Vall d'Hebron
Institut de Recerca

INTRODUCTION TO OMICS DATA ANALYSIS RNA-seq

Bioinformatics Course UEB-VHIR

November 2023

Esther Camacho[1], Mireia Ferrer[1], Álex Sánchez[1,2], Berta Miró[1]

[1] Unitat d'Estadística i Bioinformàtica (UEB) VHIR [2] Departament de Genètica Microbiologia i Estadística, UB

Vall d'Hebron Hospital · Generalitat de Catalunya Departament de Salut · Institut Català de la Salut · UAB Universitat Autònoma de Barcelona · Generalitat de Catalunya Departament d'Economia i Coneixement · Institució CERCA Centres de Recerca de Catalunya · Instituto de Salud Carlos III

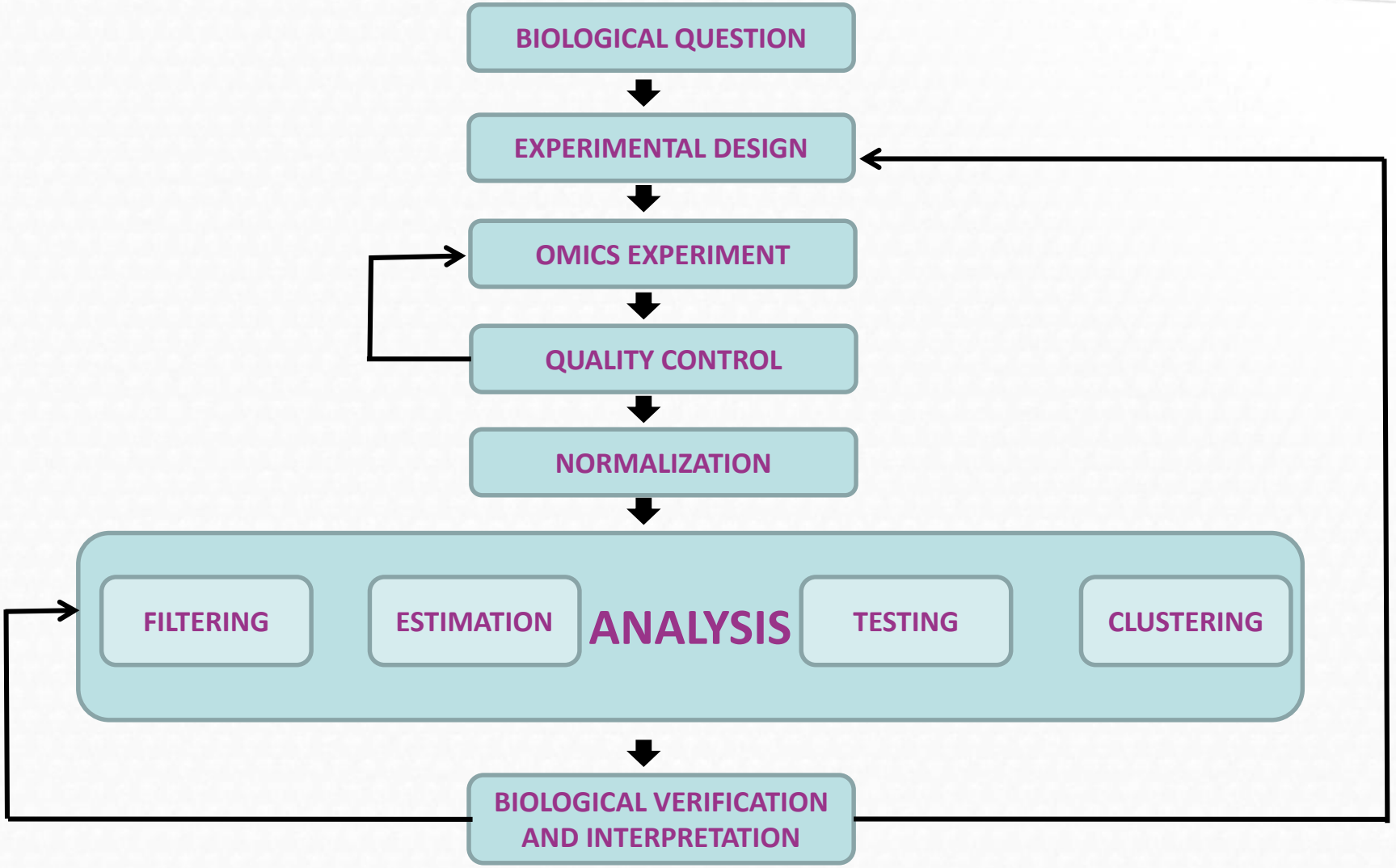
Vall d'Hebron
Institut de Recerca

1. **Introduction to omics data analysis**
2. **An example of omics data analysis. RNA-seq**
    1. **What is RNA-seq**
    2. **Basic key concepts**
    3. **Main challenges in RNA-seq**
    4. **RNA-seq vs Microarrays**
    5. **RNA-seq analysis pipeline(s)**
    6. **Alignment**
    7. **Transcript assembly**
    8. **DEG Analysis**
    9. **Visualization**

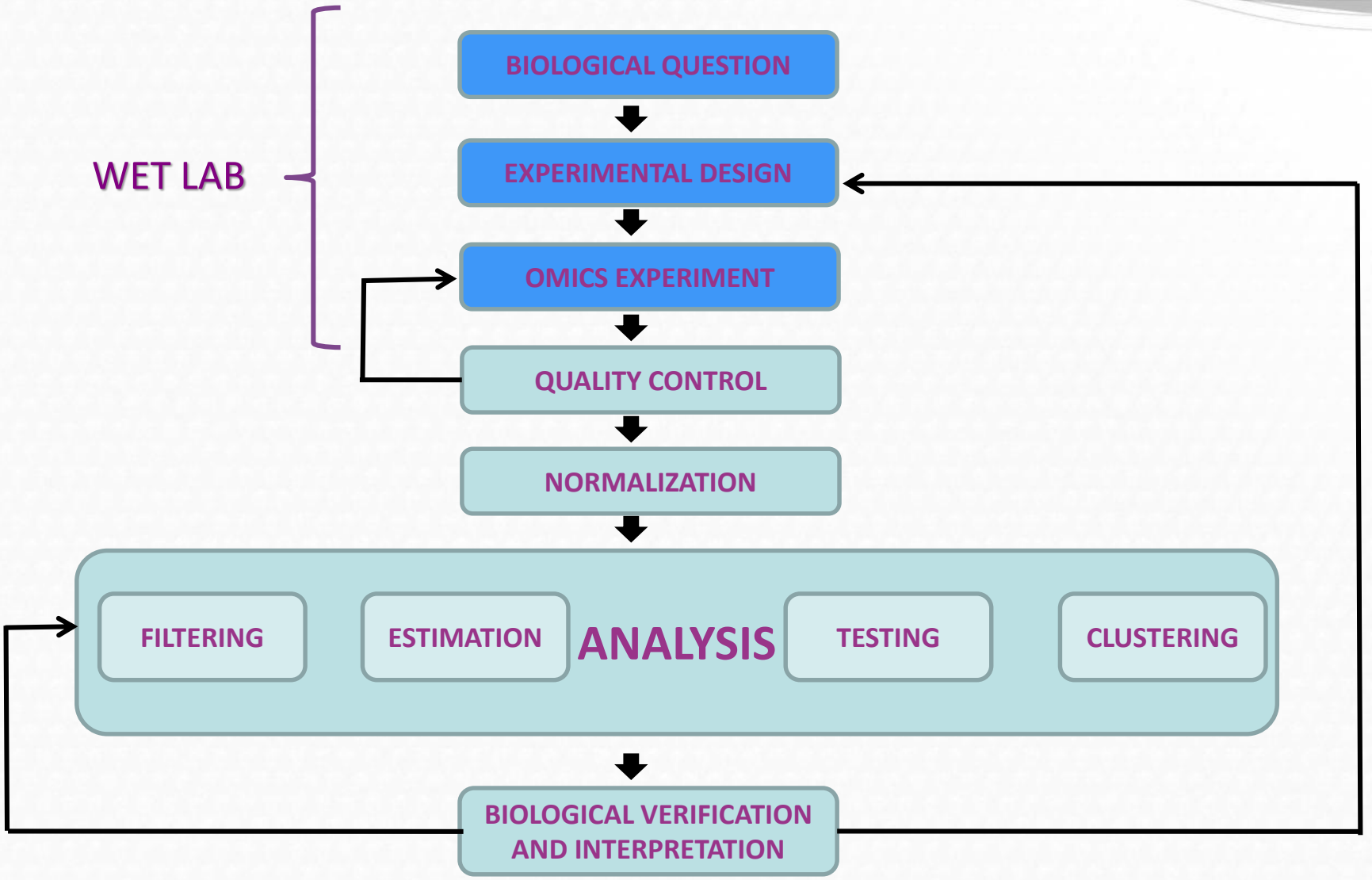
Vall d'Hebron
Institut de Recerca

1. **Introduction to omics data analysis**
2. **An example of omics data analysis. RNA-seq**
    1. **What is RNA-seq**
    2. **Basic key concepts**
    3. **Main challenges in RNA-seq**
    4. **RNA-seq vs Microarrays**
    5. **RNA-seq analysis pipeline(s)**
    6. **Alignment**
    7. **Transcript assembly**
    8. **DEG Analysis**
    9. **Visualization**

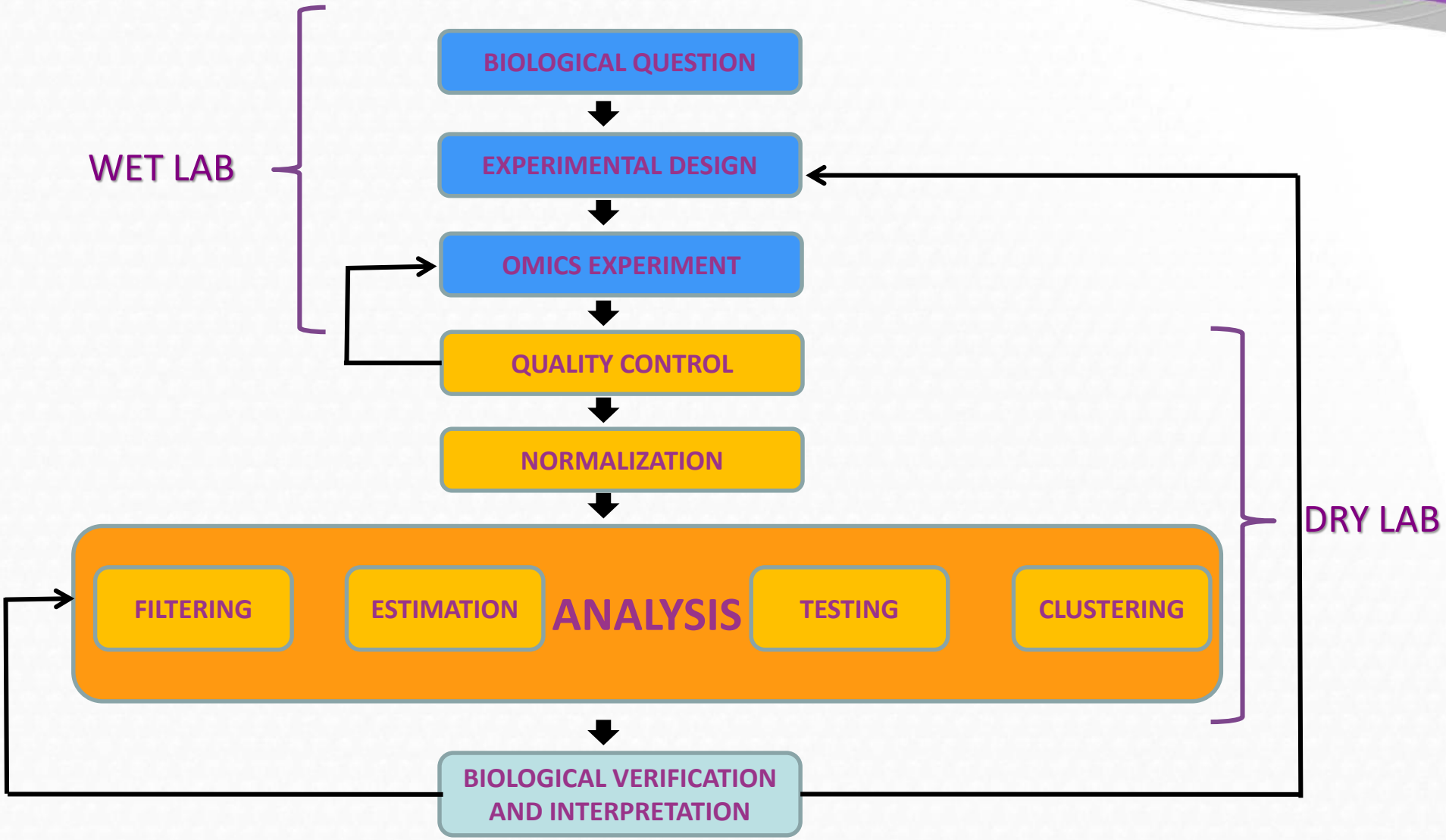
1. Introduction to omic data analysis



1. Introduction to omic data analysis



1. Introduction to omic data analysis



Vall d'Hebron
Institut de Recerca

1. **Introduction to omics data analysis**
2. **An example of omics data analysis. RNA-seq**
    1. **What is RNA-seq**
    2. **Basic key concepts**
    3. **Main challenges in RNA-seq**
    4. **RNA-seq vs Microarrays**
    5. **RNA-seq analysis pipeline(s)**
    6. **Alignment**
    7. **Transcript assembly**
    8. **DEG Analysis**
    9. **Visualization**

2.1 What is RNA-seq?

Vall d'Hebron
Institut de Recerca

RNA-Seq

- Sequencing technique of **NGS**.
- It reveals the **presence and quantity of RNA** in a sample.
- It lets to the determination/analysis of the **transcriptome**.
- This sample is sequenced in a **particular moment**, so the transcriptome obtained is limited to this **precise moment**.

2.1 What is RNA-seq?

Vall d'Hebron
Institut de Recerca

RNA-Seq

- Sequencing technique of **NGS**.
- It reveals the **presence and quantity of RNA** in a sample.
- It lets to the determination/analysis of the **transcriptome**.
- This sample is sequenced in a **particular moment**, so the transcriptome obtained is limited to this **precise moment**.

RNA-seq enables the finding of:

- Alternative gene spliced transcripts,
- Post-transcriptional modifications,
- Gene fusion,
- Mutations/SNPs,
- Changes in gene expression

2.1 What is RNA-seq?

- Functional studies
  - **Genome may be constant** but an **experimental condition has a pronounced effect on gene expression**
    - ✓ e.g. Drug treated vs. untreated cell line
    - ✓ e.g. Wild type versus knock out mice

- Some **molecular features** can only be observed at the RNA level
  - Alternative isoforms, fusion transcripts, RNA editing

- Predicting transcript sequence from genome sequence is **difficult**
  - Alternative splicing, RNA editing, etc.

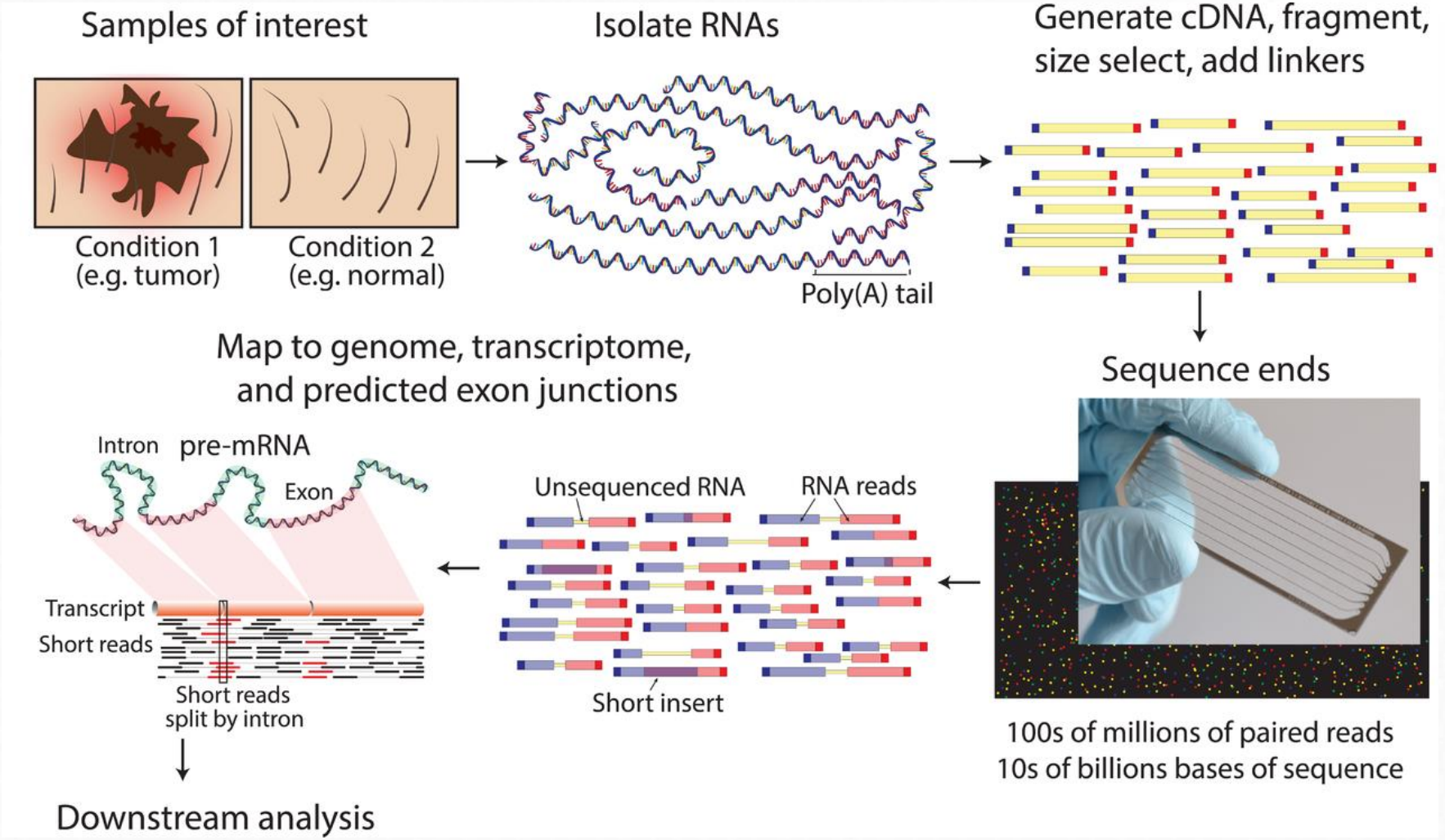
2.1 What is RNA-seq?

- RNA-seq is the high throughput sequencing of **cDNA** using NGS technologies
- RNA-seq works by sequencing every RNA molecule and profiling the expression of a particular gene by counting the number of time its transcripts have been sequenced.
- The summarized RNA-seq data is widely known as ***count table***

	Condition A			Condition B		
Gene1	4	0	2	12	14	13
Gene2	0	23	50	47	22	0
Gene3	0	2	6	13	11	15
...	...	...	...	...	...	...
GeneG	156	238	37	129	51	118

2.1 What is RNA-seq?

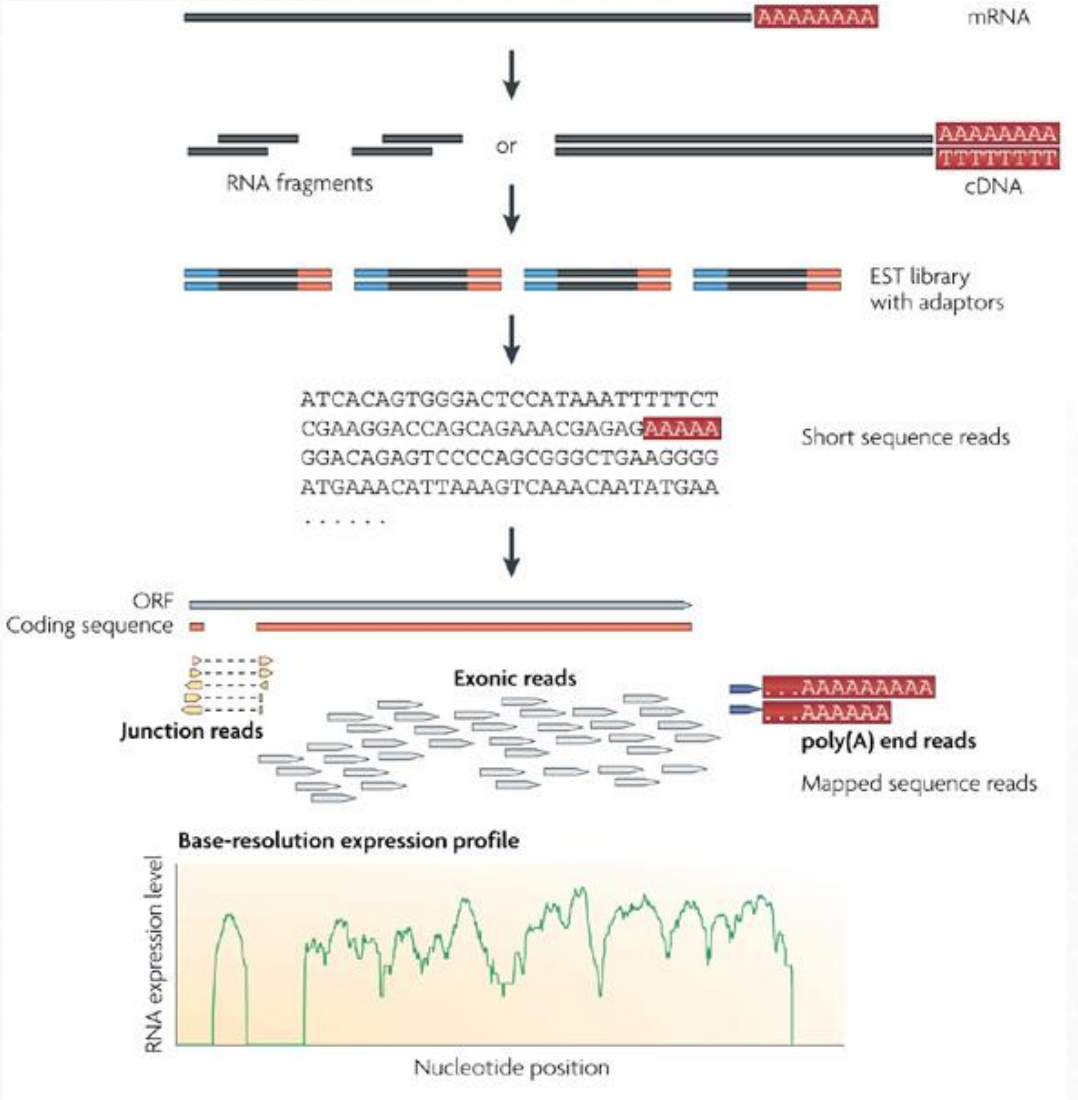
RNA-seq (SGS)



https://en.wikipedia.org/wiki/RNA-Seq

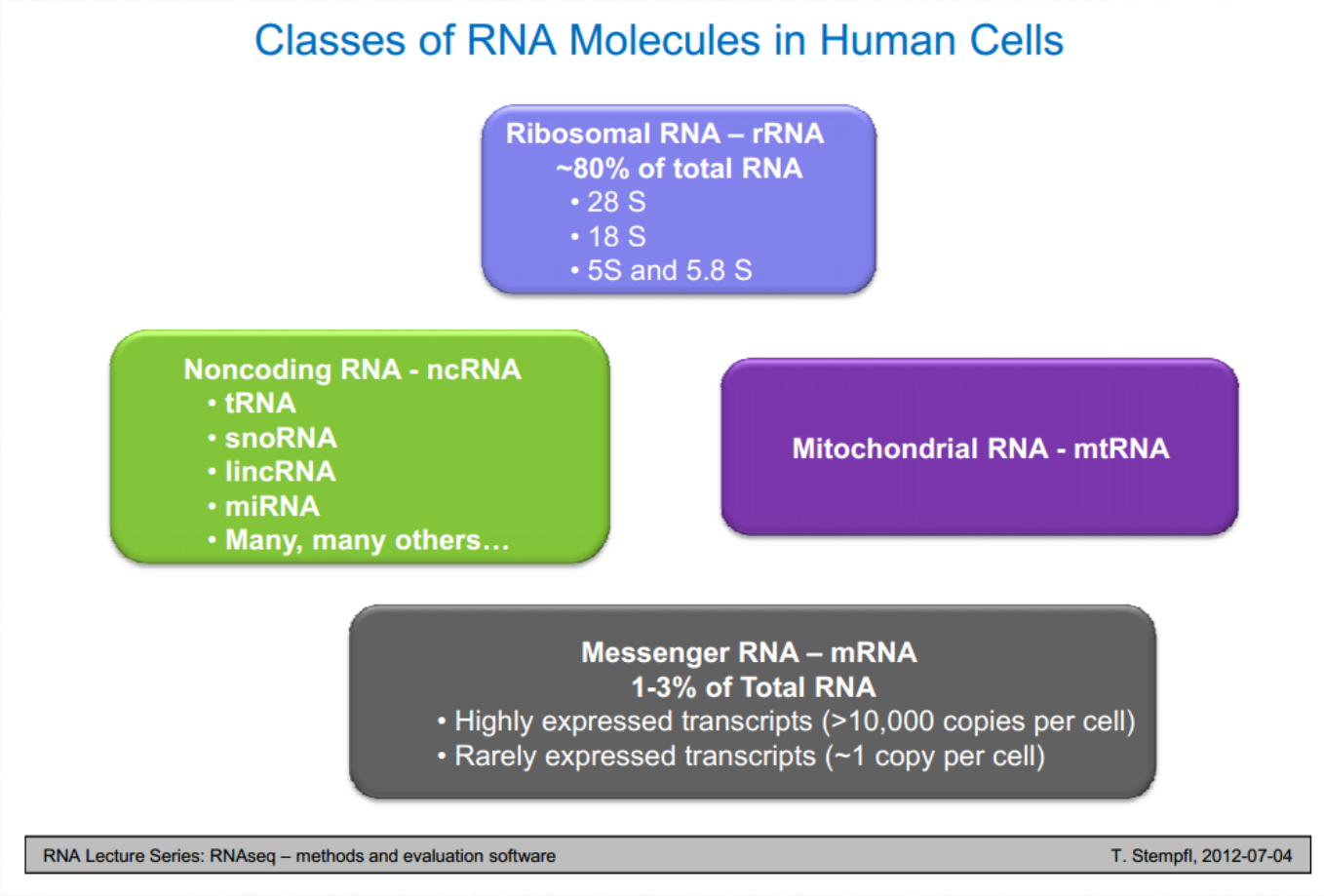
2.1 What is RNA-seq?

- Long RNAs are first converted into a library of **cDNA** fragments through either RNA fragmentation or DNA fragmentation.
- Sequencing **adaptors (blue) are added** to each cDNA fragment and a short sequence is obtained from each cDNA using high-throughput sequencing technology.
- The resulting sequence reads are **aligned** with the reference genome or transcriptome, and classified as three types: exonic reads, junction reads and poly(A) end-reads. These three types are used to generate **a base-resolution expression profile** for each gene, as illustrated at the bottom; a yeast ORF with one intron is shown.

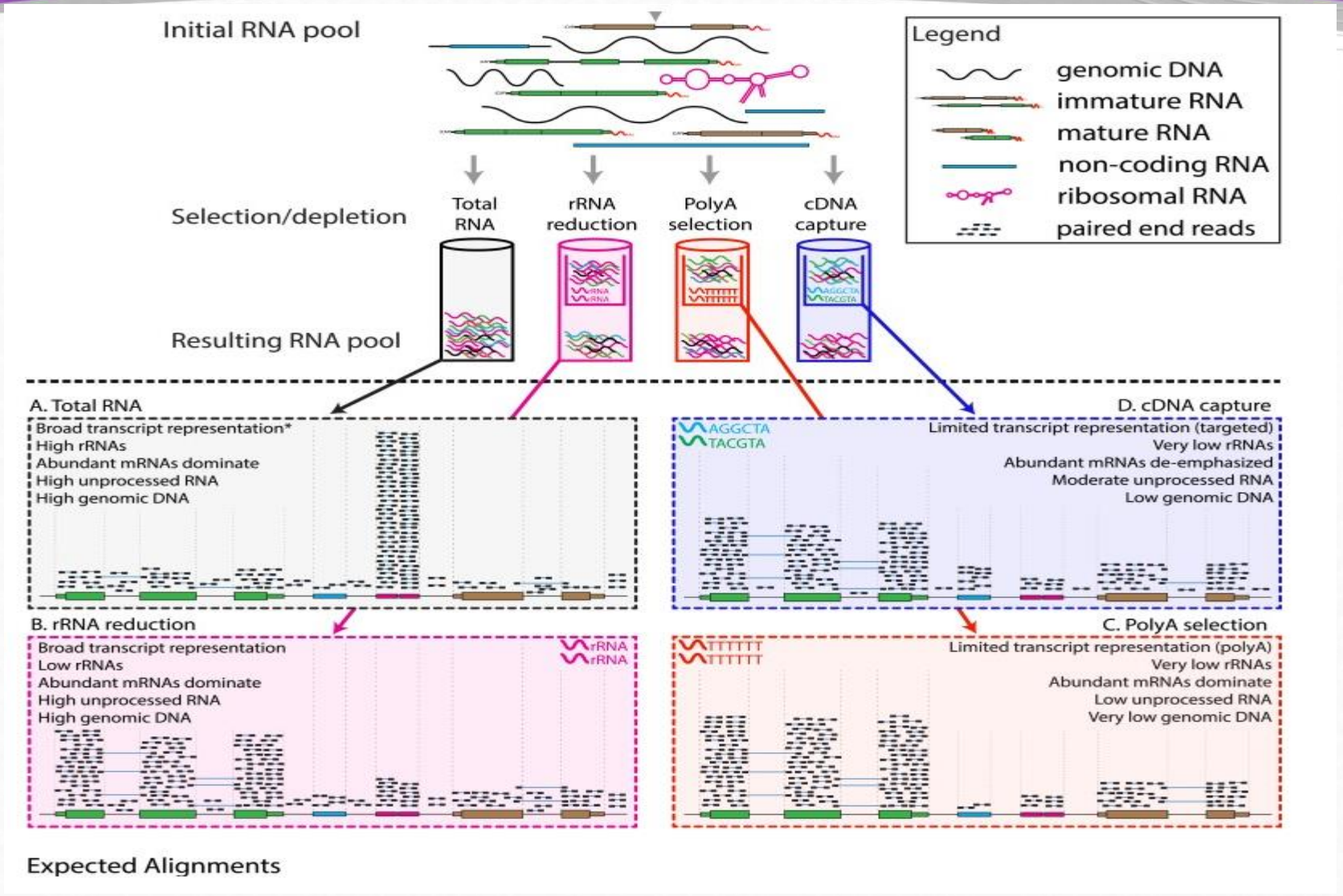


Nature Reviews Genetics **10**, 57-63 (January 2009)

2.1 What is RNA-seq?



3.3 Main challenges in RNA-seq



Vall d'Hebron
Institut de Recerca

1. **Introduction to omics data analysis**
2. **An example of omics data analysis. RNA-seq**
    1. What is RNA-seq
    2. Basic key concepts
    3. Main challenges in RNA-seq
    4. RNA-seq vs Microarrays
    5. RNA-seq analysis pipeline(s)
    6. Alignment
    7. Transcript assembly
    8. DEG Analysis
    9. Vizualization

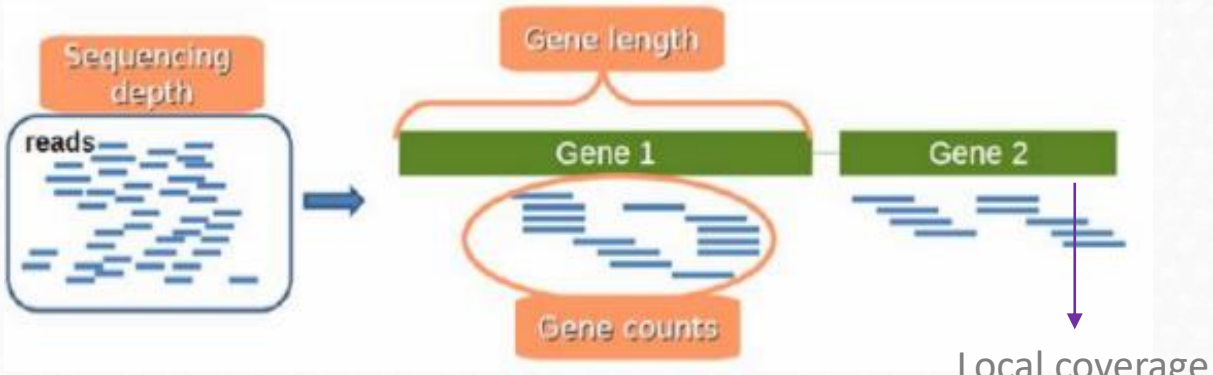
2.2 Basic key concepts?

Sequencing depth: Total number of reads mapped to the genome. (Library size) Could also be applied to samples.

Coverage: Number of reads mapped to a specific region (average of them if we are talking about the whole genome...)

Gene length: Number of bases that a gene has.

Gene counts: Number of reads mapping to that gene (expression measurement)



Vall d'Hebron
Institut de Recerca

1. **Introduction to omics data analysis**
2. **An example of omics data analysis. RNA-seq**
    1. **What is RNA-seq**
    2. **Basic key concepts**
    3. **Main challenges in RNA-seq**
    4. **RNA-seq vs Microarrays**
    5. **RNA-seq analysis pipeline(s)**
    6. **Alignment**
    7. **Transcript assembly**
    8. **DEG Analysis**
    9. **Vizualization**

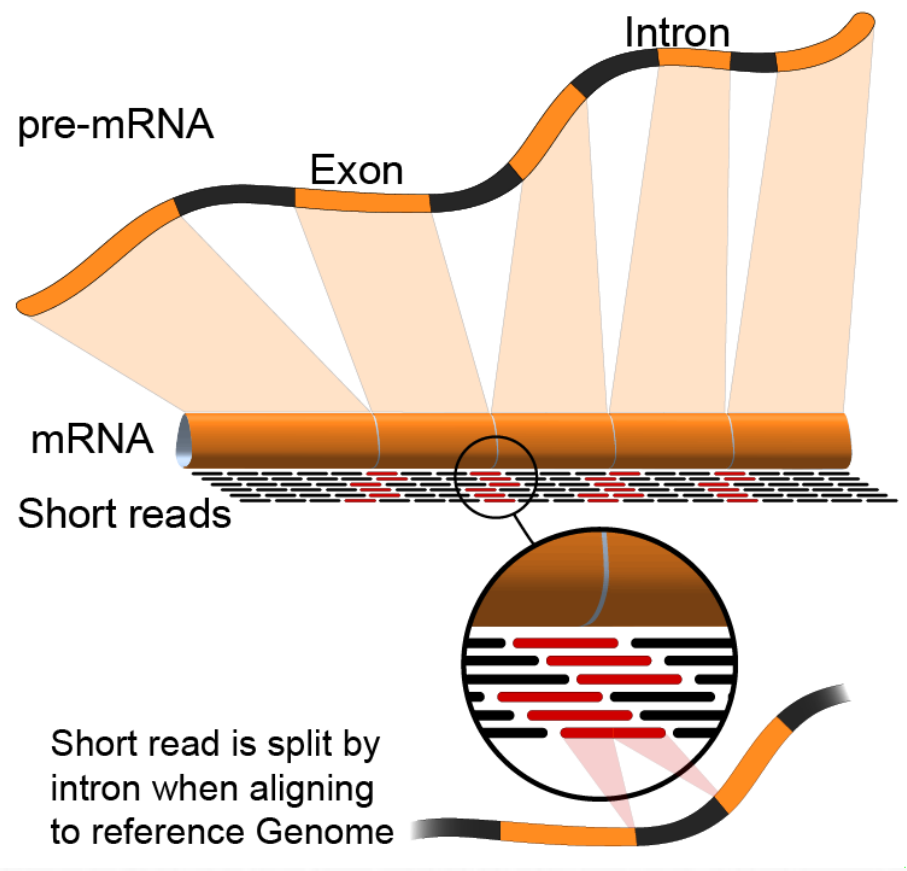
3.3 Main challenges in RNA-seq

- **Sample**
  - Purity? Quantity? Quality?

- **RNAs consist of small exons that may be separated by large introns**
  - Mapping reads to genome is challenging
  - Non-uniformity coverage of the genome

3.3 Main challenges in RNA-seq

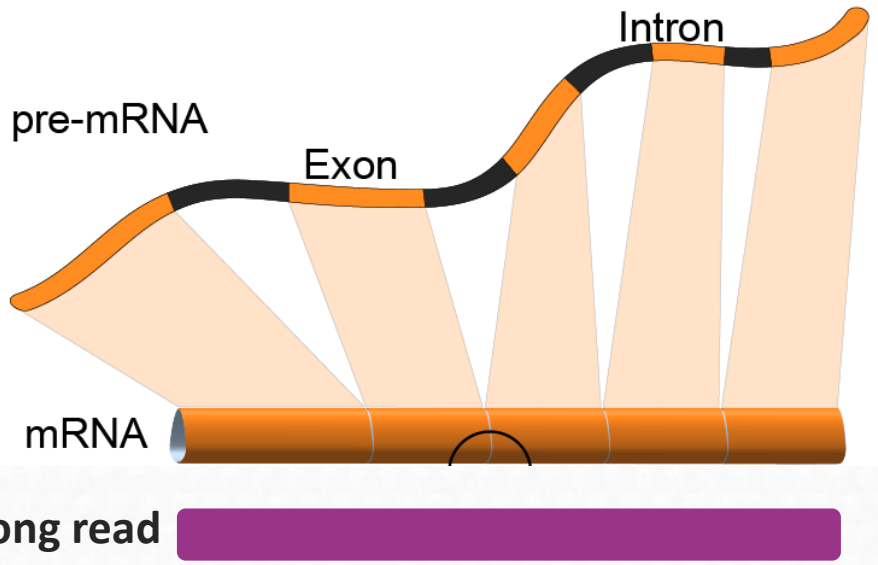
- **Sample**
  - Purity? Qua
- **RNAs consist of** arge introns
  - Mapping rea
  - Non-uniform



RNA-Seq
second generation
sequencing

3.3 Main challenges in RNA-seq

- **Sample**
  - Purity? Qua
- **RNAs consist of** arge introns
  - Mapping rea
  - Non-uniform



Molecule = Read = Transcript

RNA-Seq
third generation
sequencing

3.3 Main challenges in RNA-seq

- **Sample**
  - Purity? Quantity? Quality?

- **RNAs consist of small exons that may be separated by large introns**
  - Mapping reads to genome is challenging
  - Non-uniformity coverage of the genome

- **The relative abundance of RNAs vary wildly**
  - 10^5 – 10^7 orders of magnitude
  - Since RNA sequencing works by random sampling, a small fraction of highly expressed genes may consume the majority of reads (rRNA)

- **RNAs come in a wide range of sizes**
  - Small RNAs must be captured separately
  - PolyA selection of large RNAs may result in 3' end bias

- **RNA is fragile compared to DNA** (easily degraded)

- **Library preparation**

3.3 Main challenges in RNA-seq

- **Capture of less abundant RNAs**
- **Avoid the conversion to cDNA prior to library construction**

- **RNA-Seq TGS** → **Increasing the yield**
  → **Avoid degradation at 5' end**

3.3 Main challenges in RNA-seq

- Independently of the software used, one needs to think about

DATA STORAGE & MANAGEMENT!!

1 Illumina Flow Cell equals up to
1.5 Bn individual Clusters
= 3 Bn Reads
= 300 Gbases raw sequence
= 2.5 TByte of disk space (raw data)
> 100 GByte of disk space (fastq data)

T. Stempfl, 2012

Vall d'Hebron
Institut de Recerca

1. **Introduction to omics data analysis**
2. **An example of omics data analysis. RNA-seq**
    1. What is RNA-seq
    2. Basic key concepts
    3. Main challenges in RNA-seq
    4. RNA-seq vs Microarrays
    5. RNA-seq analysis pipeline(s)
    6. Alignment
    7. Transcript assembly
    8. DEG Analysis
    9. Vizualization

2.4 RNA-seq vs Microarrays

Published online 15 October 2008 | *Nature* **455**, 847 (2008) |
doi:10.1038/455847a

News

The death of microarrays?

High-throughput gene sequencing seems to be stealing a march on microarrays. Heidi Ledford looks at a genome technology facing intense competition.

Announcing the death of the Micro-array

30 August 2010 by Anthony Fejes, posted in Uncategorized

| **Anthony Fejes**
| About the blog
| Blog homepage

- reproducibility
- only show you what you're looking for
- what about 'indels', inversions, translocations...
- accuracy
- sensitivity

2.4 RNA-seq vs Microarrays

PLoS One. 2013 Aug 20;8(8):e71462. doi: 10.1371/journal.pone.0071462. eCollection 2013.

Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data.

Guo Y[1], Sheng Q, Li J, Ye F, Samuels DC, Shyr Y.

Abstract

RNAseq and microarray methods are frequently used to measure gene expression level. While similar in purpose, there are fundamental differences between the two technologies. Here, we present the largest comparative study between microarray and RNAseq methods to date using The Cancer Genome Atlas (TCGA) data. **We found high correlations between expression data obtained from the Affymetrix one-channel microarray and RNAseq (Spearman correlations coefficients of ~0.8).** We also observed that the low abundance genes had poorer correlations between microarray and RNAseq data than high abundance genes. As expected, due to measurement and normalization differences, Agilent two-channel microarray and RNAseq data were poorly correlated (Spearman correlations coefficients of only ~0.2). By examining the differentially expressed genes between tumor and normal samples we observed reasonable concordance in directionality between Agilent two-channel microarray and RNAseq data, although a small group of genes were found to have expression changes reported in opposite directions using these two technologies. Overall, RNAseq produces comparable results to microarray technologies in term of expression profiling. The RNAseq normalization methods RPKM and RSEM produce similar results on the gene level and reasonably concordant results on the exon level. Longer exons tended to have better concordance between the two normalization methods than shorter exons.

2.4 RNA-seq vs Microarrays

Vall d'Hebron
Institut de Recerca

Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells

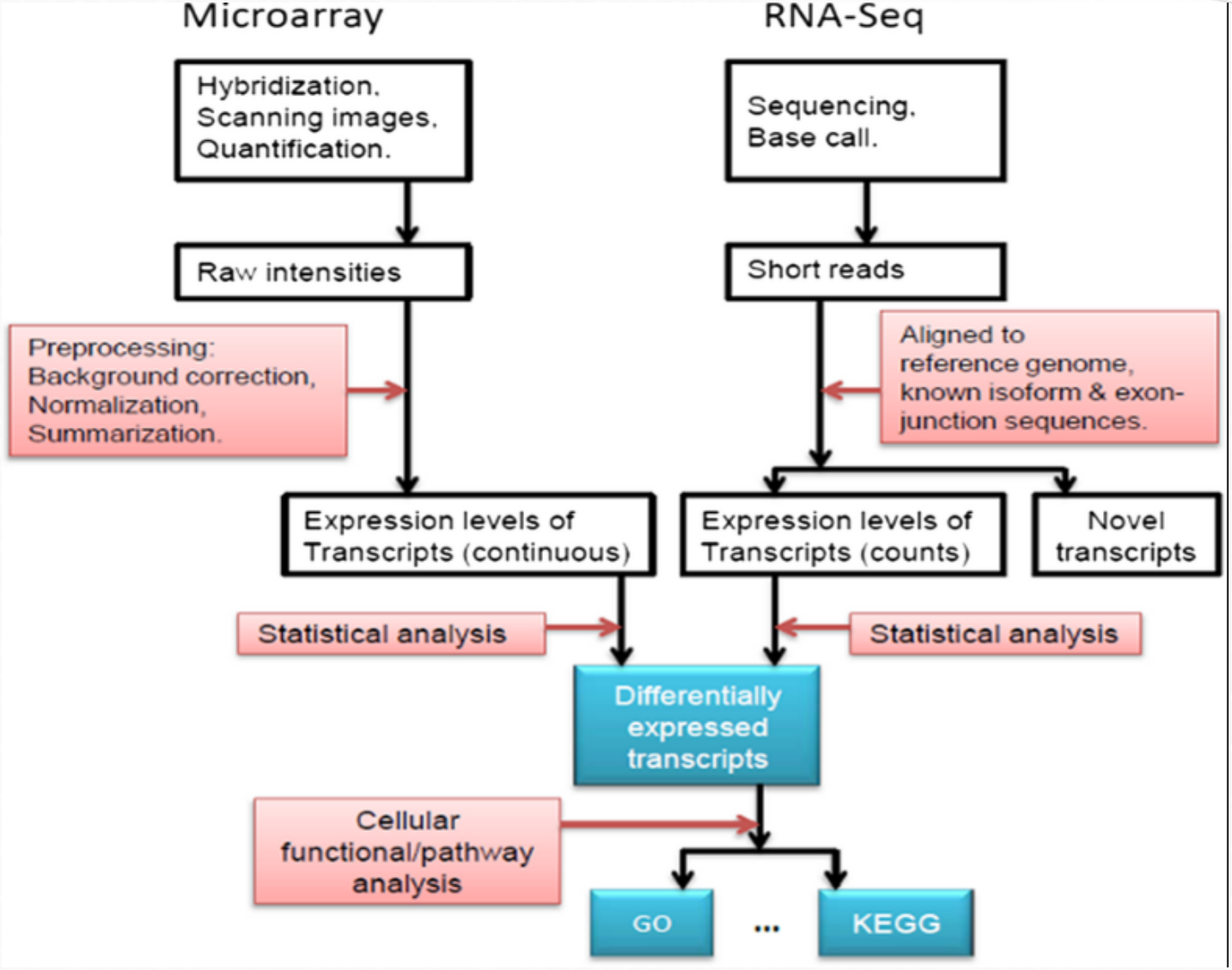
Shanrong Zhao, Wai-Ping Fung-Leung, Anton Bittner, Karen Ngo, Xuejun Liu

Published: January 16, 2014 • DOI: 10.1371/journal.pone.0078644

- RNA-Seq was **superior in detecting low abundance** transcripts
- also **better detecting differentiating biologically isoforms**
- RNA-Seq demonstrated a **broader dynamic range** than microarray.
- RNA-Seq **avoid problems inherent to microarray probe performance**

!!!The study try to demonstrate the benefits of RNA-Seq over microarray in transcriptome profiling

2.4 RNA-seq vs Microarrays



2.4 RNA-seq vs Microarrays

Vall d'Hebron
Institut de Recerca

Pros and cons of both technologies

Microarrays	**RNA-seq**
Costs,	High reproducibility,
well established methods, small data	not limited to expression
Hybridization bias,	Costs,
sequence must be known	complexity of analysis

"**High correlation** between gene expression profiles generated by the two platforms."

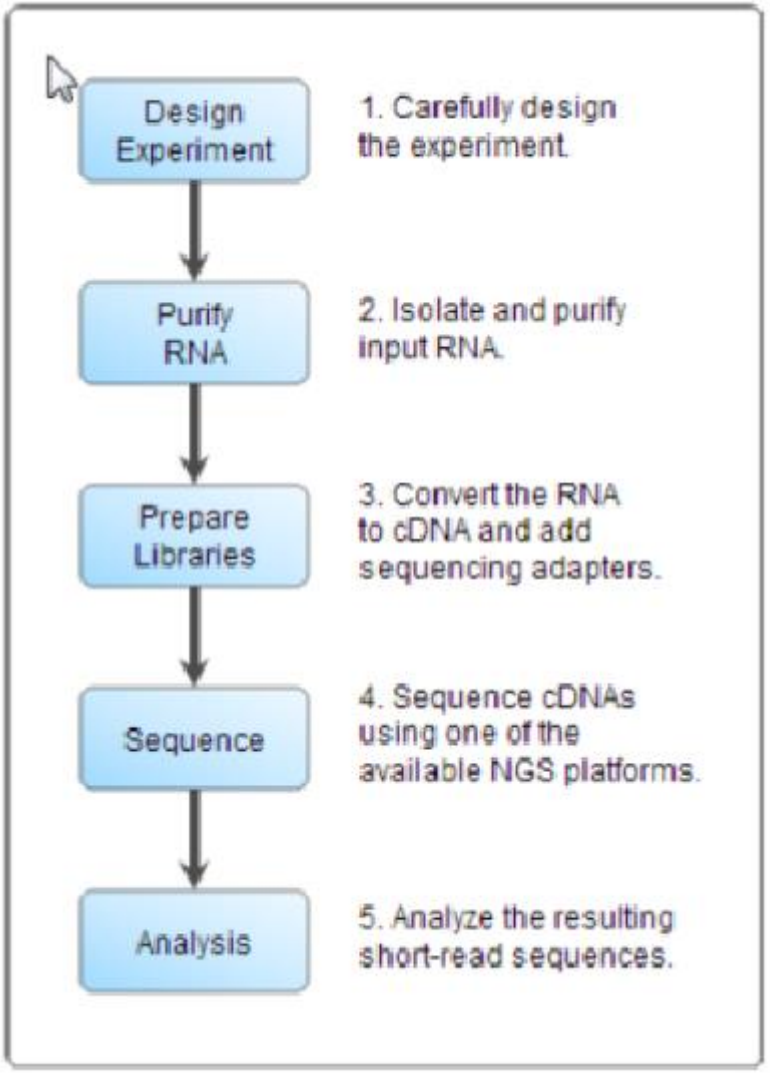
"RNA-Seq sequencing technology is new to most researchers, more expensive than microarray, **data storage is more challenging** and **analysis is more complex**."

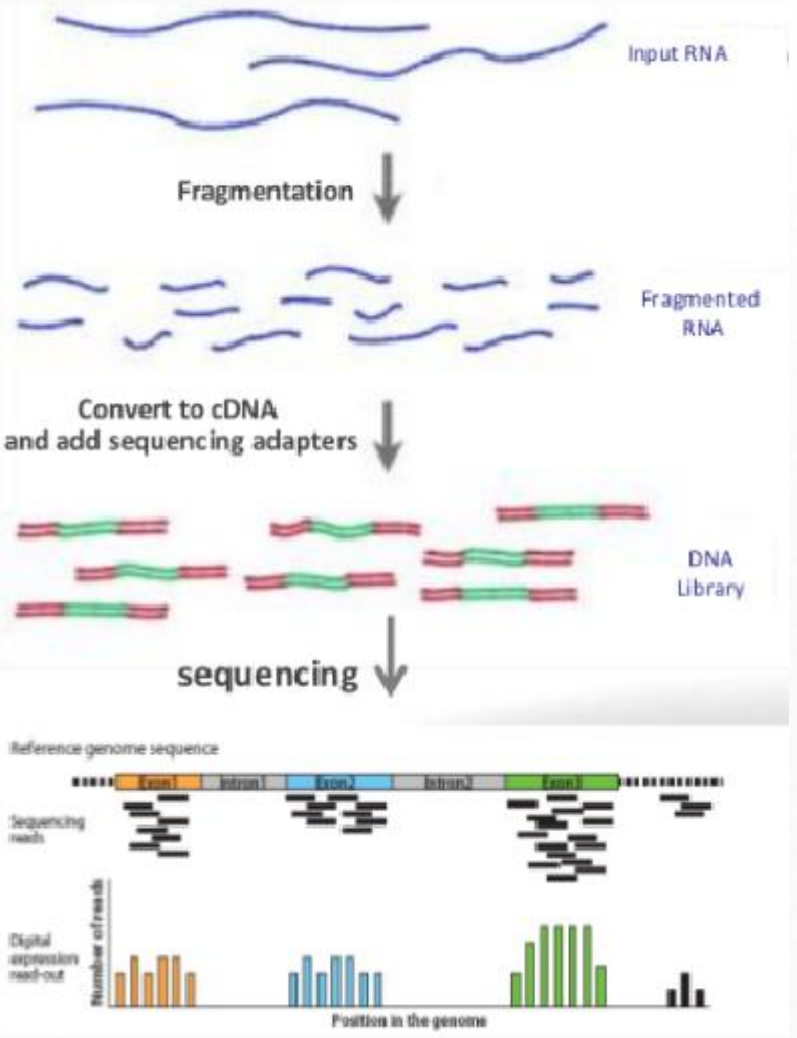
Zhao et al, 2014

Vall d'Hebron
Institut de Recerca

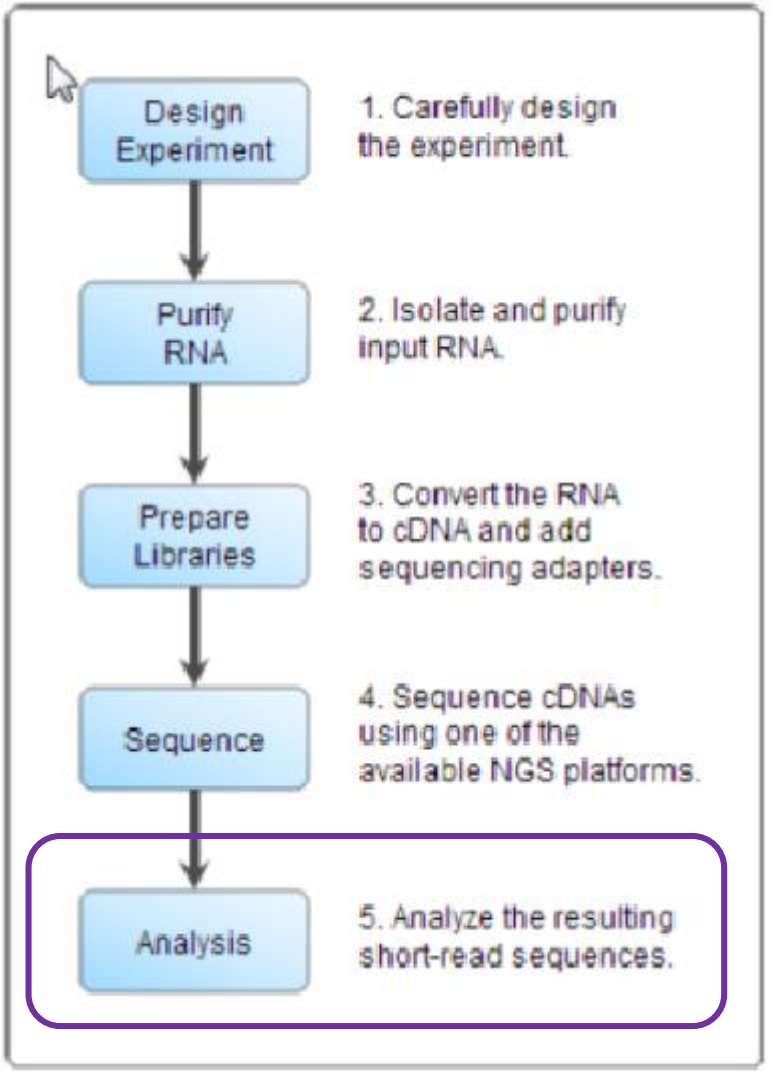
1. Introduction to omics data analysis
2. An example of omics data analysis. RNA-seq
    1. What is RNA-seq
    2. Basic key concepts
    3. Main challenges in RNA-seq
    4. RNA-seq vs Microarrays
    5. RNA-seq analysis pipeline(s)
    6. Alignment
    7. Transcript assembly
    8. DEG Analysis
    9. Vizualization

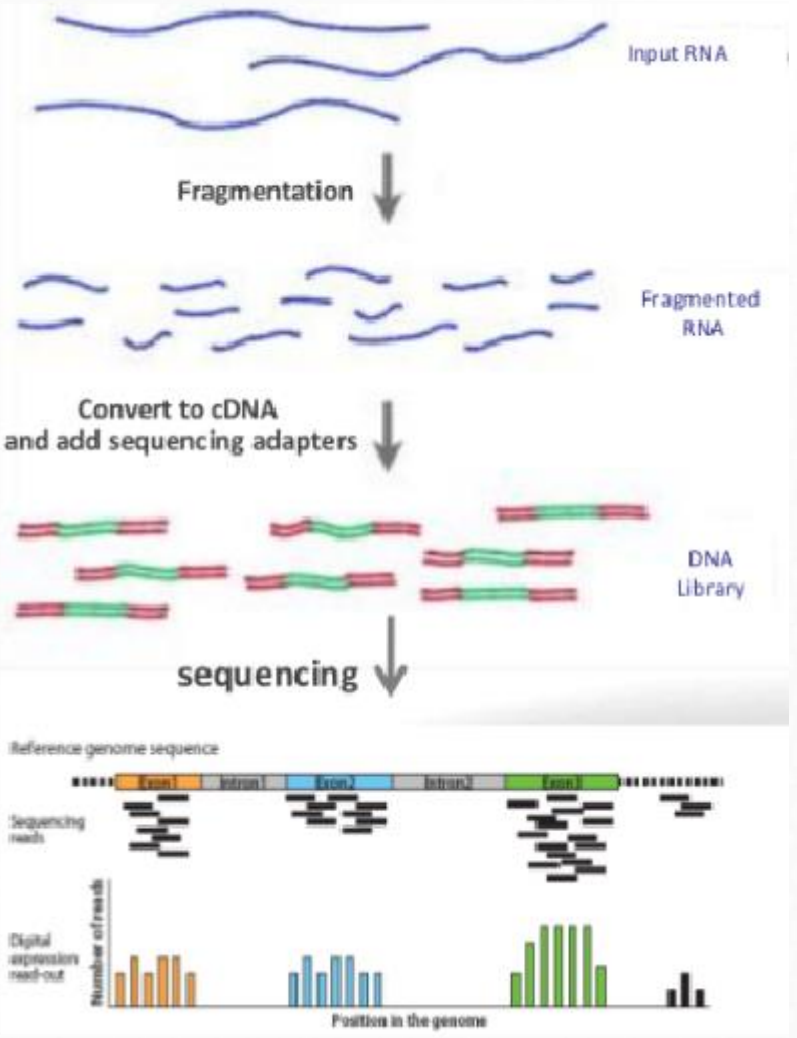
2.5 RNA-seq analysis pipeline(s)



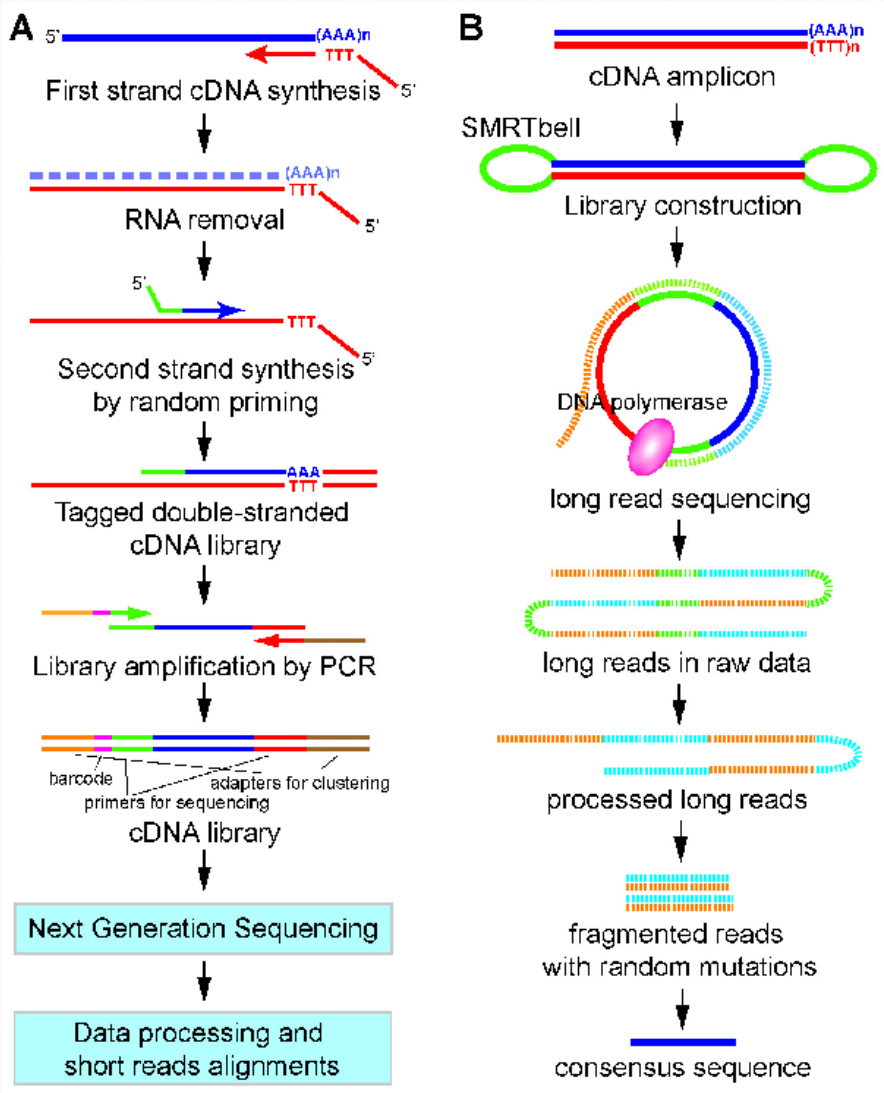


2.5 RNA-seq analysis pipeline(s)





2.5 RNA-seq analysis pipeline(s)



RNA-Seq workflow

A= RNA-Seq (SGS)

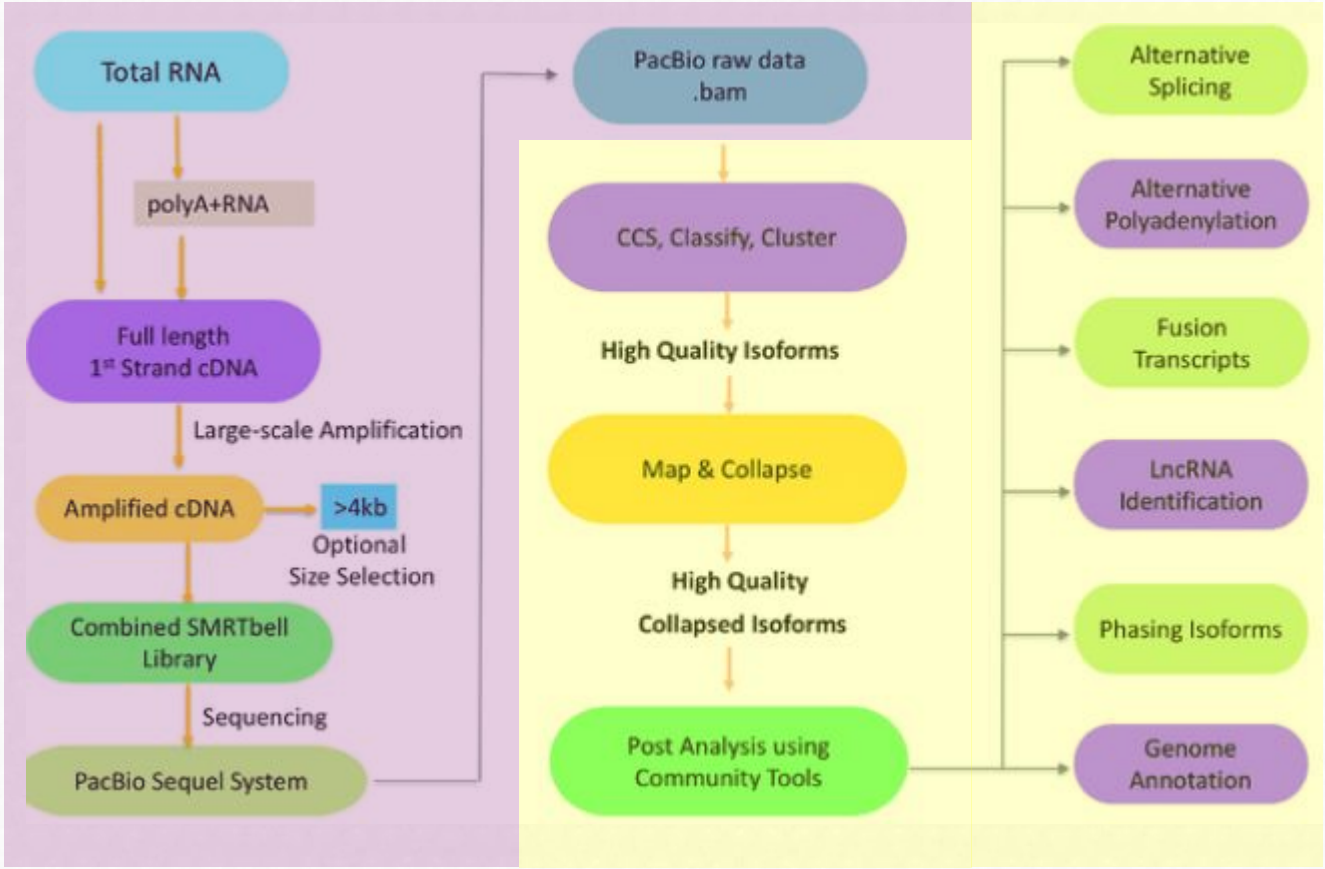
B= RNA-Seq (TGS) -> Iso-Seq (PacBio)

https://pubmed.ncbi.nlm.nih.gov/28148393/

Analysis

2.5 RNA-seq analysis pipeline(s)

RNA-Seq (TGS) pipeline exemple: **Iso-Seq**



https://www.frontiersin.org/articles/10.3389/fgene.2019.00384/full

2.5 RNA-seq analysis pipeline(s)

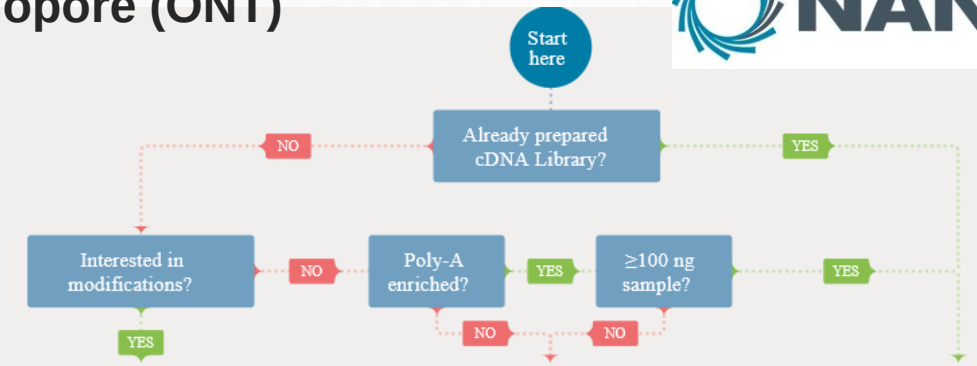
Vall d'Hebron Institut de Recerca

Oxford NANOPORE Technologies

RNA-Seq (TGS) : **Oxford nanopore (ONT)**

Tenen 3 kits!

https://nanoporetech.com/applications/rna-sequencing



	Direct RNA Sequencing Kit	cDNA-PCR Sequencing Kit	Direct cDNA Sequencing Kit
Preparation time	105 mins	165 mins	275 mins
Input requirement	500 ng RNA (poly-A$^+$)	1 ng RNA (poly-A$^+$)	100 ng RNA (poly-A$^+$)
RT required	Optional	Yes	Yes
PCR required	No	Yes	No
Read length	Equal to RNA length	Enriched for full-length cDNA	Enriched for full-length cDNA
Typical throughput	●○○	●●●	●●○
Typical number of reads (MinION, 1 kb avg. length)	1 million	7 - 12 million	5 - 10 million
Typical number of reads (PromethION, 1 kb avg. length)	8 million	> 60 million	30 - 60 million
Multiplexing options	In development	PCR Barcoding Kit	Native Barcoding Kit

2.5 RNA-seq analysis pipeline(s)

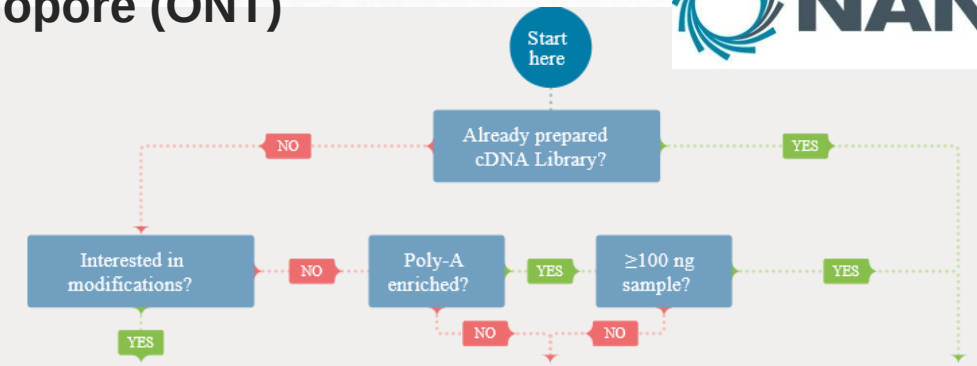
Vall d'Hebron Institut de Recerca

Oxford NANOPORE Technologies

RNA-Seq (TGS) : **Oxford nanopore (ONT)**

Tenen 3 kits!

https://nanoporetech.com/applications/rna-sequencing



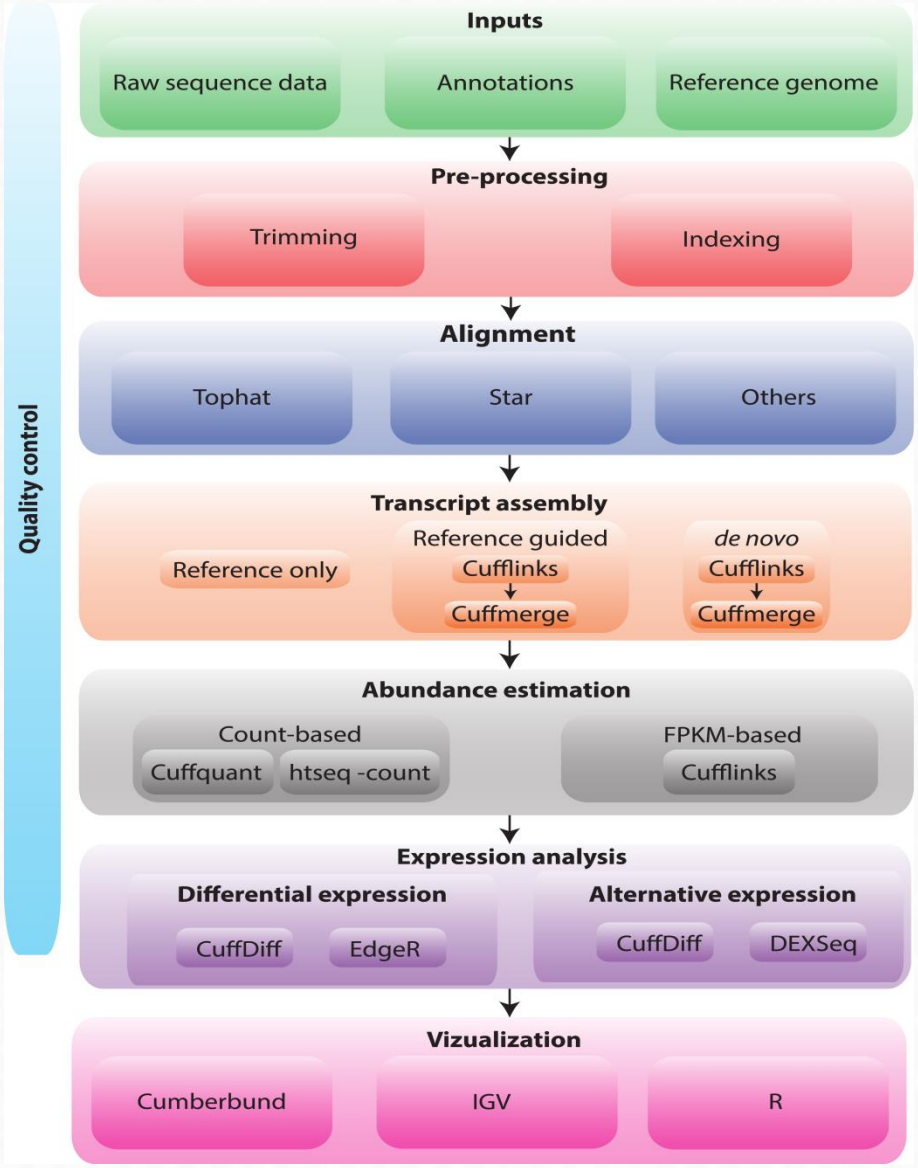
	Direct RNA Sequencing Kit	cDNA-PCR Sequencing Kit	Direct cDNA Sequencing Kit
Preparation time	105 mins	165 mins	275 mins
Input requirement	500 ng RNA (poly-A$^+$)	1 ng RNA (poly-A$^+$)	100 ng RNA (poly-A$^+$)
RT required	Optional	Yes	Yes
PCR required	No	Yes	No
Read length	Equal to RNA length	Enriched for full-length cDNA	Enriched for full-length cDNA
Typical throughput	●○○	●●●	●●○
Typical number of reads (MinION, 1 kb avg. length)	1 million	7 - 12 million	5 - 10 million
Typical number of reads (PromethION, 1 kb avg. length)	8 million	> 60 million	30 - 60 million
Multiplexing options	In development	PCR Barcoding Kit	Native Barcoding Kit

2.5 RNA-seq analysis pipeline(s)

RNA-Seq (SGS)

Steps Standard pipeline

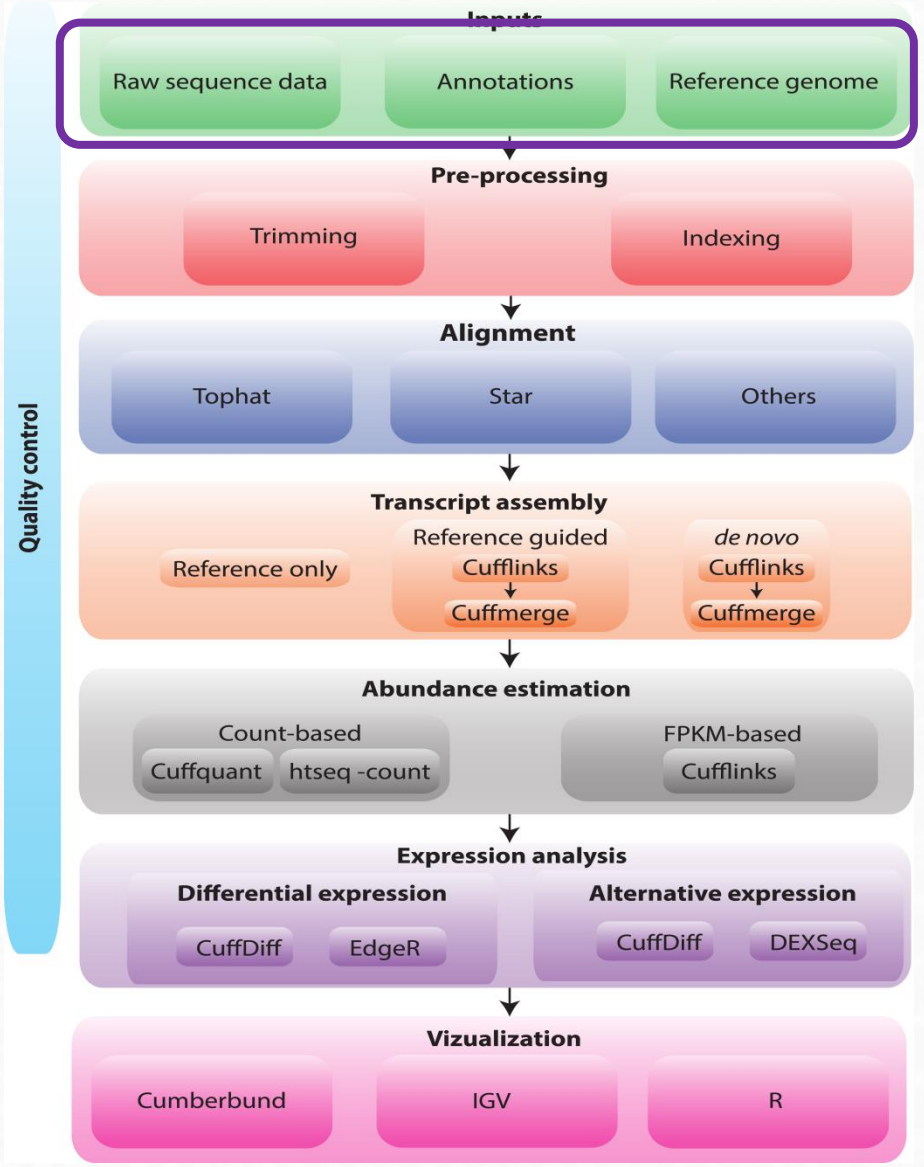
2.5 RNA-seq analysis pipeline(s)



RNA-Seq (SGS)

Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

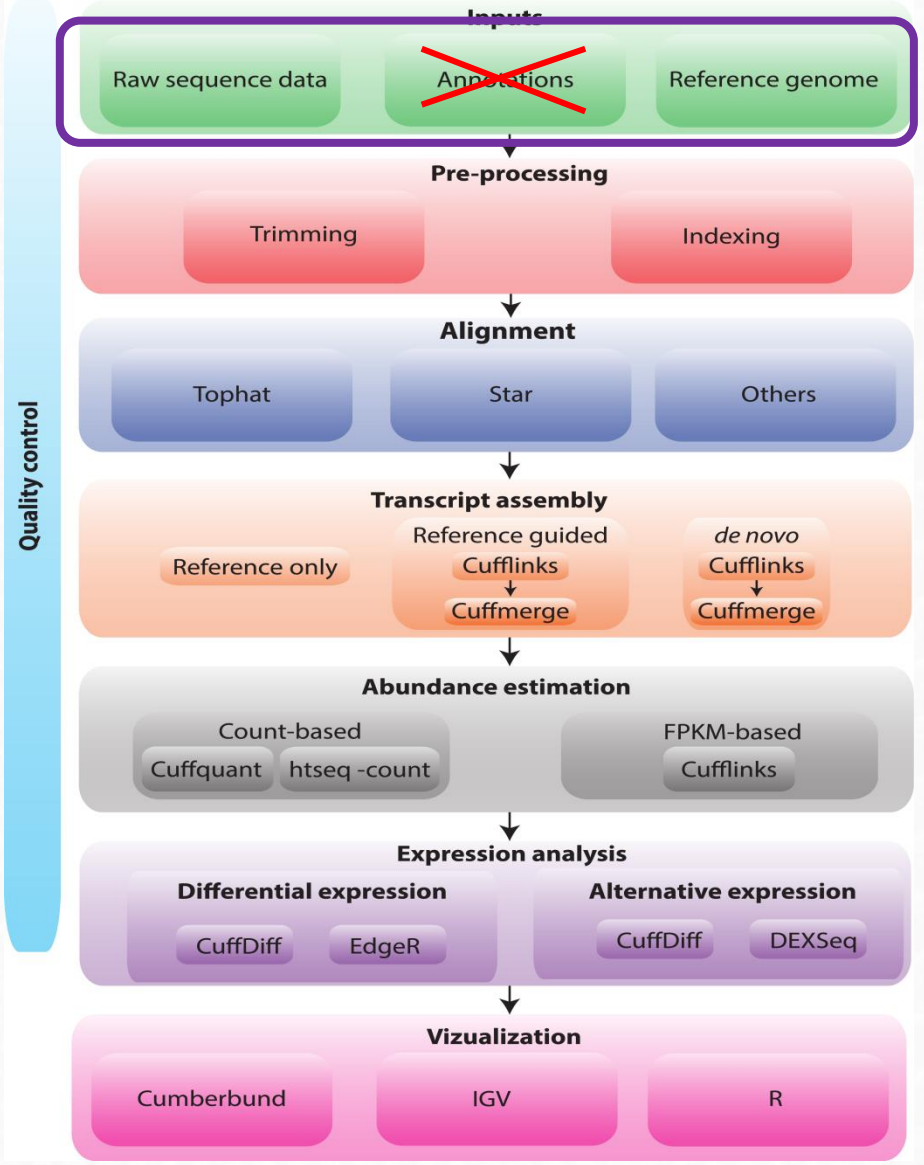
2.5 RNA-seq analysis pipeline(s)



- Fastq files
- FASTQC (quality control)

Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

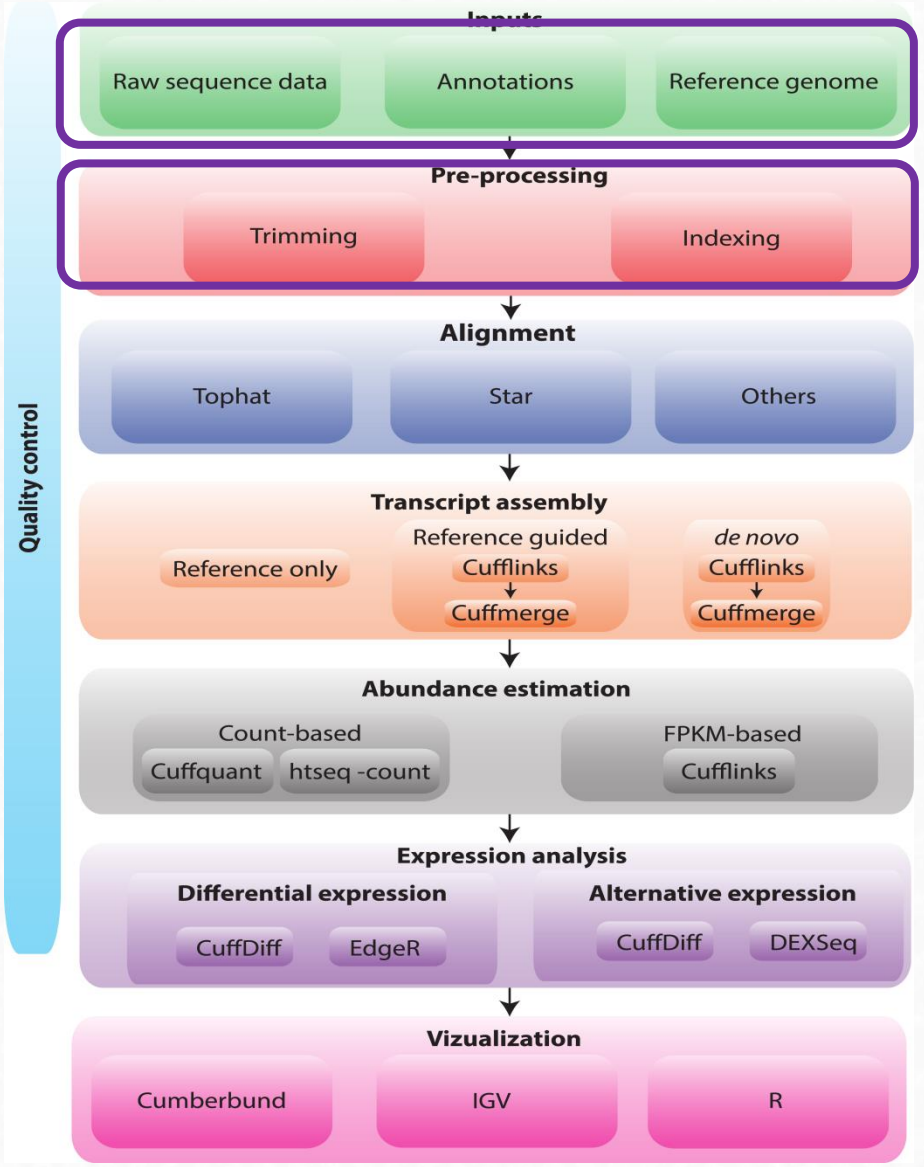
2.5 RNA-seq analysis pipeline(s)



- Fastq files
- FASTQC (quality control)

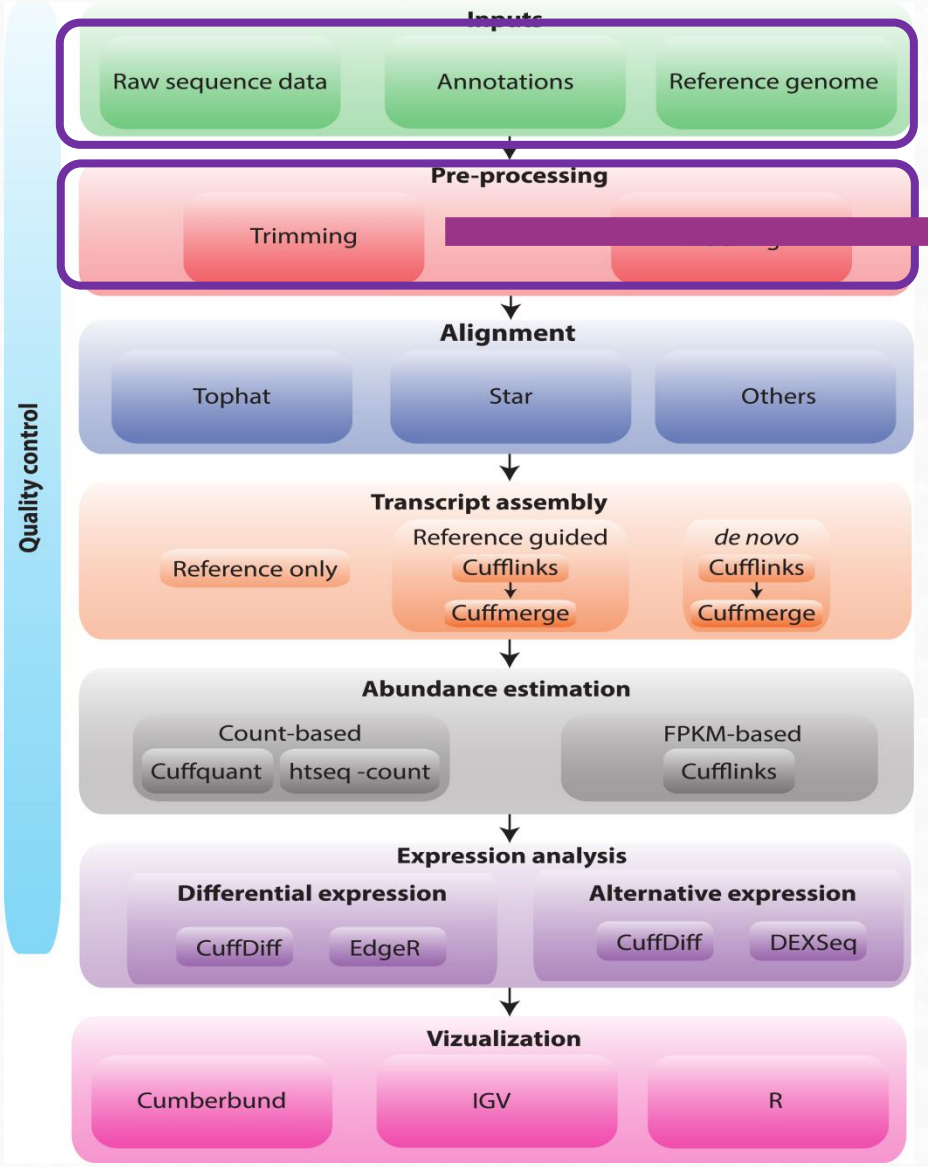
Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

2.5 RNA-seq analysis pipeline(s)



Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

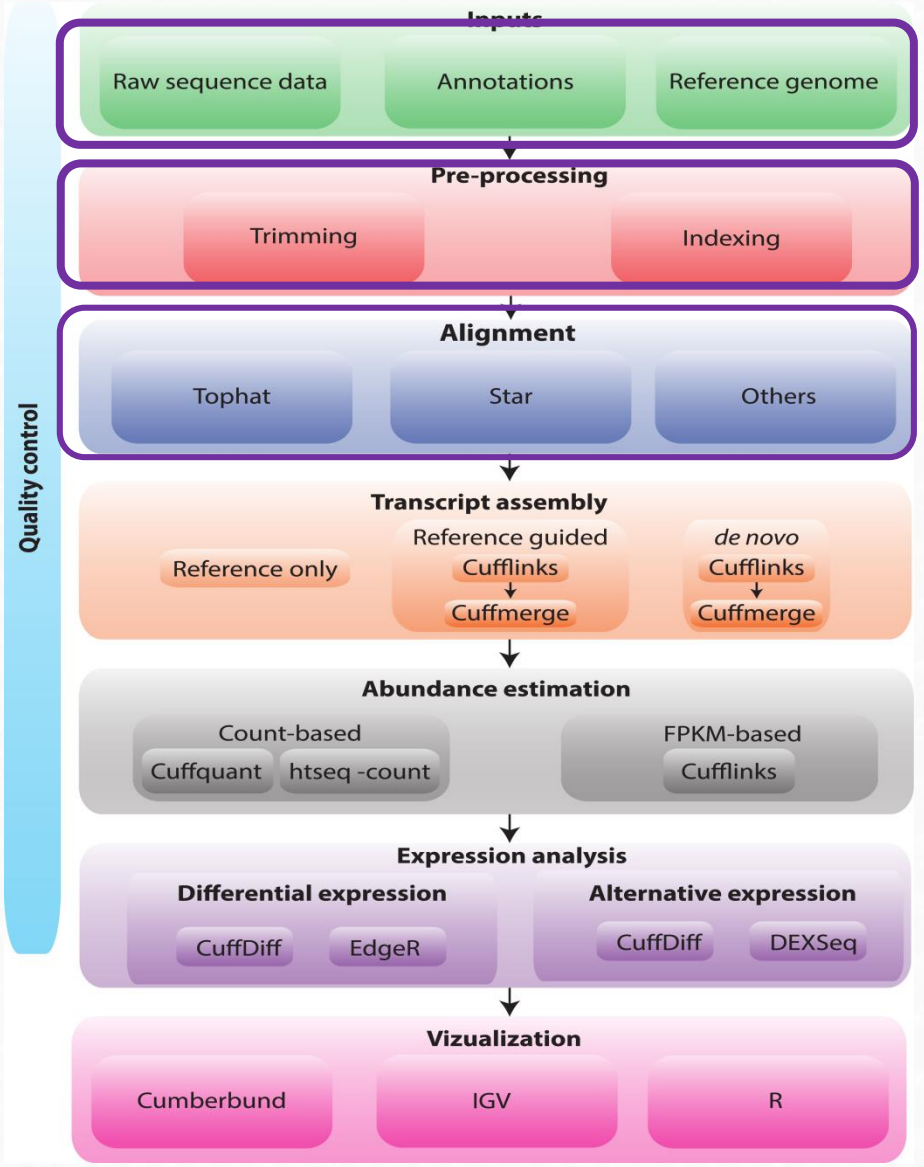
2.5 RNA-seq analysis pipeline(s)



Cutadapt
Trimmomatic
...

Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

2.5 RNA-seq analysis pipeline(s)



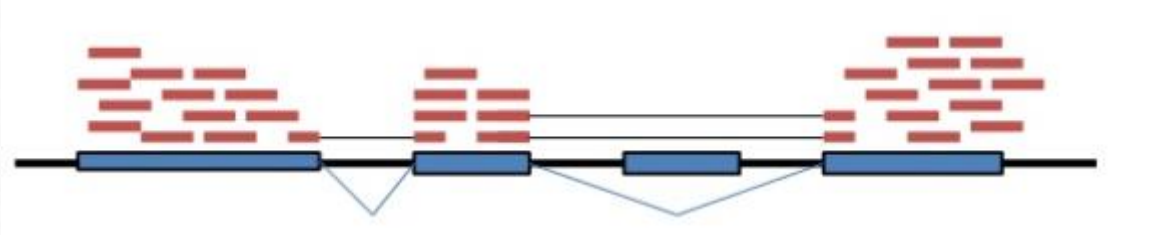
Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

Vall d'Hebron
Institut de Recerca

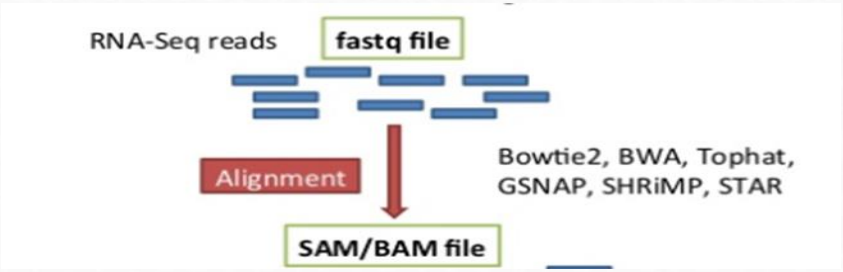
1. Introduction to omics data analysis
2. An example of omics data analysis. RNA-seq
    1. What is RNA-seq
    2. Basic key concepts
    3. Main challenges in RNA-seq
    4. RNA-seq vs Microarrays
    5. RNA-seq analysis pipeline(s)
    6. Alignment
    7. Transcript assembly
    8. DEG Analysis
    9. Vizualization

What to map to?

Map to the genome, with knowledge of transcript annotations



- **Well annotated genome reference is required**.
- To effectively map to exon junctions, you need a **mapping algorithm** that can divide the sequencing reads and map portions independently.
- Identifying **alternative transcript isoforms** involves complex algorithms



Vall d'Hebron
Institut de Recerca

Which sequence mappers to use?

- RNASeq Alignment algorithm must be
  - **Fast**
  - Able to **handle SNPs, indels, and sequencing errors**
  - Maintain **accurate quantification**
  - Allow for introns for reference genome alignment (spliced alignment detection)

Vall d'Hebron
Institut de Recerca

Which sequence mappers to use?

- RNASeq Alignment algorithm must be
  - **Fast**
  - Able to **handle SNPs, indels, and sequencing errors**
  - Maintain **accurate quantification**
  - Allow for introns for reference genome alignment (spliced alignment detection)

- Burrows-Wheeler Transform (BWT) mappers
  - **Fast**
  - **Limited mismatches** allowed (<3)
  - **Limited indel detection** ability
  - Examples: Bowtie2, BWA, Tophat, HISAT2
  - Use cases: **large and conserved genome and transcriptomes**

- Hash Table mappers
  - Require **large amount of RAM** for indexing
  - **More mismatches allowed**
  - **Indel detection**
  - Examples: GSNAP, SHRiMP, STAR
  - Use case: highly **variable or smaller genomes, transcriptomes**

Vall d'Hebron
Institut de Recerca

Which sequence mappers to use?

- RNASeq Alignment algorithm must be
  - **Fast**
  - Able to **handle SNPs, indels, and sequencing errors**

Front Genet. 2018; 9: 35. PMCID: PMC5834436
Published online 2018 Feb 26. doi: [10.3389/fgene.2018.00035] PMID: 29535759

Comparison of Burrows-Wheeler Transform-Based Mapping Algorithms Used in High-Throughput Whole-Genome Sequencing: Application to Illumina Data for Livestock Genomes[1]

Brittney N. Keel* and Warren M. Snelling

- Hash Table mappers
  - Require **large amount of RAM** for indexing
  - **More mismatches allowed**
  - **Indel detection**
  - Examples: GSNAP, SHRiMP, STAR
  - Use case: highly **variable or smaller genomes, transcriptomes**

2.6 Alignment

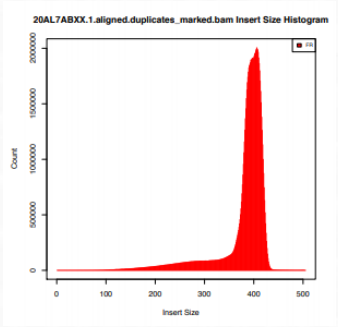
Steps with TopHat

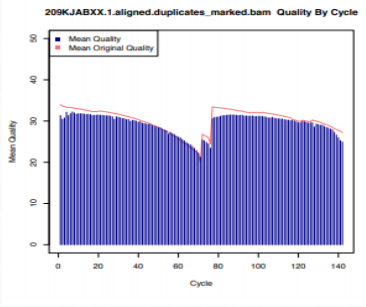
1. **Unspliced reads** are mapped to locate exons (with Bowtie)

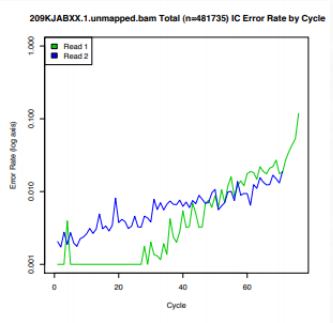
2. Unmapped reads are then split and aligned independently to identify exon

Important to **check the quality** of mapping process (percentage of mapped reads)

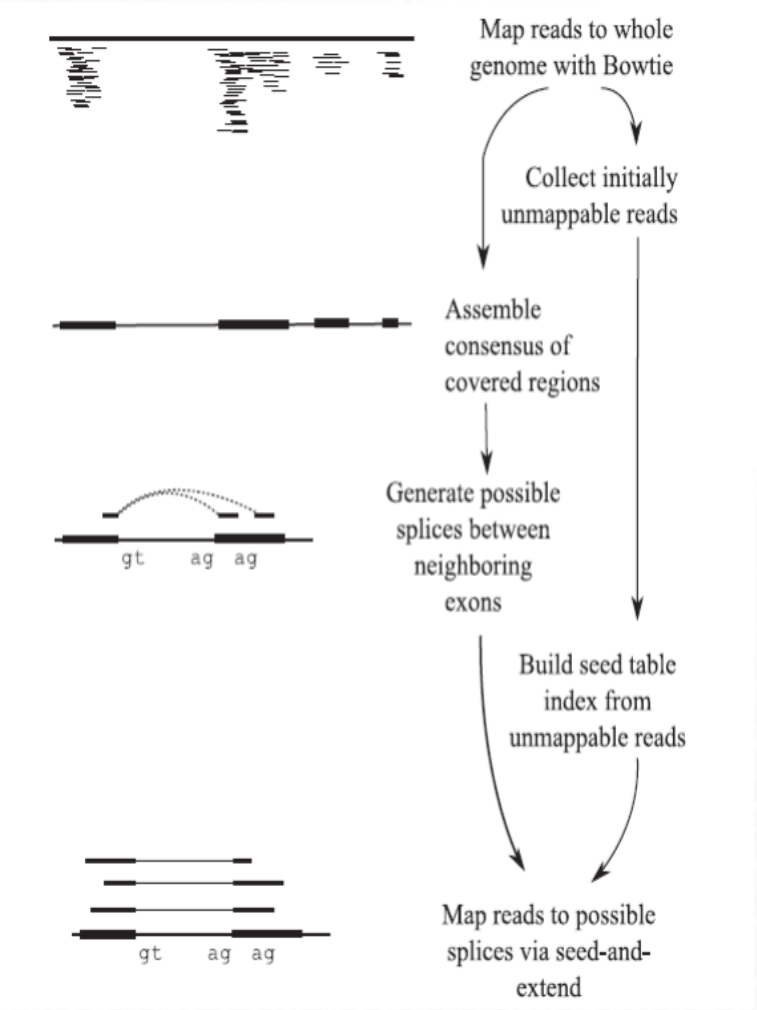
Picard can be used for quality control of mapping







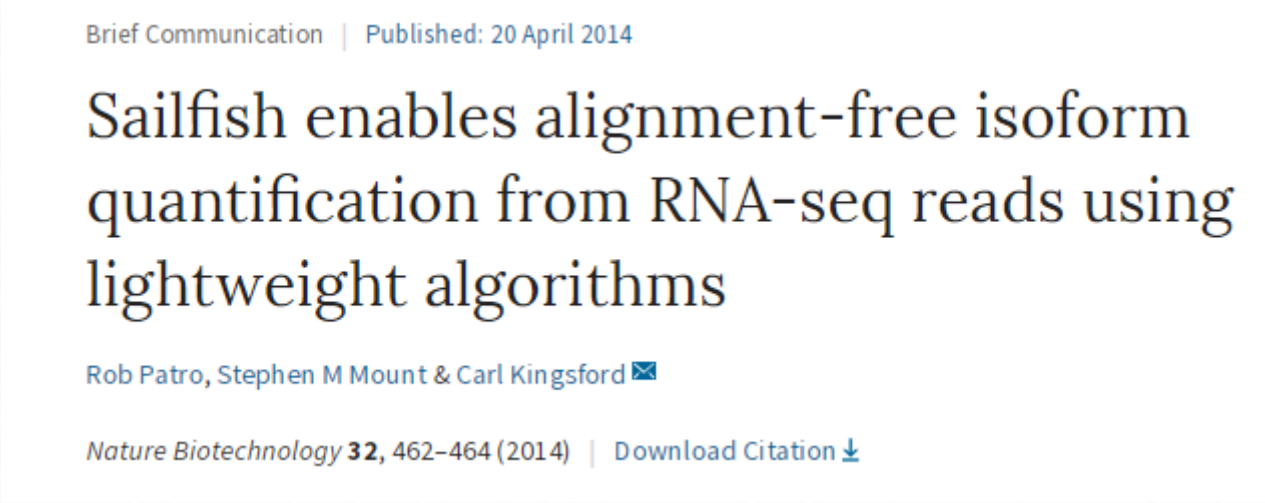
TopHat Pipeline



Vall d'Hebron
Institut de Recerca

Alignment-independent quantification for RNA-Seq

- Alignment steps are **computationally heavy** and can be very **time-consuming** even with multi-threading.
- In 2014, **Sailfish** method, demonstrated that it was not necessary to actually **align each read to the genome** in order to obtain accurate transcript each read.



Brief Communication | Published: 20 April 2014

Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms

Rob Patro, Stephen M Mount & Carl Kingsford

Nature Biotechnology **32**, 462–464 (2014) | Download Citation

https://cgatoxford.wordpress.com/2015/10/12/road-testing-kallisto/

Vall d'Hebron
Institut de Recerca

Alignment-independent quantification for RNA-Seq

- Alignment steps are **computationally heavy** and can be very **time-consuming** even with multi-threading.
- In 2014, **Sailfish** method, demonstrated that it was not necessary to actually **align each read to the genome** in order to obtain accurate transcript each read.
- All you actually need to do is establish the **most likely transcript** for each read

1. shredding the transcriptome and reads into **kmers** (short overlapping sequences)
2. **matching the transcriptome and read kmers** (is a very fast and low memory usage)

https://cgatoxford.wordpress.com/2015/10/12/road-testing-kallisto/

Vall d'Hebron
Institut de Recerca

Alignment-independent quantification for RNA-Seq

- Nowadays there exist various tools:

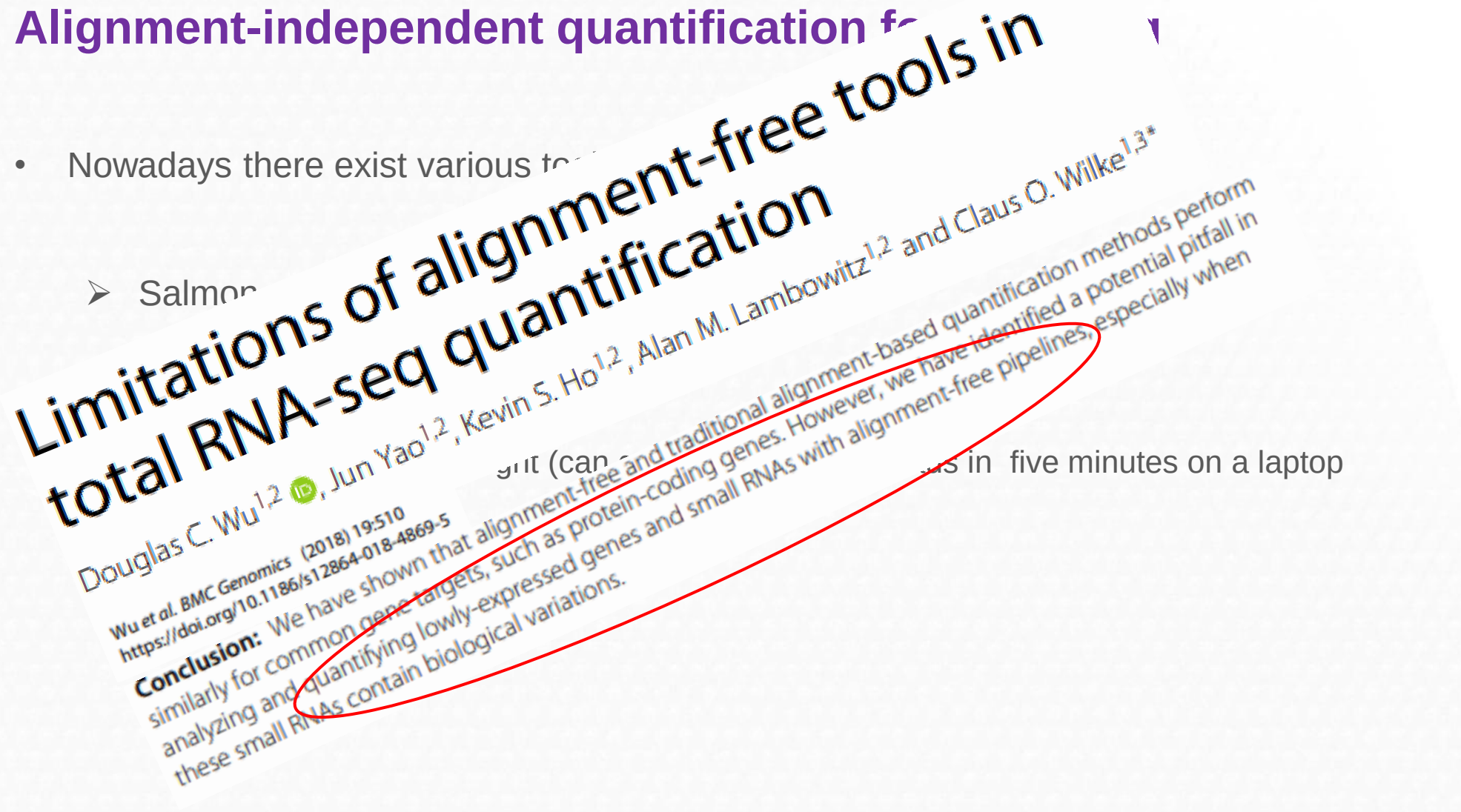
  - Salmon, Kallisto, Sailfish

PROS:

- Extremely Fast & Lightweight (can quantify 20 million reads in five minutes on a laptop computer)
- Easy to use

Alignment-independent quantification fo

- Nowadays there exist various to
  - Salmon



Limitations of alignment-free tools in total RNA-seq quantification

Douglas C. Wu[1,2], Jun Yao[1,2], Kevin S. Ho[1,2], Alan M. Lambowitz[1,2] and Claus O. Wilke[1,3*]

Wu et al. BMC Genomics (2018) 19:510
https://doi.org/10.1186/s12864-018-4869-5

Conclusion: We have shown that alignment-free and traditional alignment-based quantification methods perform similarly for common gene targets, such as protein-coding genes. However, we have identified a potential pitfall in analyzing and quantifying lowly-expressed genes and small RNAs with alignment-free pipelines, especially when these small RNAs contain biological variations.

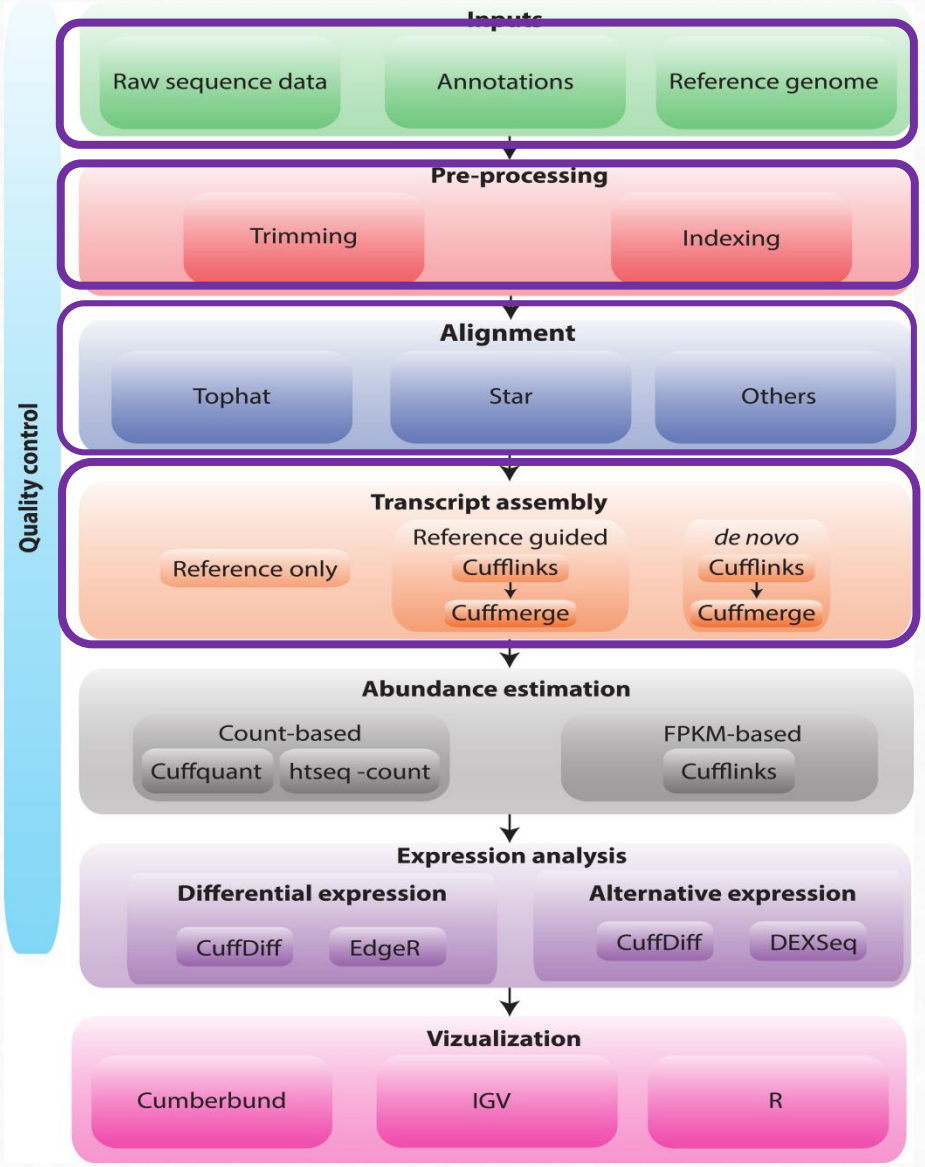
gnit (can

s in five minutes on a laptop

Vall d'Hebron
Institut de Recerca

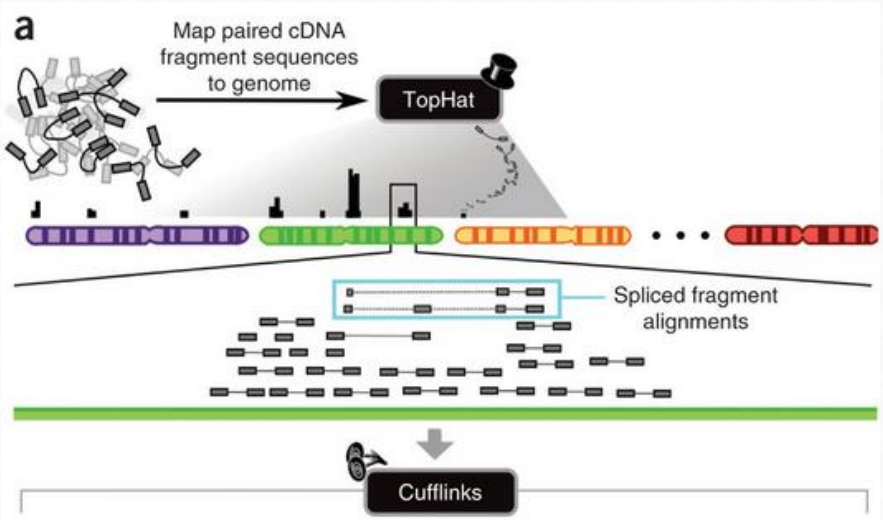
1. Introduction to omics data analysis
2. An example of omics data analysis. RNA-seq
    1. What is RNA-seq
    2. Basic key concepts
    3. Main challenges in RNA-seq
    4. RNA-seq vs Microarrays
    5. RNA-seq analysis pipeline(s)
    6. Alignment
    7. Transcript assembly
    8. DEG Analysis
    9. Vizualization

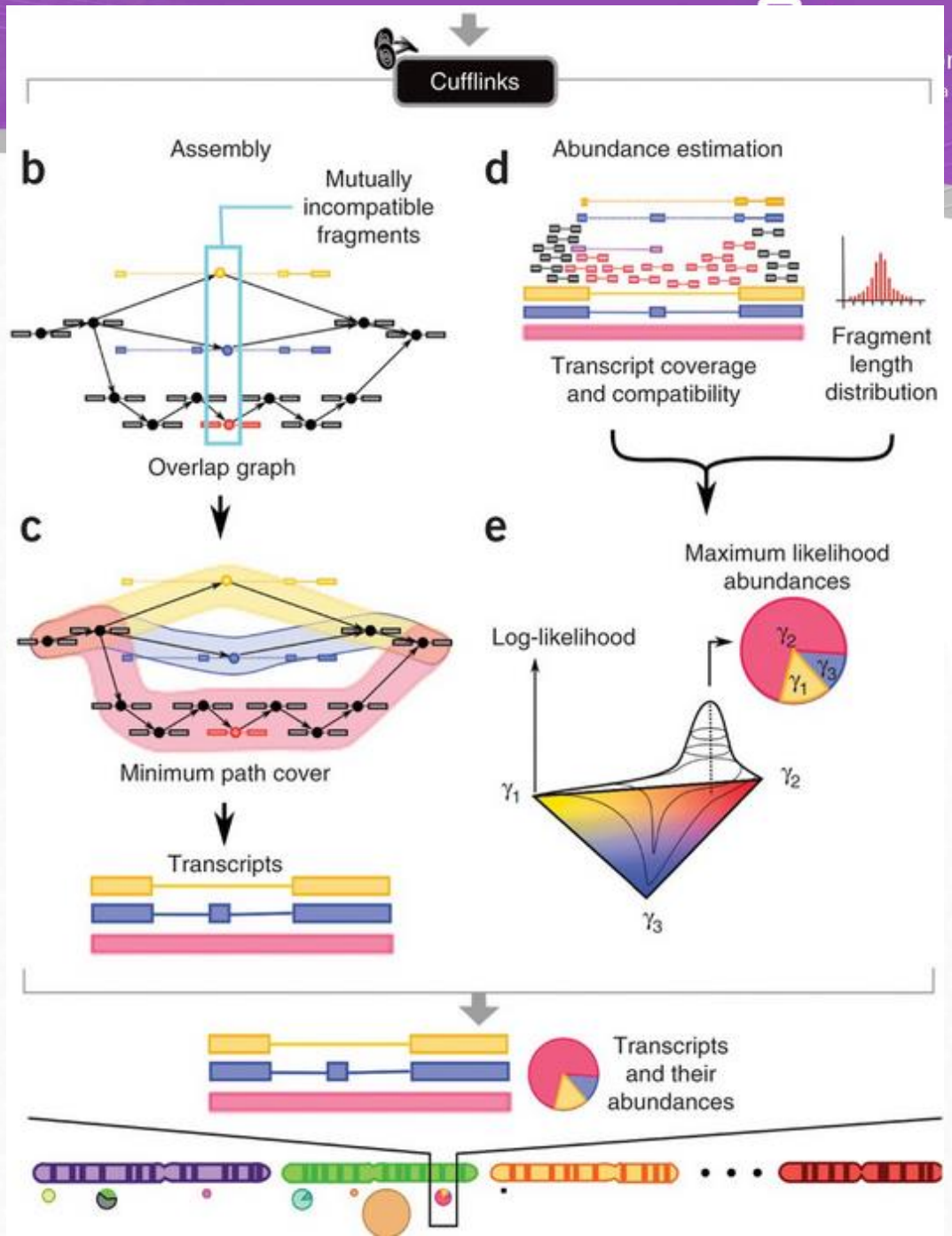
2.5 RNA-seq analysis pipeline(s)



Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

2.5 RNA-seq analysis pipeline(s)



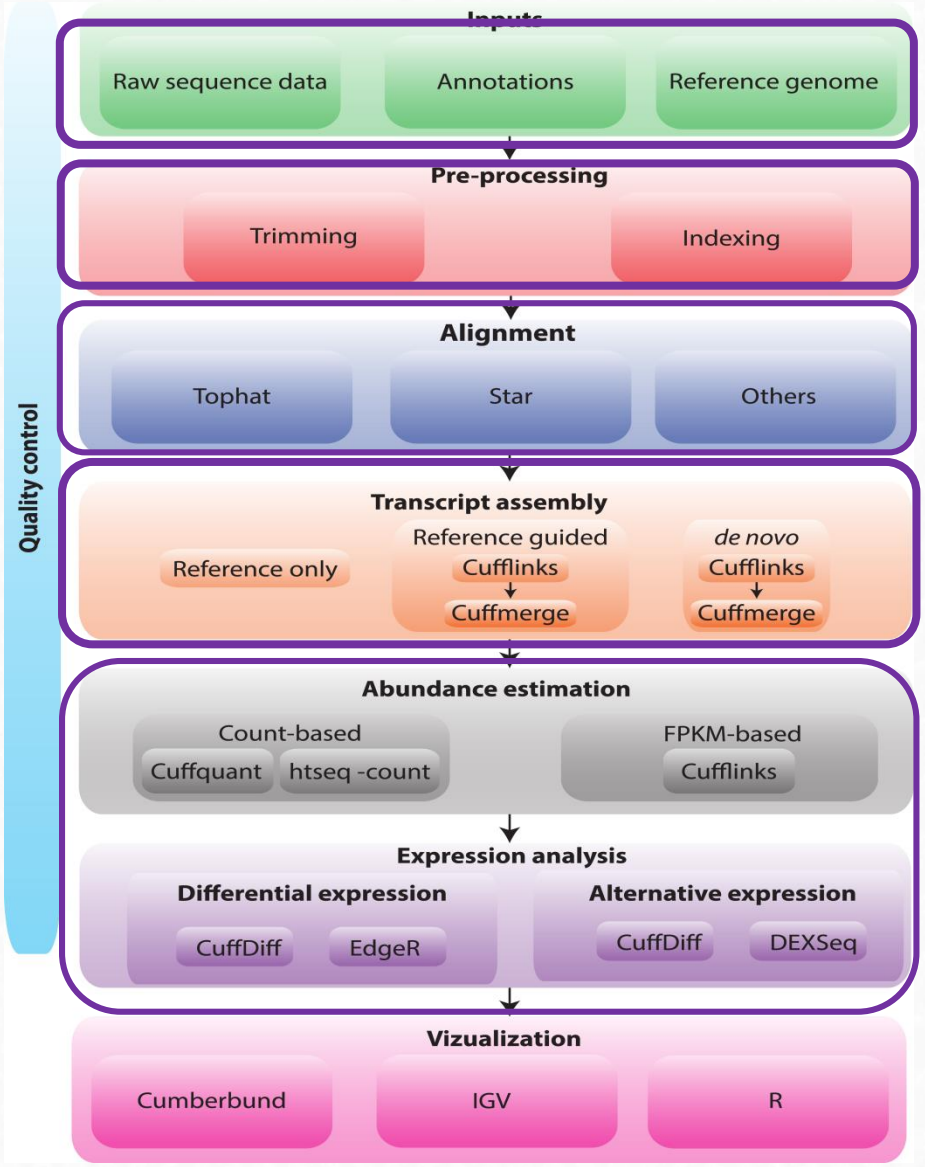


https://www.nature.com/articles/nbt.1621

Vall d'Hebron
Institut de Recerca

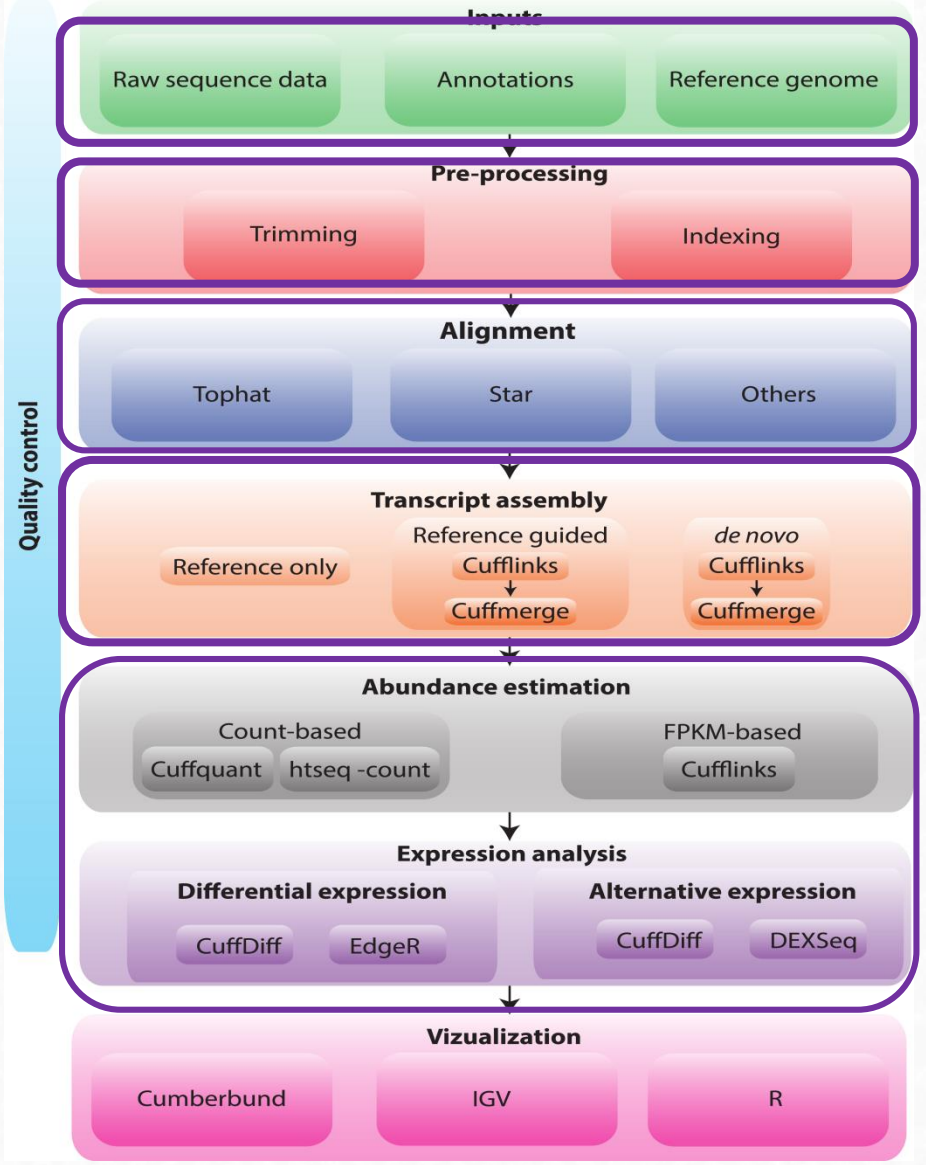
1. Introduction to omics data analysis
2. An example of omics data analysis. RNA-seq
    1. What is RNA-seq
    2. Basic key concepts
    3. Main challenges in RNA-seq
    4. RNA-seq vs Microarrays
    5. RNA-seq analysis pipeline(s)
    6. Alignment
    7. Transcript assembly
    8. DEG Analysis
    9. Vizualization

2.5 RNA-seq analysis pipeline(s)



Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

2.5 RNA-seq analysis pipeline(s)

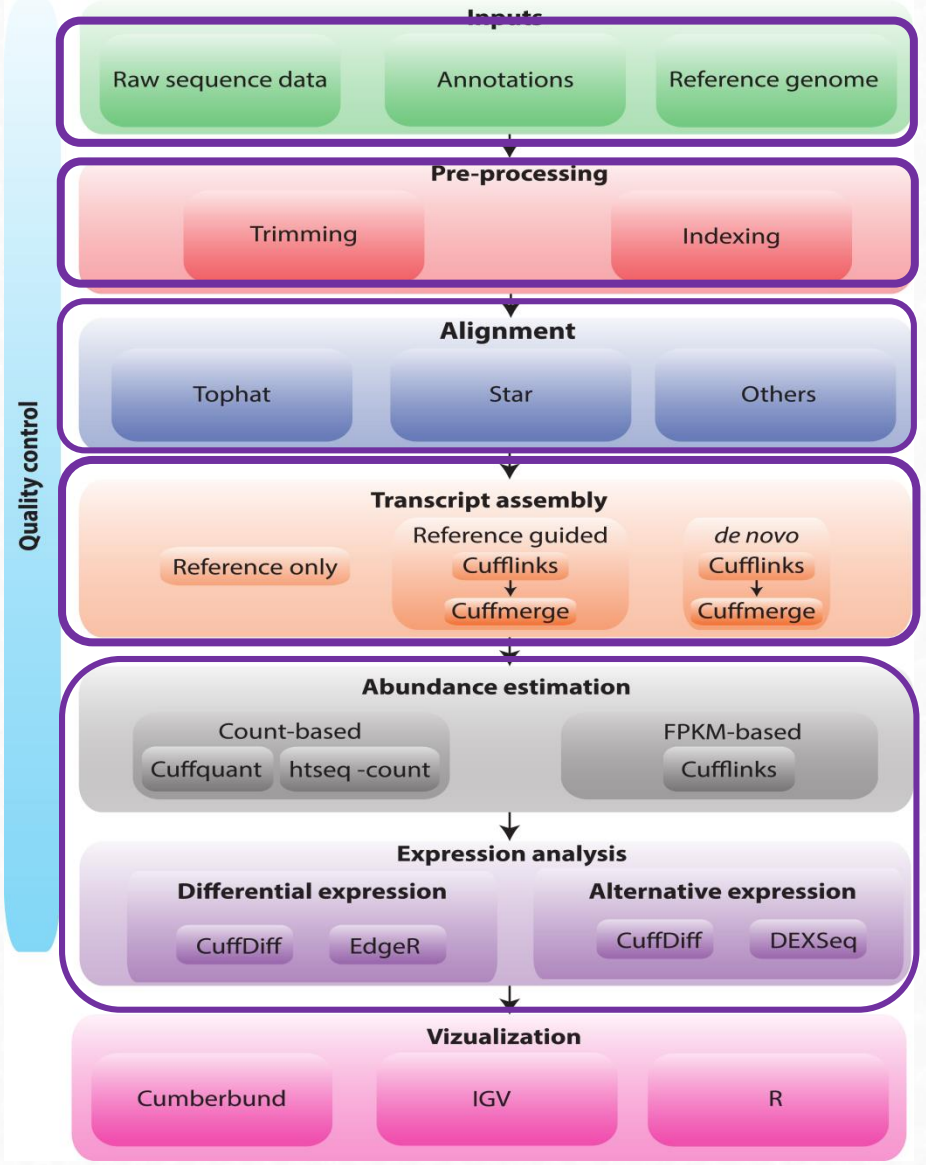


Count-based methods: assign the reads to transcripts directly.

Abundance based methods: assign abundance of each transcript with a probabilistic model that makes use of info such as fragment length distribution etc. Many software tools want raw counts for input because they do their own normalization.

Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

2.5 RNA-seq analysis pipeline(s)



Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

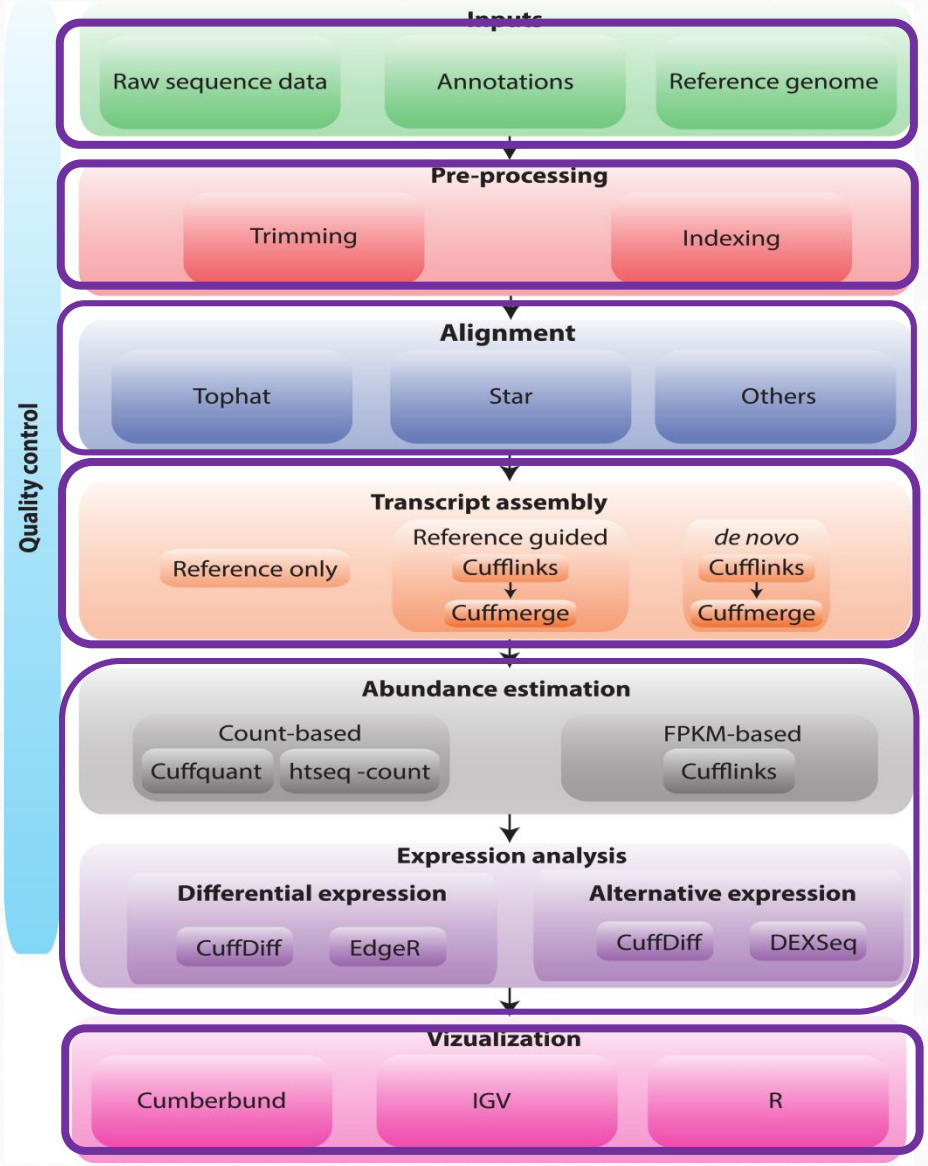
Differential expression: taking the normalized read count data and performing statistical analysis to discover quantitative changes in expression levels between experimental groups.

Alternative expression: analysis of RNA-seq data to catalog transcripts and assess alternative expression of known and predicted mRNA isoforms in cells and tissues

Vall d'Hebron
Institut de Recerca

1. Introduction to omics data analysis
2. An example of omics data analysis. RNA-seq
    1. What is RNA-seq
    2. Basic key concepts
    3. Main challenges in RNA-seq
    4. RNA-seq vs Microarrays
    5. RNA-seq analysis pipeline(s)
    6. Alignment
    7. Transcript assembly
    8. DEG Analysis
    9. Visualization

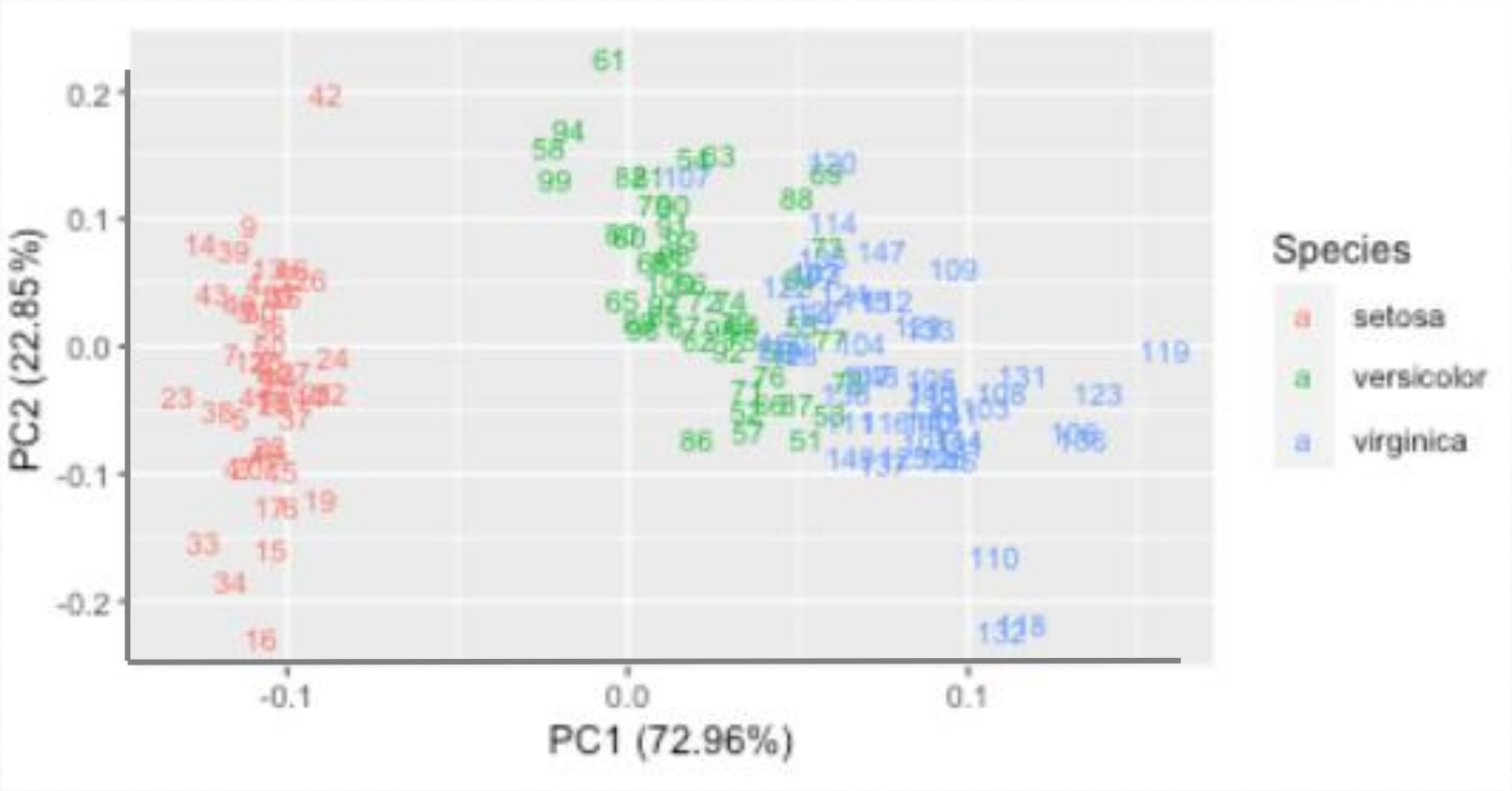
2.5 RNA-seq analysis pipeline(s)



Griffith, 2015 DOI: 10.1371/journal.pcbi.1004393

2.10 Visualization

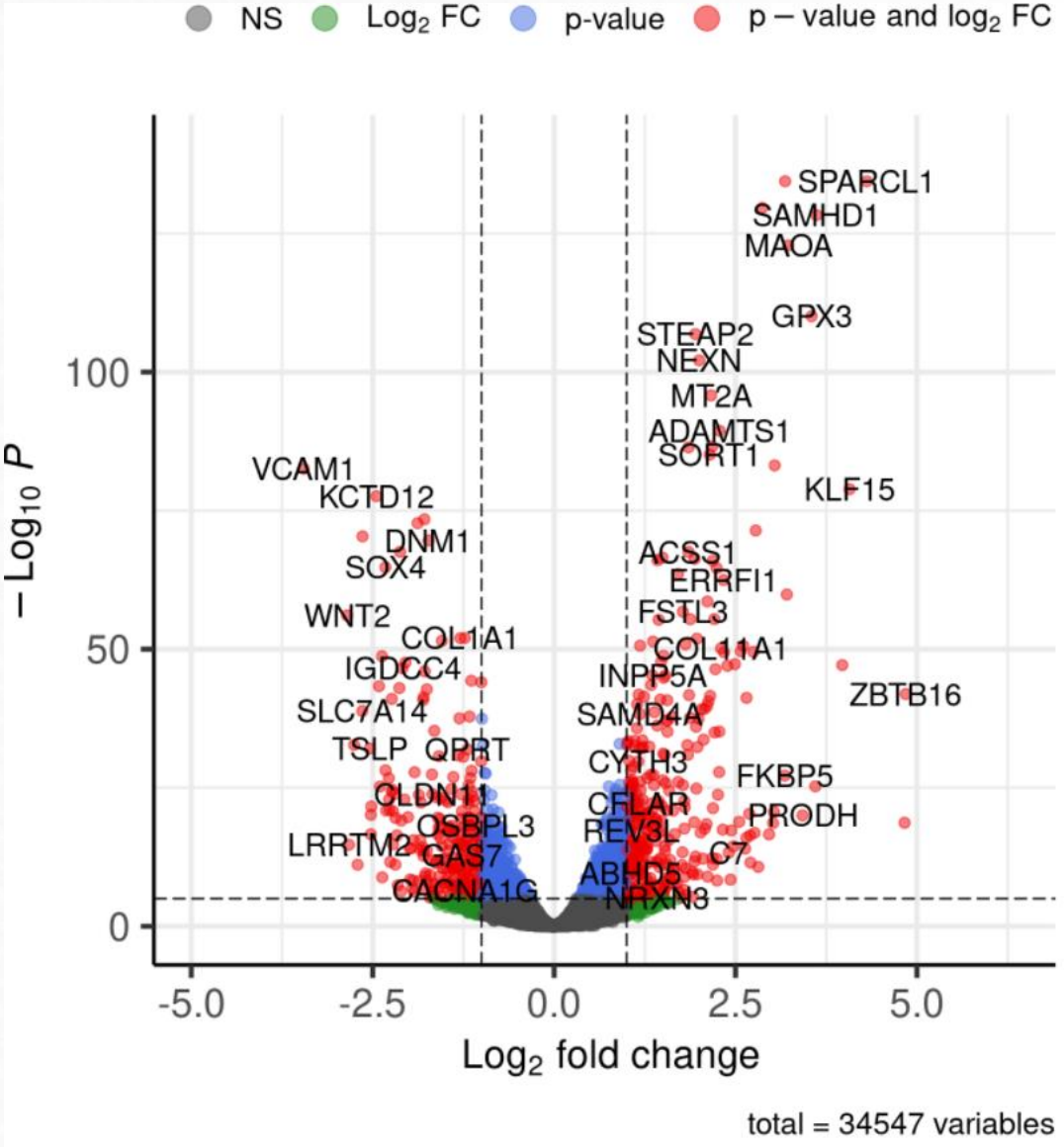
PCA



https://cran.r-project.org/web/packages/ggfortify/vignettes/plot_pca.html

2.10 Visualization

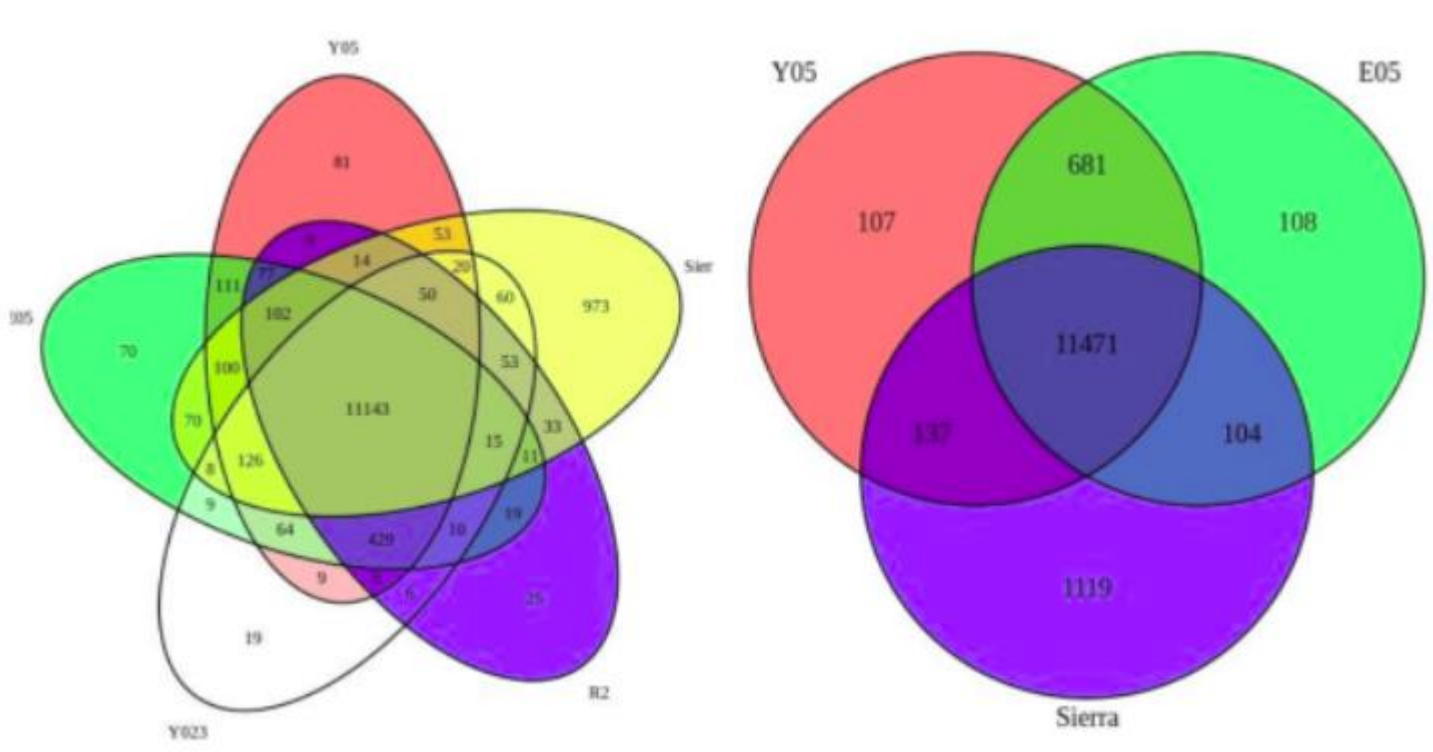
Volcano plots



https://www.bioconductor.org/packages/release/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html

2.10 Visualization

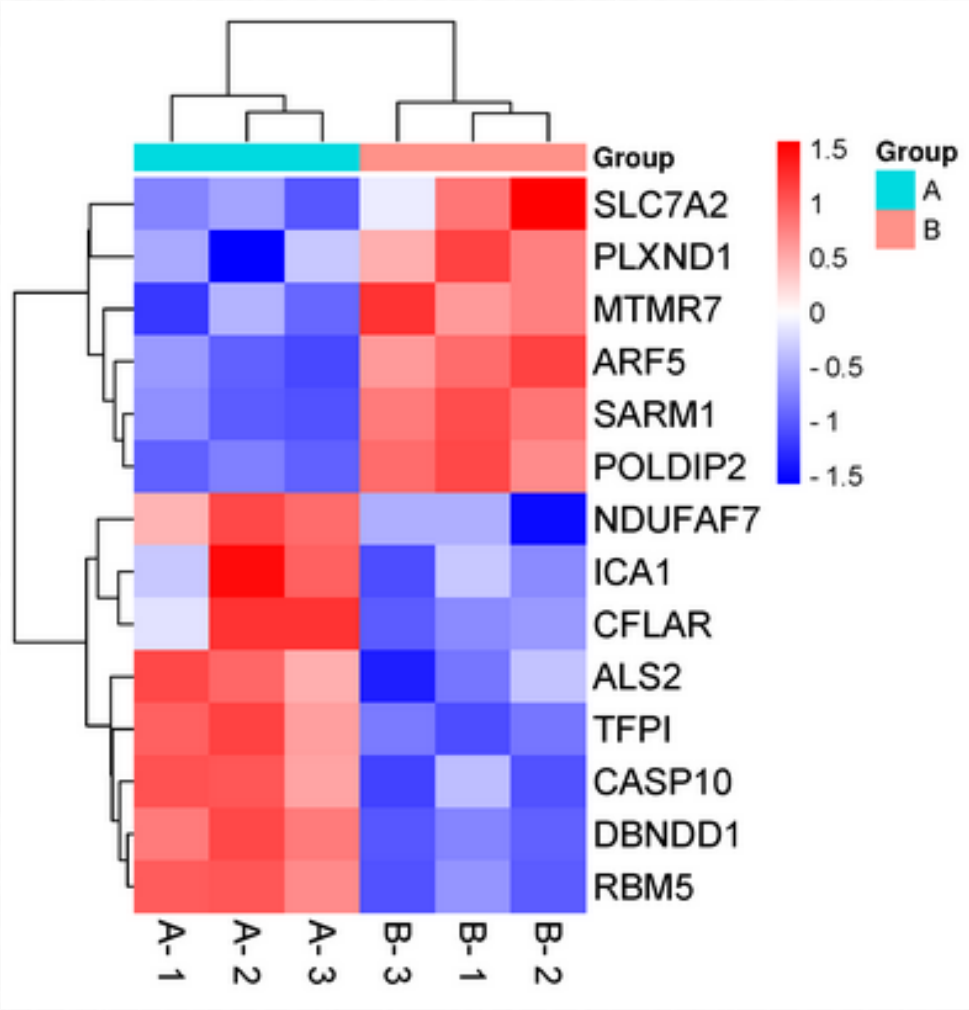
Venn diagrams



https://www.r-bloggers.com/2020/08/comparing-data-sets-with-venn-diagrams/

2.10 Visualization

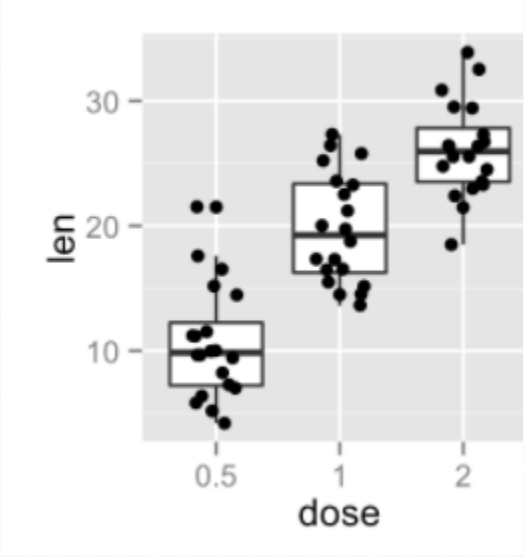
Heatmap

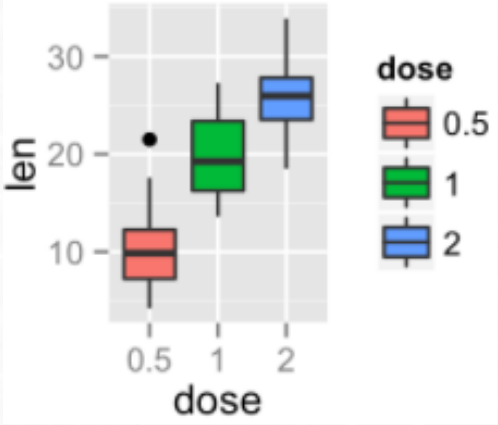


https://www.bioinformatics.com.cn/plot_basic_cluster_heatmap_plot_024_en

2.10 Visualization

Boxplot

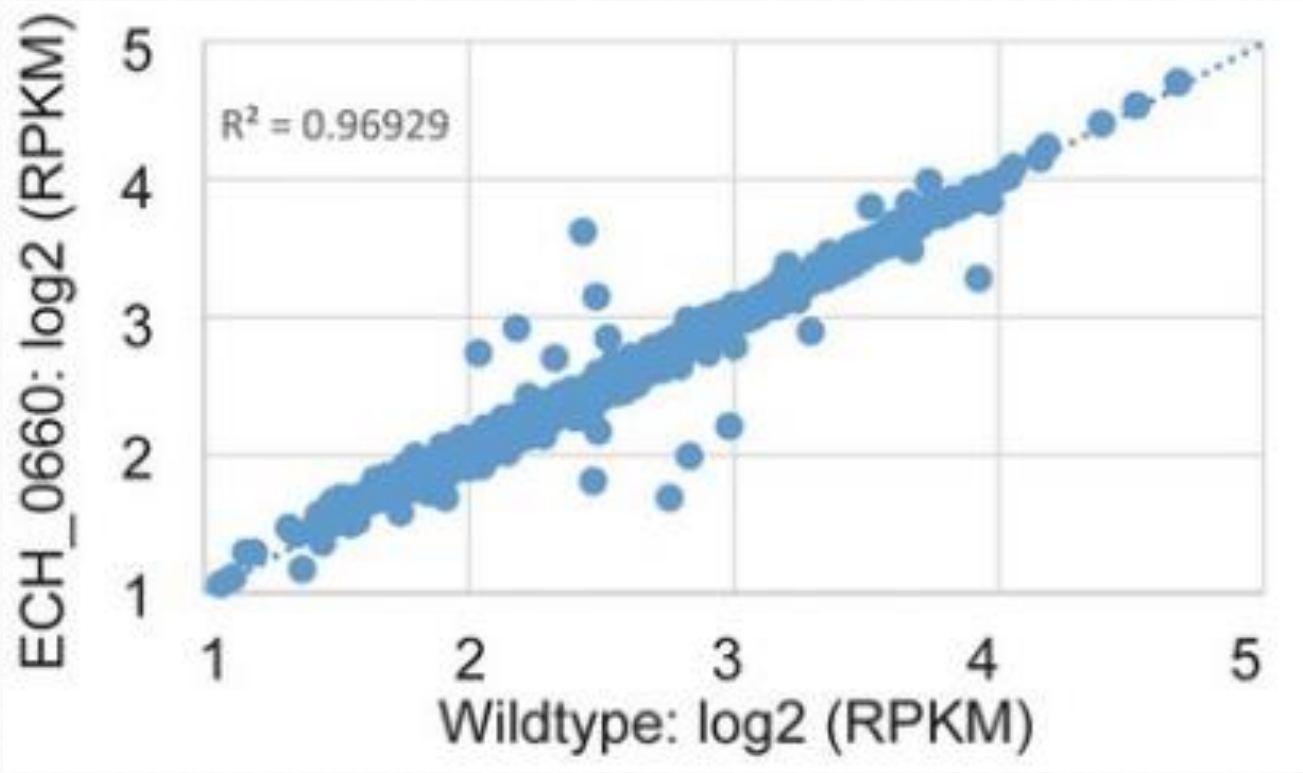




http://www.sthda.com/english/wiki/ggplot2-box-plot-quick-start-guide-r-software-and-data-visualization

2.10 Visualization

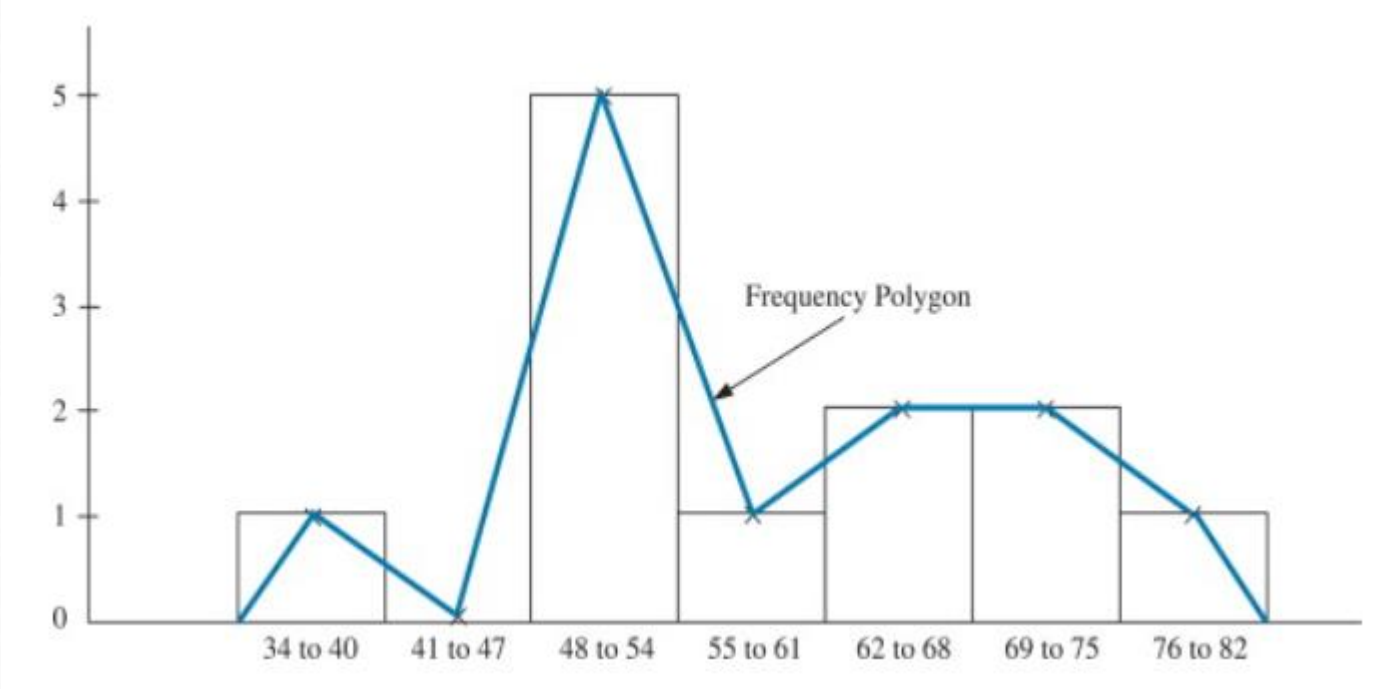
Scatter plot



https://www.nature.com/articles/s41598-018-24471-3

Vall d'Hebron
Institut de Recerca

Frequency plots

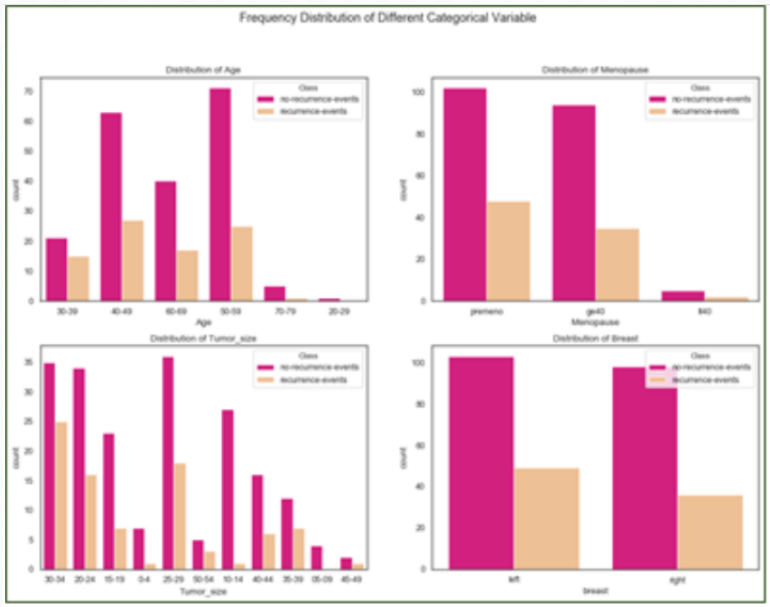


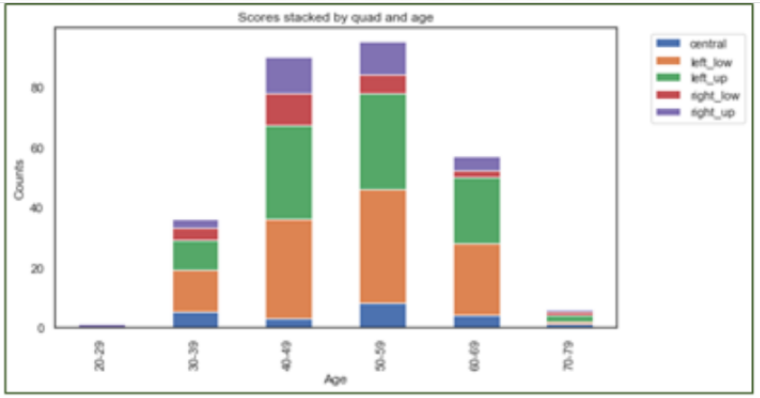
https://www.sciencedirect.com/topics/mathematics/frequency-polygon

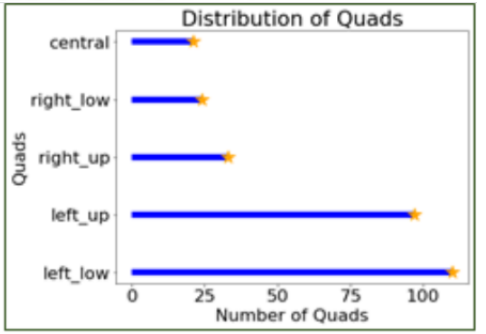
2.10 Visualization

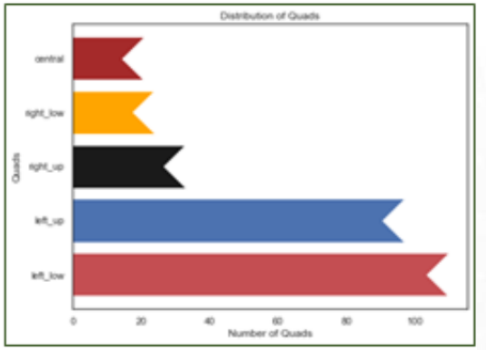
Barplots

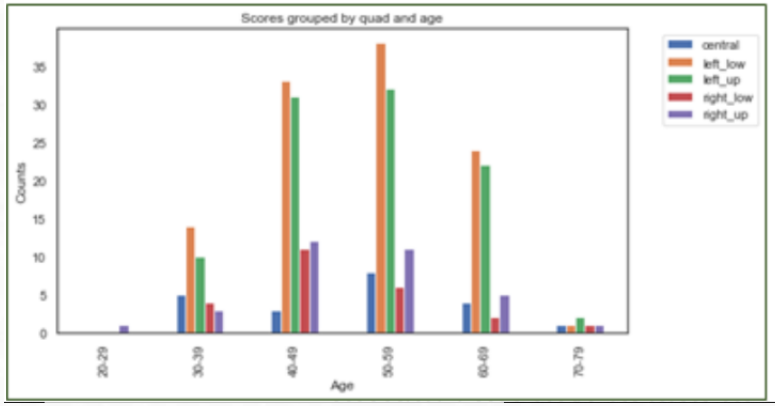
https://towardsdatascience.com/different-bar-charts-in-python-6d984b9c6b17

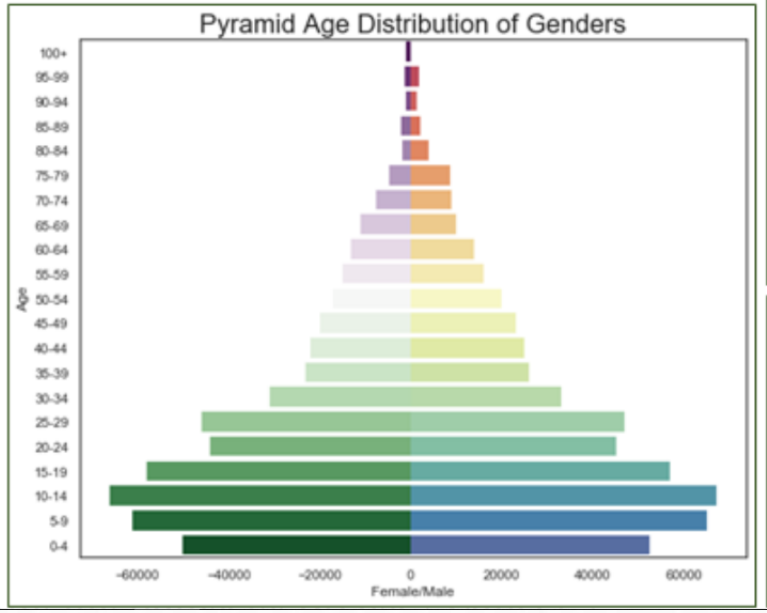


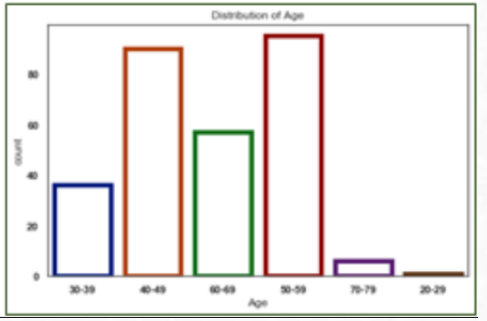










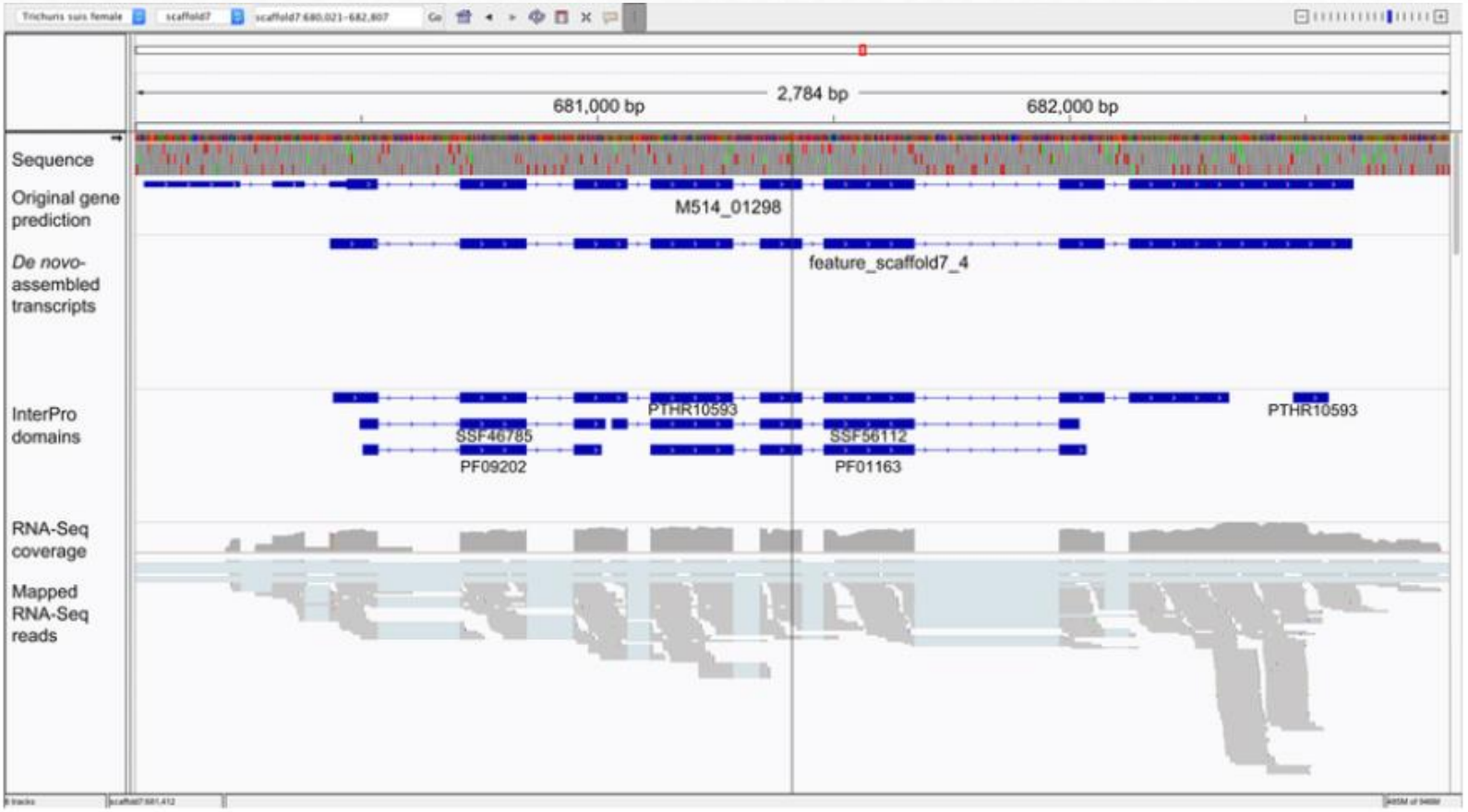


2.10 Visualization

Viewers

UCSC genome browser, **IGV**...

https://pubmed.ncbi.nlm.nih.gov/29717207/



2.10 Visualization

RNA-seq paired and stranded reads

https://bioinformatics.stackexchange.com/questions/2466/interpreting-intergrative-genomic-viewer-igv

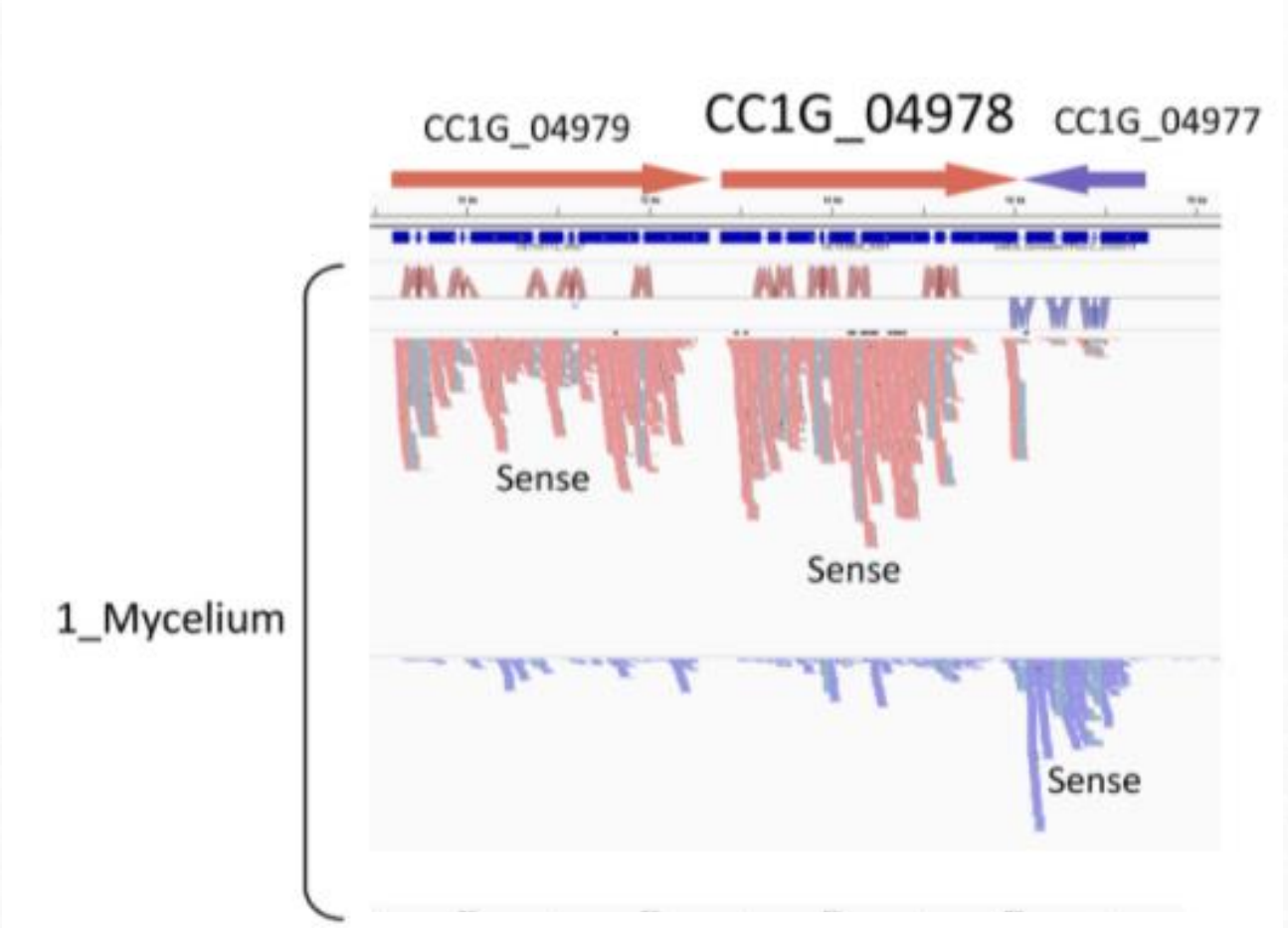
2.10 Visualization

RNA-seq paired and stranded reads

https://bioinformatics.stackexchange.com/questions/2466/interpreting-intergrative-genomic-viewer-igv

2.10 Visualization

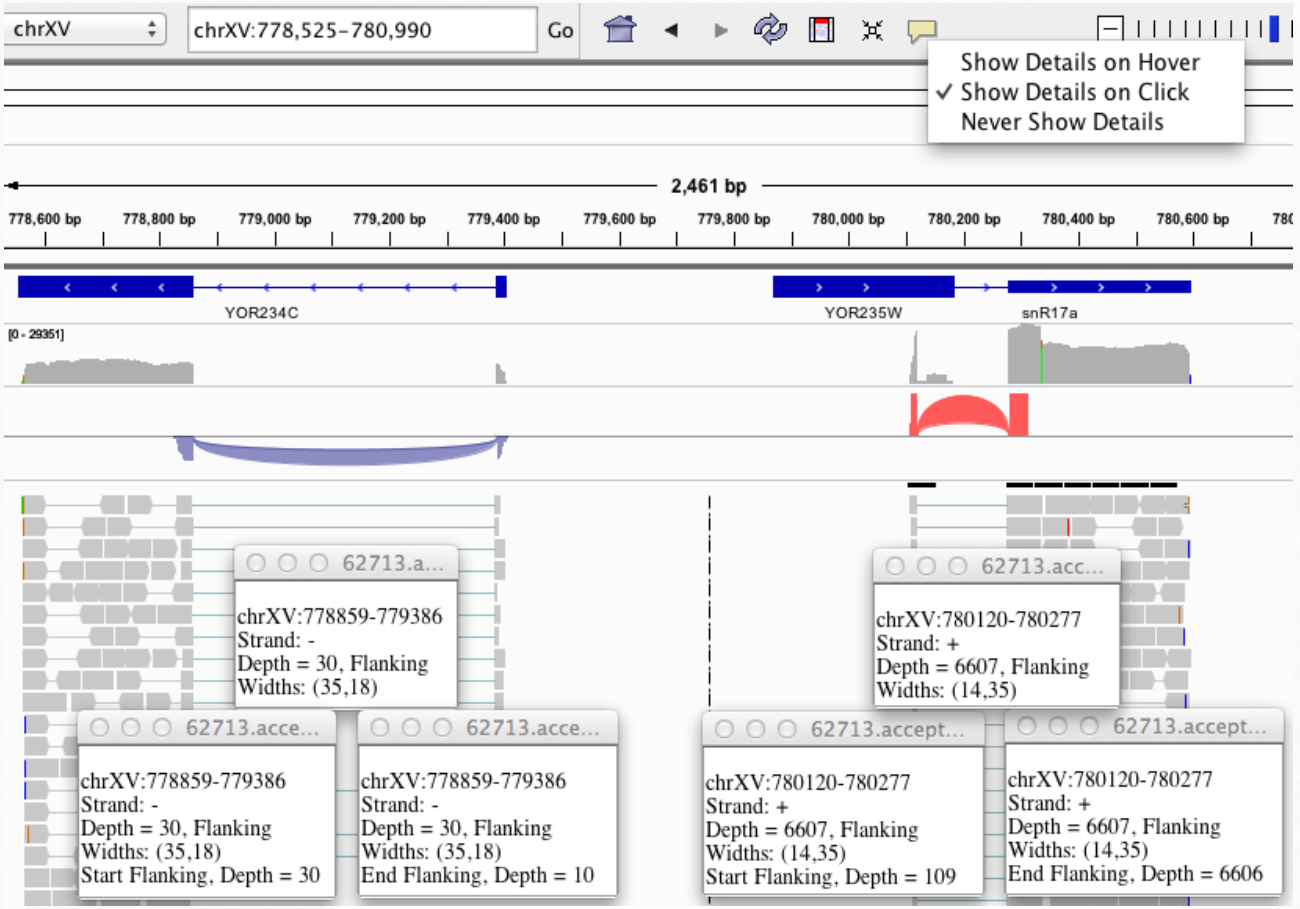
RNA-seq paired and stranded reads



https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0141586.g008

2.10 Visualization

Junctions detected with TopHat



https://software.broadinstitute.org/software/igv/splice_junctions

2.10 Visualization

Iso-Seq RNA-seq (TGS)

https://www.nature.com/articles/s41598-019-42184-z