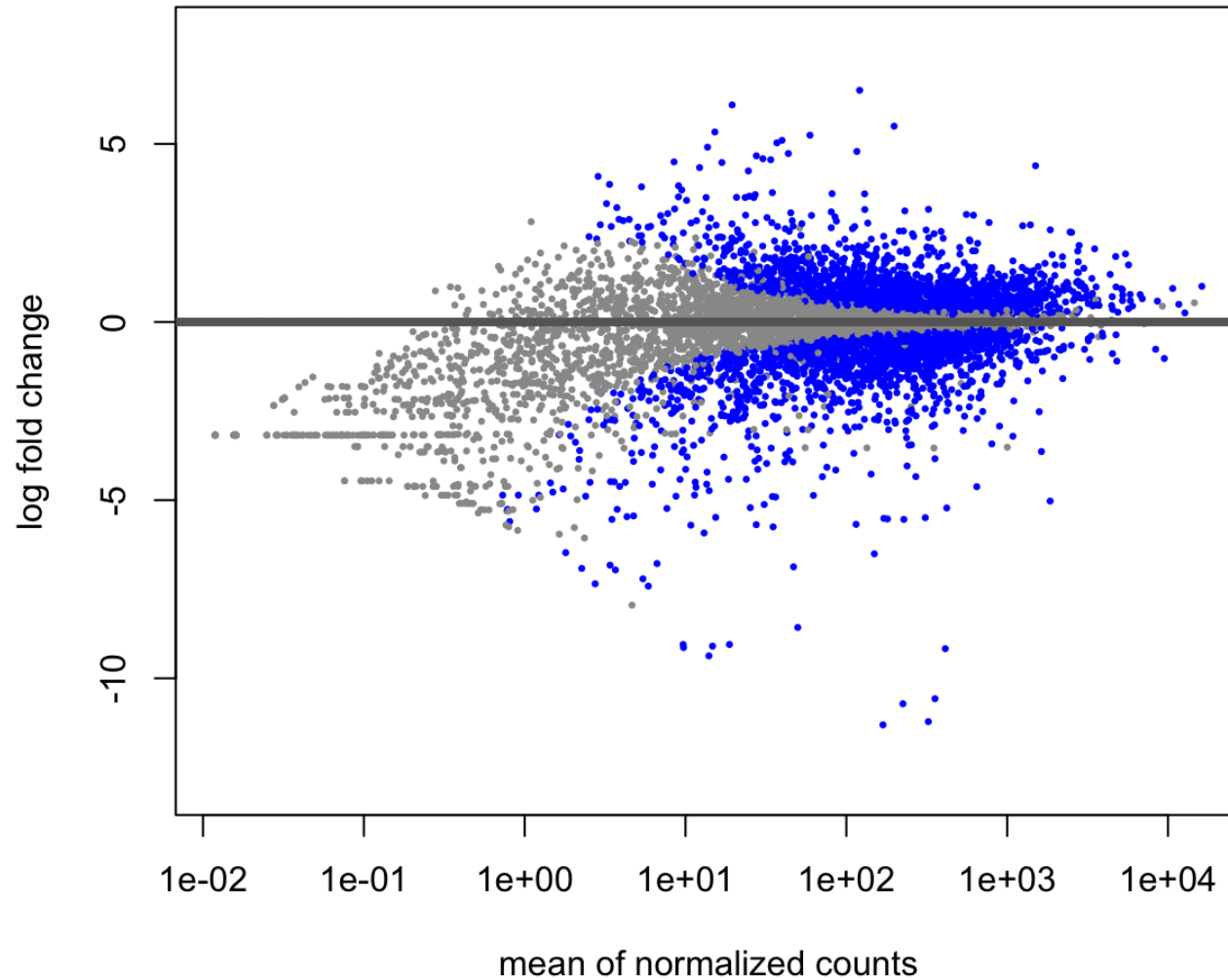
ERR506076_sub_2.fastq

SRR3223443_sub.fastq

Principal Component Analysis



MA plot



Tricky question ...

```
Res_1 <- results(dds, lfcThreshold = 0, alpha = 0.01)
```

*Is there **any** difference?*

*P-values are calculated assuming the null hypothesis is $LFC = 0$, so you're still testing for **any** difference in expression, no matter how small.*

```
Res_2 <- results(dds, lfcThreshold = 1, alpha=0.01)
```

Is the difference larger / smaller than one?

P-values are calculated assuming the null hypothesis is $LFC > 1$ or $LFC < 1$

<FULL_PATH>

replace this with

/home/manager/Transcriptomics/

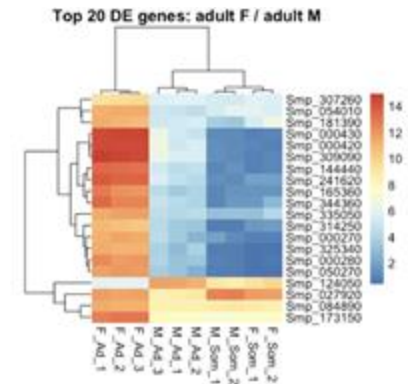
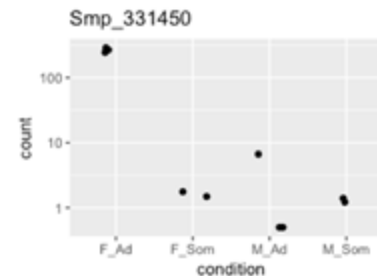
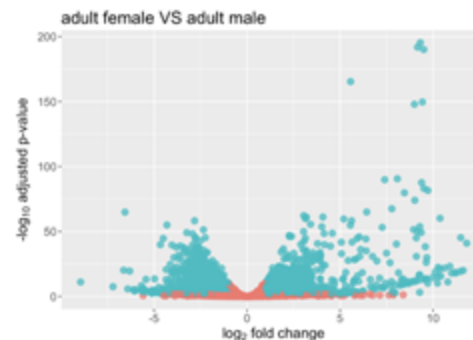
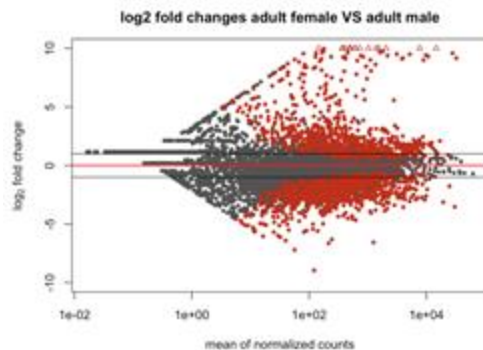
Data visualisation

Big result tables are hard/scary to look at

Sometimes we want to compare between datasets

Nice plots are your friends

- MA plots
- Volcano plots
- Gene expression plot
 - single gene
 - multiple genes (e.g. using heatmap)



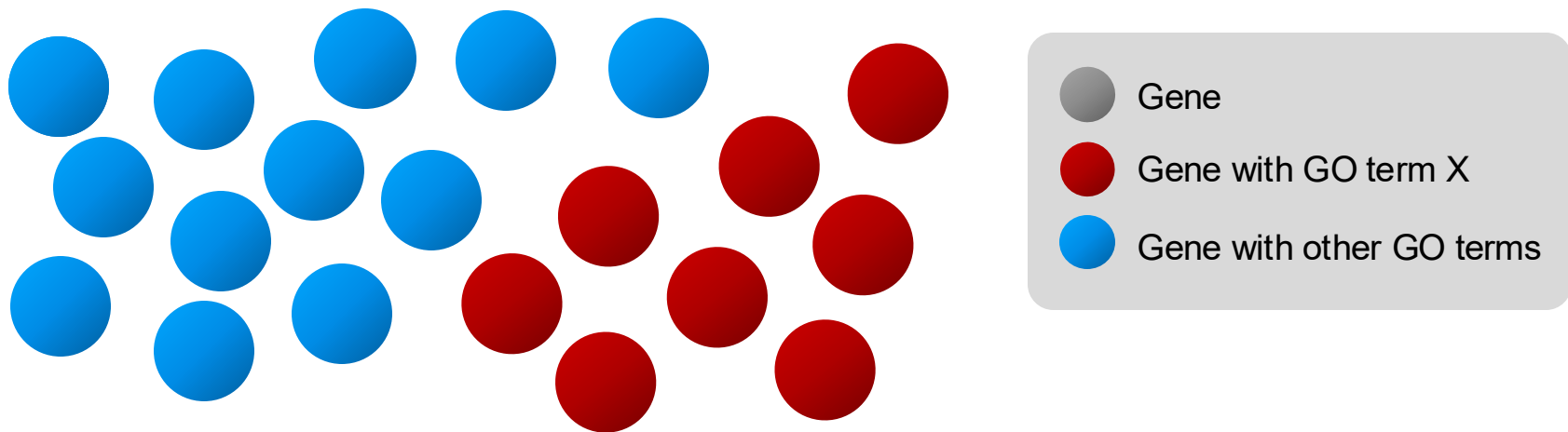
Functional analysis

- Rather than going through the list of differentially expressed genes to find genes that you expect to see changes
 - Do functional analysis
 - Let data guide the way
- Possibly the most common = GO enrichment

GO term enrichment

Genes often have associated GO terms (Gene Ontology terms). GO terms describe functions of a gene, and can be derived from sequence similarity, experiment, homology etc.

GO term enrichment: “Are there any GO terms present in my data more frequently than expected by chance alone?”



GO term enrichment

Genes often have associated GO terms (Gene Ontology terms).

WormBase ParaSite Version: WBPS9 (WS258)

Search WormBase ParaSite...

Genome List BLAST BioMart REST API VEP Downloads WormBase

Schistosoma mansoni (PRJEA36577) Location: Smp.Chr_3:12,709,526-12,722,895 Gene: Smp_013040

Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
- Sequence
- Literature
- Comparative genomics
 - Gene tree
 - Orthologues
 - Paralogues
- Gene Ontology**
 - Molecular function**
 - Cellular component
 - Biological process
- External references
- Expression
- Variation
 - Variation Table

Gene: Smp_013040

Description Cathepsin D (A01 family) [Source:UniProtKB/TrEMBL;Acc:[G4VEV6](#)]

Location [Scaffold Smp.Chr_3:12,709,526-12,722,895](#) reverse strand.

INSDC Sequence ID [HE601626.1](#)

Gene Overview This gene has 2 transcripts ([splice variants](#)), [1048 orthologues](#) and [1 paralogue](#).

Gene Type Protein coding

Annotation Method Gene models from Wellcome Trust Sanger Institute [Reference Helminth Genomes project](#)

Transcripts [Show transcript table](#)

Molecular function

GO term	Evidence	Annotation source	Transcript IDs	Actions
Aspartic-type endopeptidase activity	IEA	UniProtKB/TrEMBL:G4VEV6 , UniProtKB/TrEMBL:P91802 , InterPro:Aspartic_peptidase_AS , InterPro:Aspartic_peptidase_A1 , InterPro:Cathepsin_D , InterPro:Aspartic_peptidase_N	Smp_013040.1 Smp_013040.2	Search BioMart View associated genes
Aspartic-type endopeptidase activity	IEA	UniProtKB/TrEMBL:G4VEV6 , UniProtKB/TrEMBL:P91802	Smp_013040.1 Smp_013040.2	Search BioMart View associated genes

GO term enrichment

Genes often have associated GO terms (Gene Ontology terms). GO terms describe functions of a gene, and can be derived from sequence similarity, experiment, homology etc.

ID number **Description**

WormBase ParaSite Version: W

Genome List BLAST BioMart REST API VEP

Schistosoma mansoni (PRJEA36577) Location: Smp.Chr_3:12

Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
- Sequence
- Literature
- Comparative genomics
 - Gene tree
 - Orthologues
 - Paralogues
- Gene Ontology
 - Molecular function**
 - Cellular component
 - Biological process
- External references
- Expression
- Variation
 - Variation Table
 - Variation Image

Gene: Smp_013040

Description

Location

INSDC Sequence ID

Gene Overview

Gene Type

Annotation Method

Transcripts

Molecular function

Accession	Term
GO:0004190	aspartic-type endopeptidase activity
GO:0008233	peptidase activity

Table with 6 columns: Accession, Term, Evidence, Annotation source, Transcript IDs, and links.

Accession	Term	Evidence	Annotation source	Transcript IDs	Links
GO:0004190	aspartic-type endopeptidase activity	IEA	UniProtKB/TrEMBL:G4VEV6, UniProtKB/TrEMBL:P91802, InterPro:Aspartic_peptidase_AS, InterPro:Aspartic_peptidase_A1, InterPro:Cathepsin_D, InterPro:Aspartic_peptidase_N	Smp_013040.1 Smp_013040.2	Search BioMart View associated genes
GO:0008233	peptidase activity	IEA	UniProtKB/TrEMBL:G4VEV6, UniProtKB/TrEMBL:P91802	Smp_013040.1 Smp_013040.2	Search BioMart View associated genes

Making use of existing transcriptome data

Databases of transcriptomics data

- Gene expression omnibus
(<https://www.ncbi.nlm.nih.gov/geo/>)
- ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>)
- Expression Atlas (<https://www.ebi.ac.uk/gxa/home>)

Example of RNA-seq data reuse:

- This paper looked at data from published RNA-seq experiments and revealed that some of the cells were contaminated with Mycoplasma (**PMID: 25712092**).
- This paper used existing RNA-seq data with some new data to study specific gene expression on Z and W chromosome of schistosomes (**PMID: 30044216**).

Finding data on the web

ENA - European Nucleotide Archive

<https://www.ebi.ac.uk/ena>

The screenshot shows the ENA website with a teal header. The header includes the EMBL-EBI logo, the ENA logo, and a search bar with a 'Search' button. Below the header is a navigation bar with links: Home, Search & Browse, Submit & Update, Software, About ENA, and Support. The main content area is divided into two columns. The left column contains the 'European Nucleotide Archive' title, a description of the archive, and two search sections: 'Text Search' and 'Sequence Search'. The right column contains a 'Popular' section with a list of links and a 'Latest ENA news' section with a news item dated 08 Jul 2019.

EMBL-EBI

Services Research Training About us

ENA
European Nucleotide Archive

Search

Examples: BN000065, histone

Advanced
Sequence

Home Search & Browse Submit & Update Software About ENA Support

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

Text Search

search

Examples: BN000065, histone

Advanced search

Sequence Search

Enter or paste a nucleotide sequence or accession number

Search

Advanced search

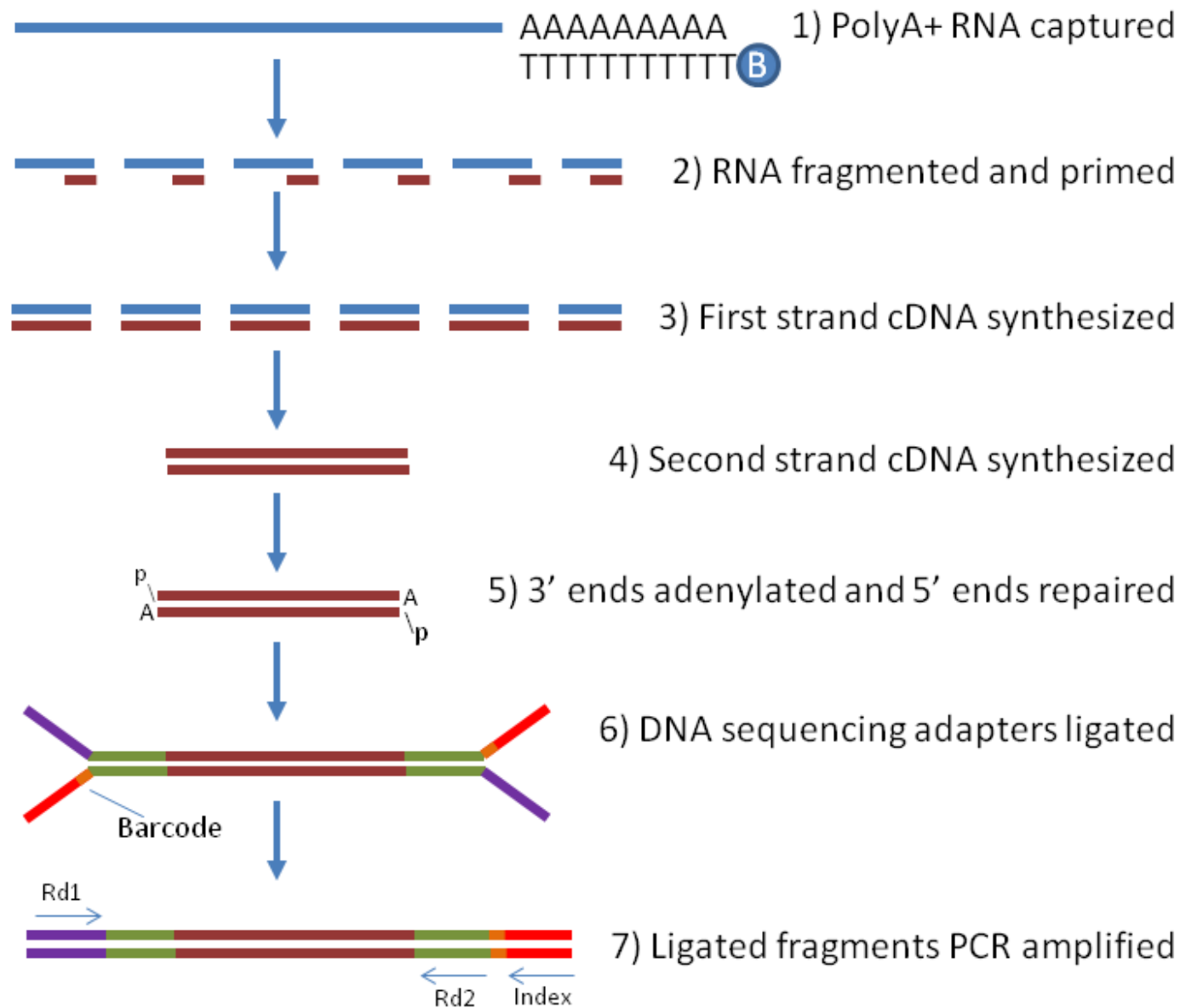
Popular

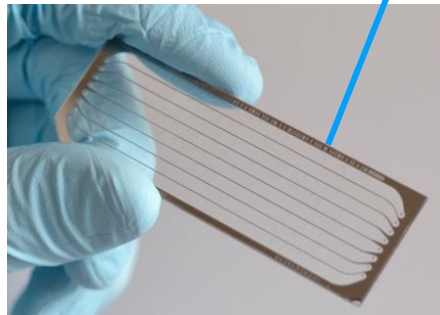
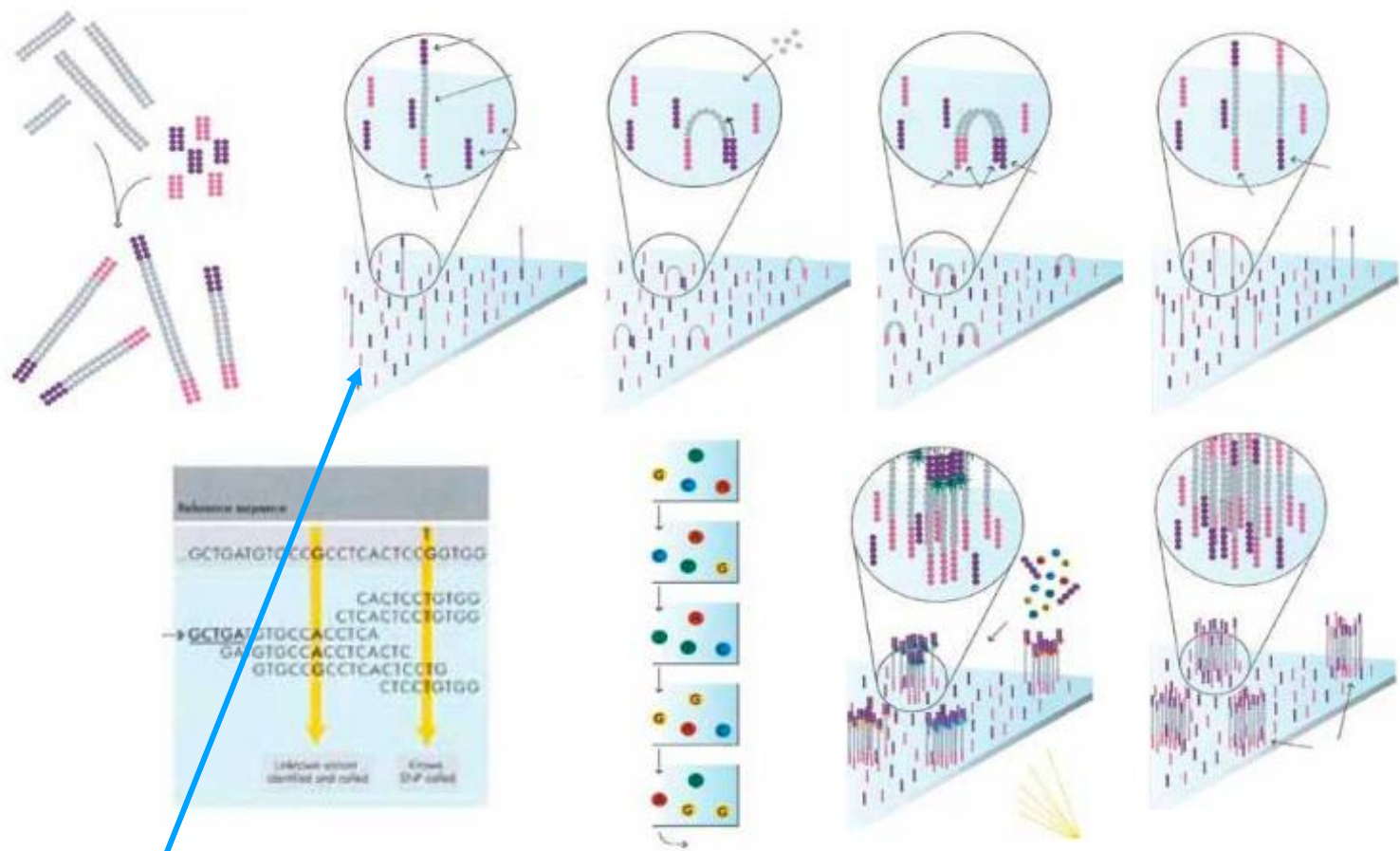
- Submit and update
- Sequence submissions
- Genome assembly submissions
- Submitting environmental sequences
- Citing ENA data
- REST URLs for data retrieval
- REST URLs to search ENA

Latest ENA news

08 Jul 2019: Release 140 of ENA's assembled/annotated sequences is now available

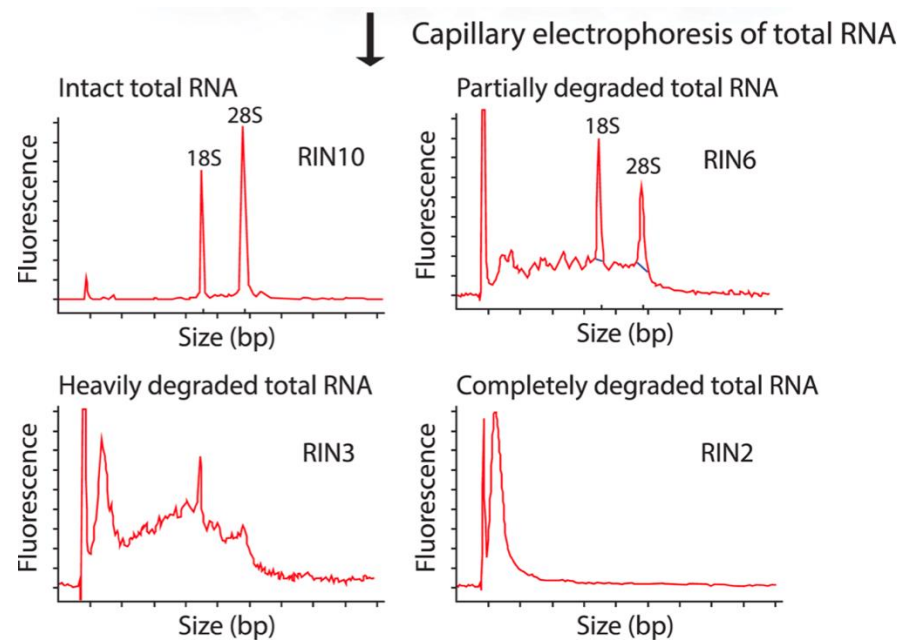
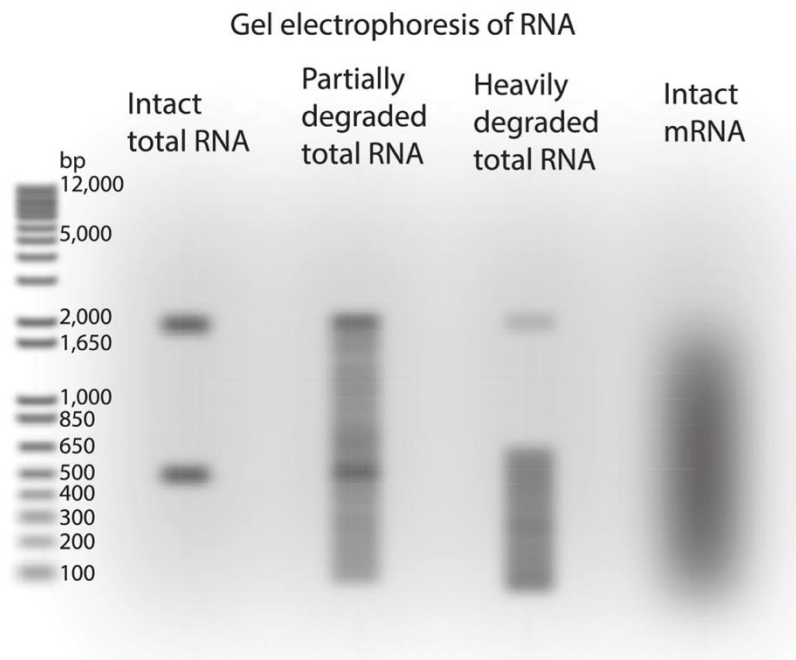
Release 140 of assembled/annotated sequences from ENA is now...





QC your RNA

- Sharpness and intensity of bands on RNA gel and RIN number are indicator of RNA integrity
- Note: depending on your organism, RIN is not always a good indicator for RNA integrity



Extra: Functional analysis pathway enrichment

Title: Interferon type I signaling pathway
Availability: CC BY 2.0
Organism: Homo sapiens

