Lecture 5: RNA-Seq

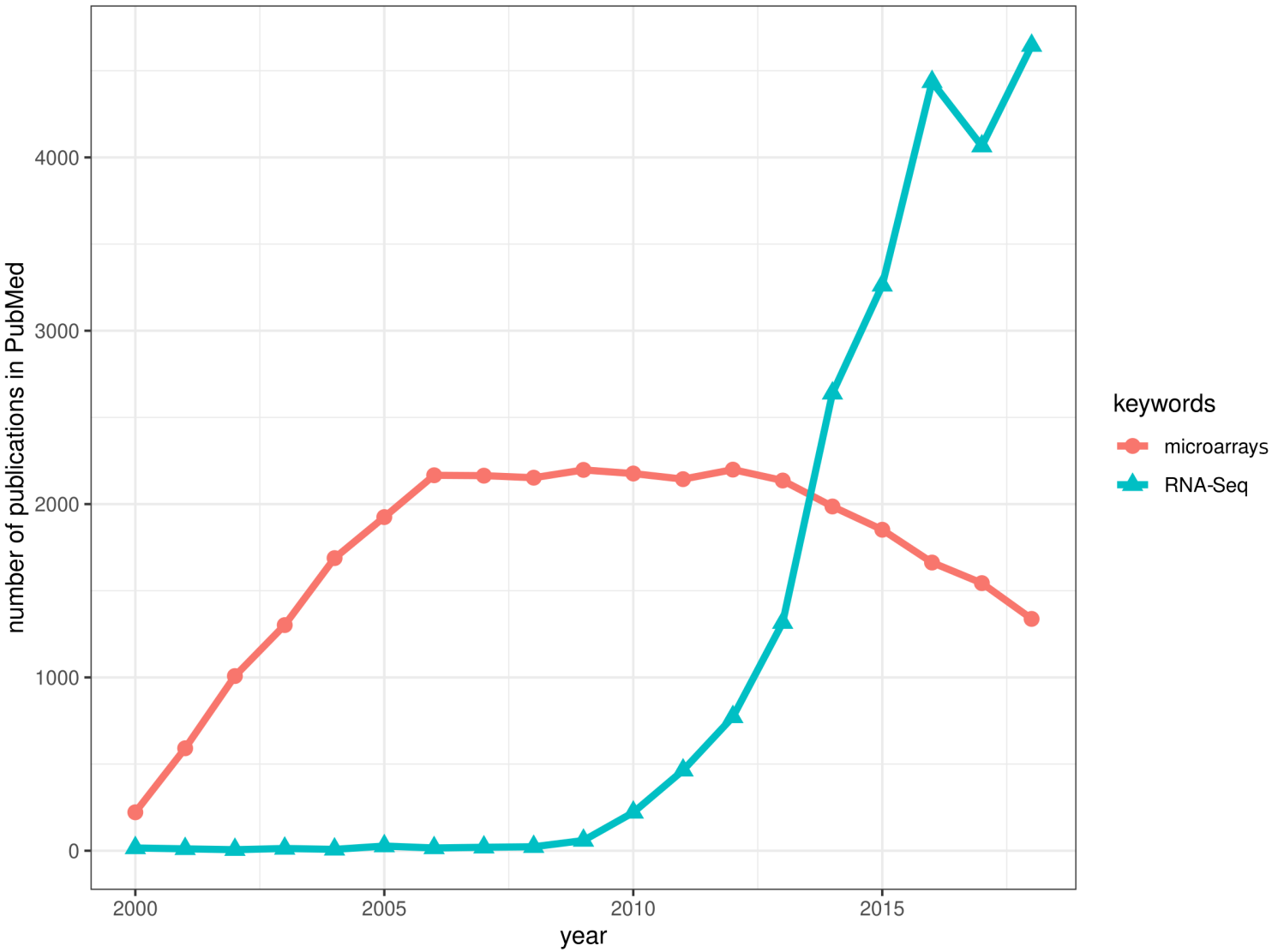
BIOINF3005/7160: Transcriptomics Applications

Zhipeng Qu

School of Biologial Sciences,
The University of Adelaide

April 6th, 2020

Trend of transcriptomics technologies



Alexandru Dan Corlan. Medline trend: automated yearly statistics of PubMed results for any query, 2004.

What is RNA-Seq?

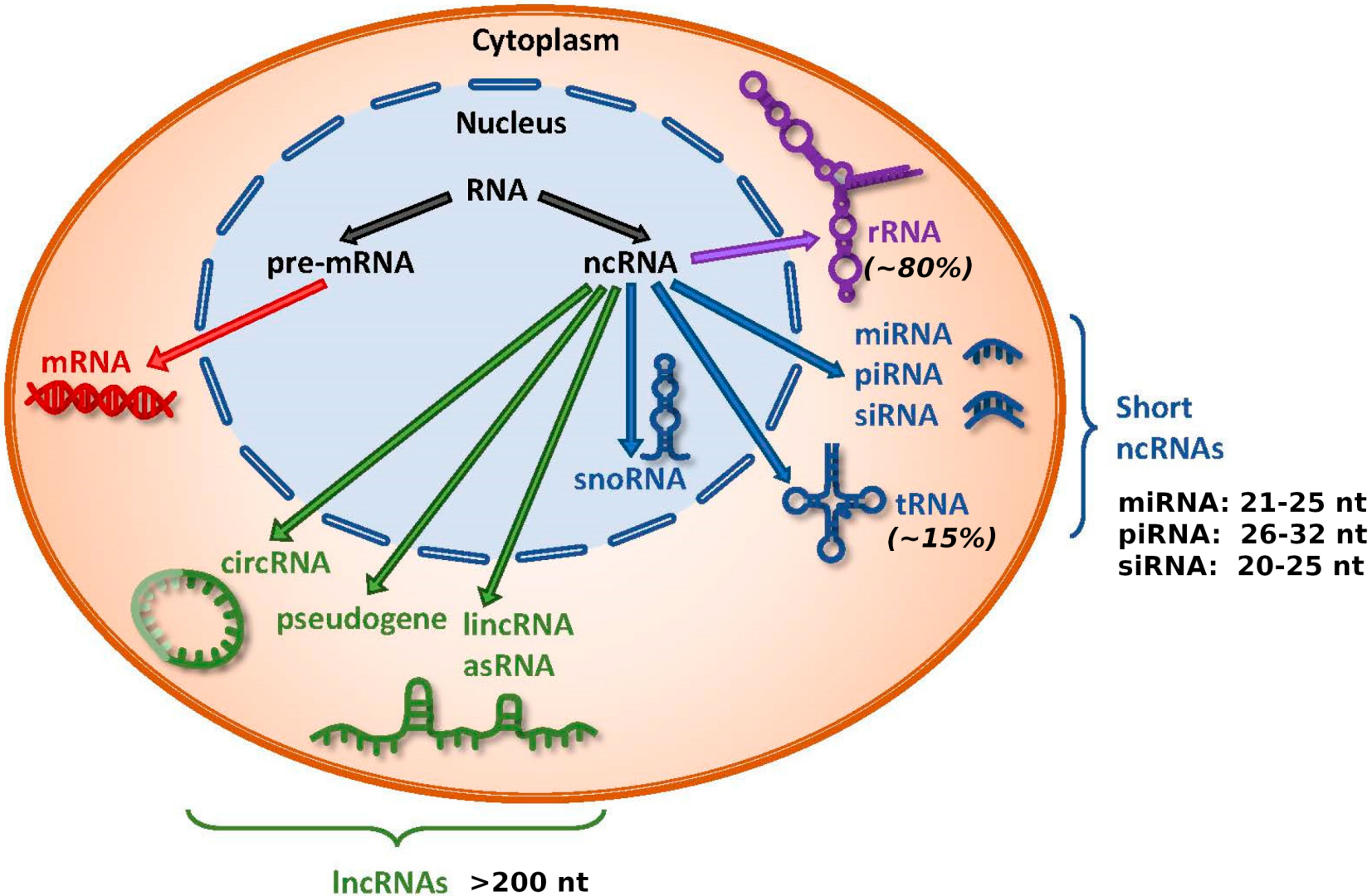
"RNA-Seq, also called RNA sequencing, is a particular technology-based sequencing technique which uses next-generation sequencing (NGS) to reveal the presence and quantity of RNA in a biological sample at a given moment, analyzing the continuously changing cellular transcriptome."

---

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

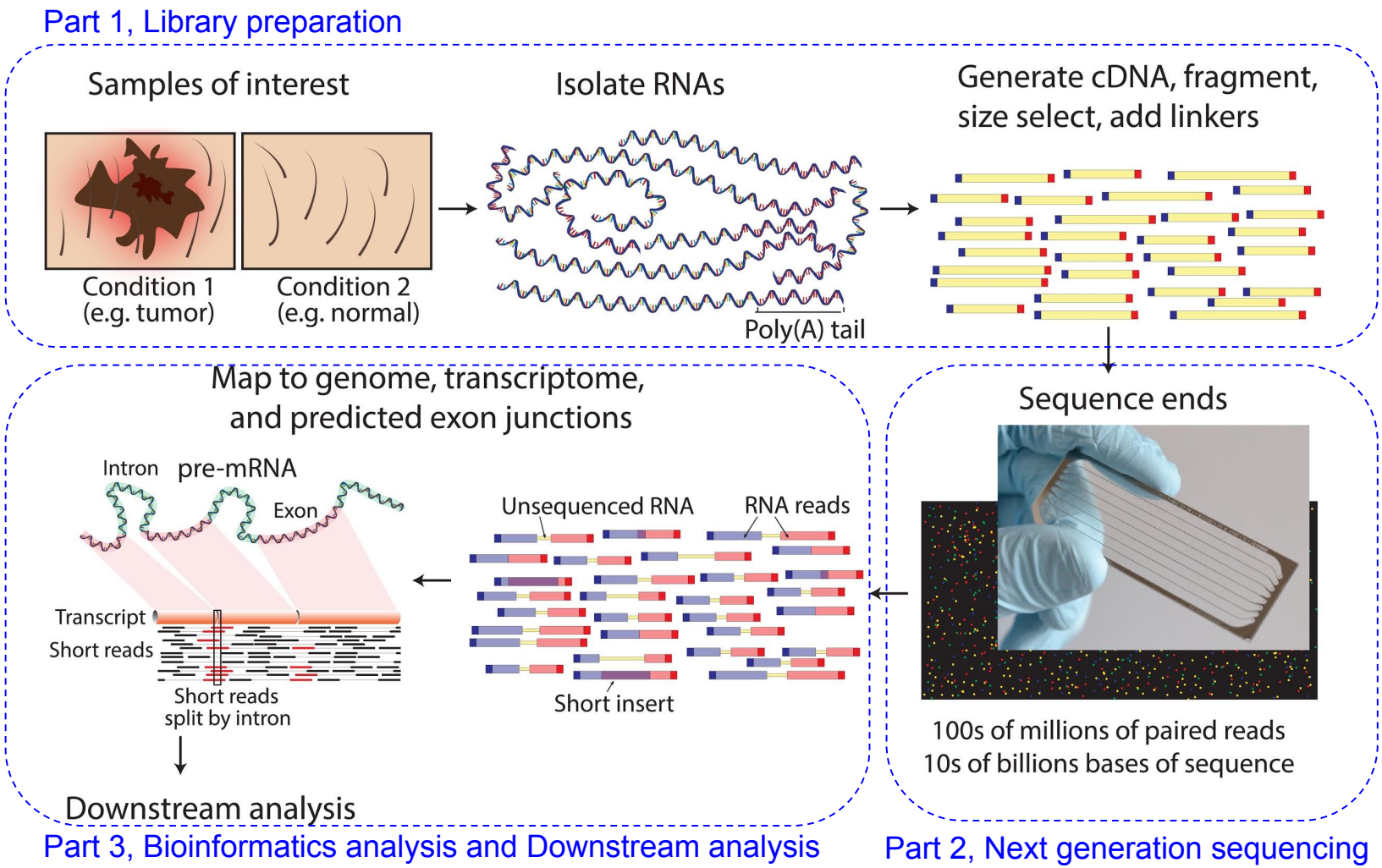
Wang Z, Gerstein M, Snyder M (January 2009). "RNA-Seq: a revolutionary tool for transcriptomics". Nature Reviews. Genetics. 10 (1): 57–63.

Expression of different RNAs in Eukaryotic cell



Chan, J.J.; Tay, Y. Noncoding RNA:RNA Regulatory Networks in Cancer. Int. J. Mol. Sci. 2018, 19, 1310.

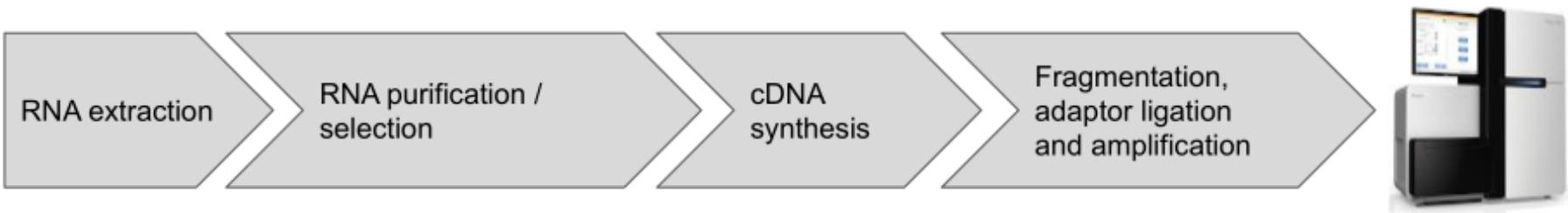
Overview of RNA-Seq



Malachi Griffith*, Jason R. Walker, Nicholas C. Spies, Benjamin J. Ainscough, Obi L. Griffith*. 2015. Informatics for RNA-seq: A web resource for analysis on the cloud. PLoS Comp Biol. 11(8).

Part 1, Library preparation

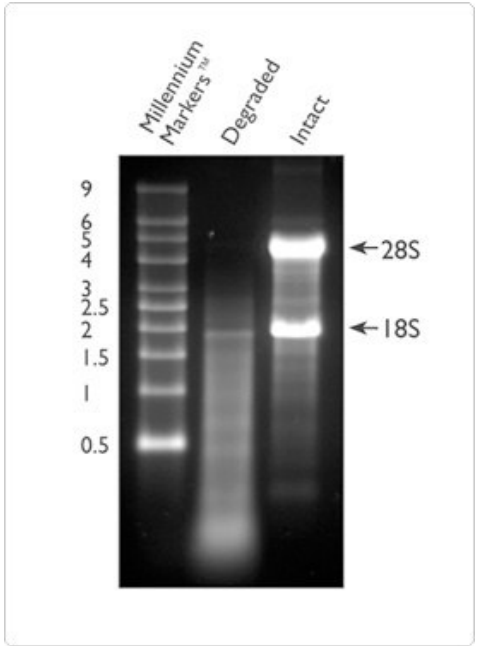
- library preparation flowchart
- RNA quality
- rRNA depletion
- mRNA library
- targeted RNA library
- small RNA library


RNA extraction
RNA purification / selection
cDNA synthesis
Fragmentation, adaptor ligation and amplification

RNA quality

- RNA degradation is a gradual process.
- It can skew measurements of gene expression levels between samples.
- It may affect the purification of small RNAs from total RNAs.

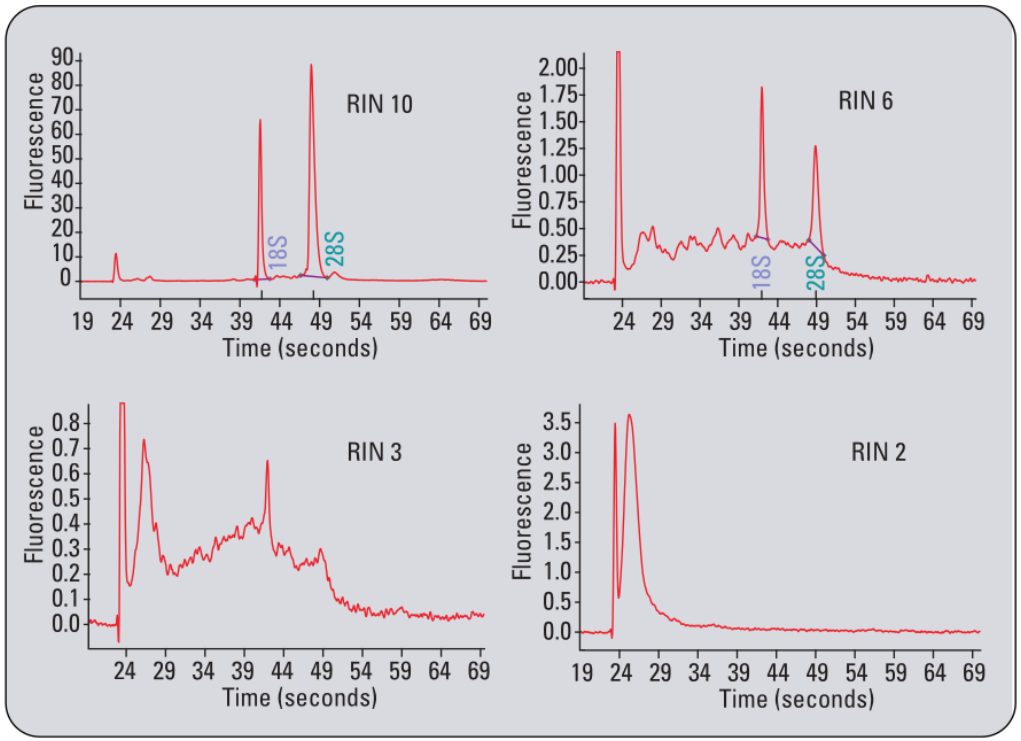
Gallego Romero, I., Pai, A.A., Tung, J. et al. RNA-seq: impact of RNA degradation on transcript quantification. BMC Biol 12, 42 (2014).



Mammalian 28S and 18S rRNAs are approximately 5 kb and 2 kb in size, the theoretical 28S:18S ratio is approximately 2.7:1; but a 2:1 ratio has long been considered the benchmark for intact RNA. mRNA quality has historically been assessed by electrophoresis of total RNA followed by staining with ethidium bromide.

---

https://www.thermofisher.com/au/en/home/references/ambion-tech-support/rna-isolation/tech-notes/assessing-rna-quality.html



RNA integrity number (RIN) was introduced by Agilent, which takes the entire electrophoretic trace into account. It scores the RNA integrity based on a numbering system from 1 to 10, with 1 being the most degraded profile and 10 being the most intact.

---

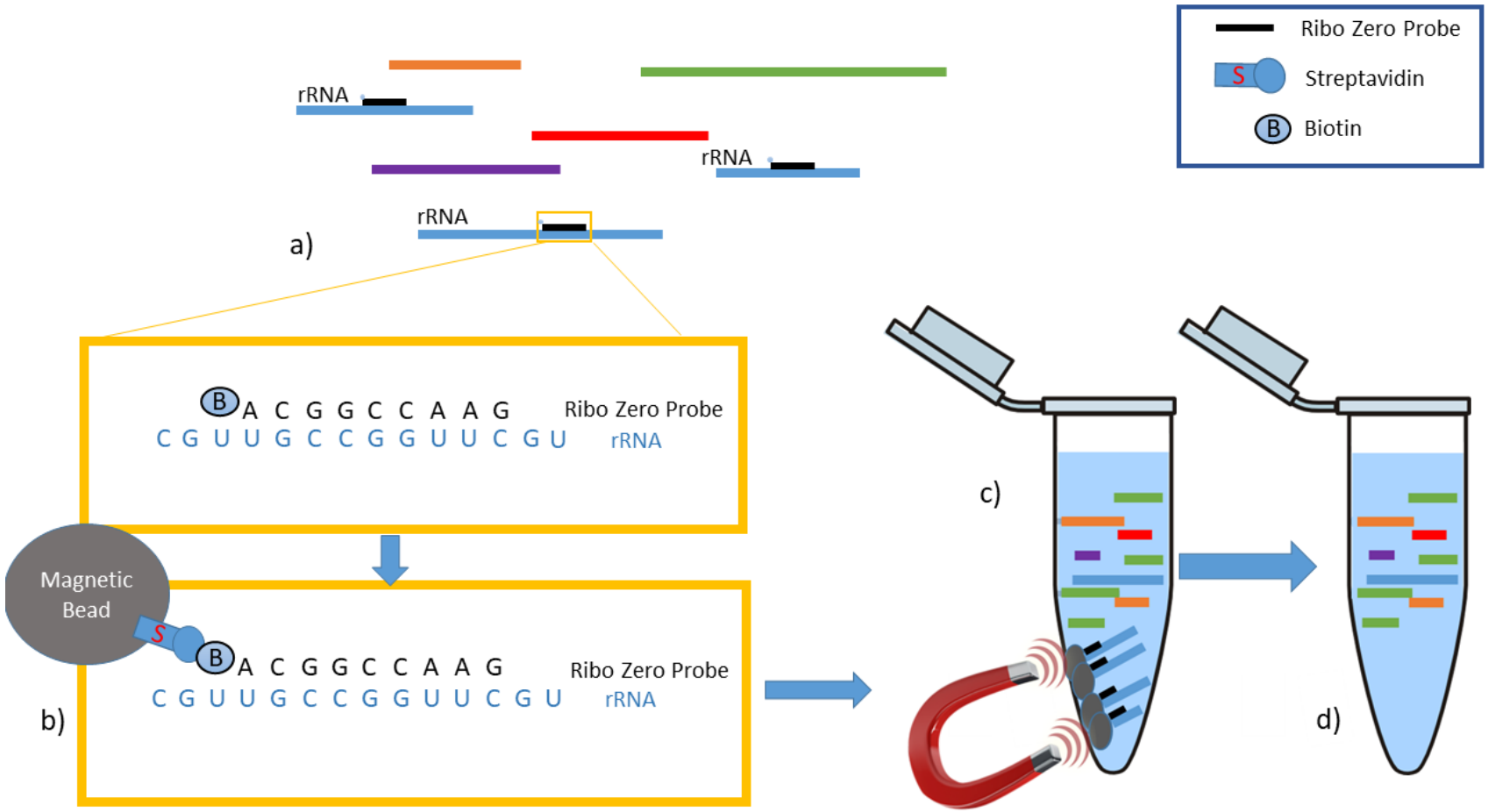
https://www.agilent.com

rRNA depletion

Three strategies to remove ribosomal RNA (rRNA) from total RNAs:

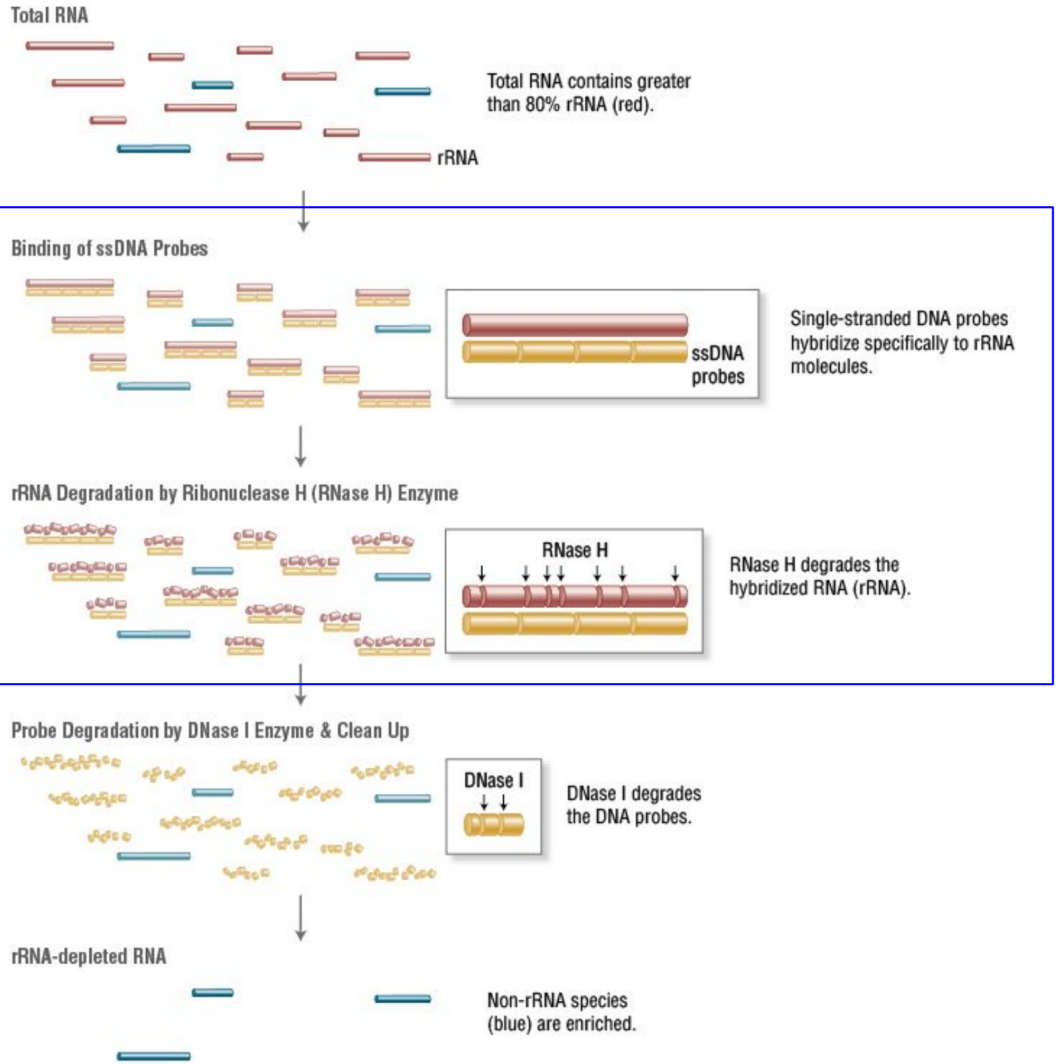
- based on complimentary oligonucleotides coupled to paramagnetic beads
  - Illumina's RiboZero, Qiagen GeneRead rRNA depletion, and Lexogen RiboCop
- based on hybridizing the rRNA to DNA oligos and degrading the RNA:DNA hybrids using RNase H.
  - NEBNext rRNA depletion, Kapa RiboErase, and Takara/Clontech's RiboGone
- targeting the rRNA sequences after conversion to cDNA
  - aimed at low input samples
  - Takara/Clontech SMARTer Pico kit

Herbert, Z.T., Kershner, J.P., Butty, V.L. et al. Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. BMC Genomics 19, 199 (2018).



illumina RiboZero rRNA depletion

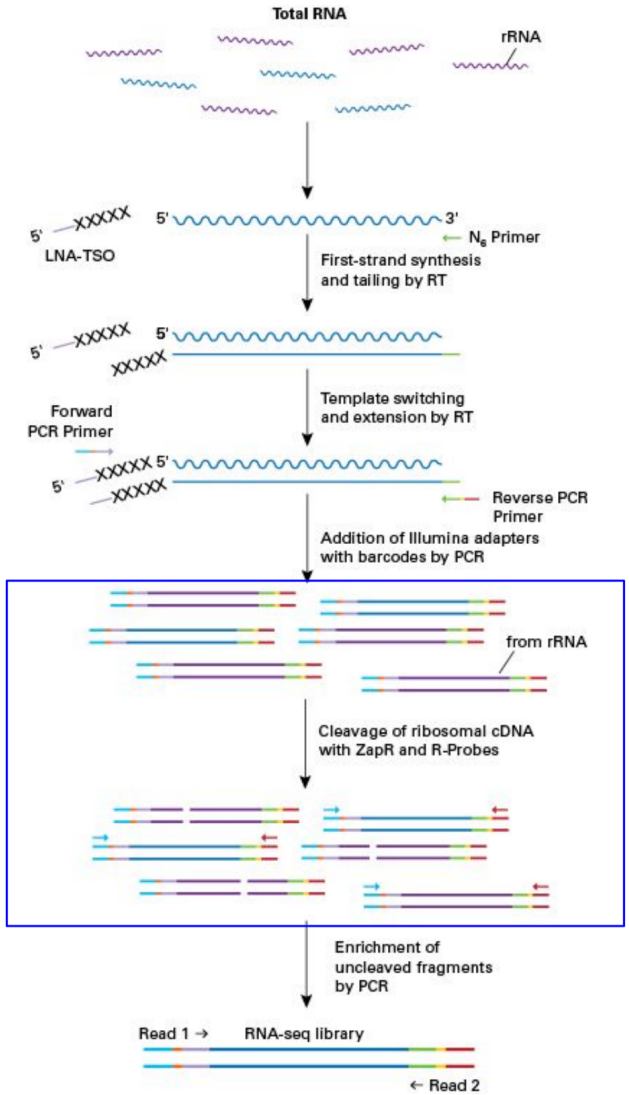
https://sapac.support.illumina.com/bulletins/2019/10/best-practices-to-minimize-rrna-contamination-in-truseq-stranded.html



NEBNext rRNA depletion

---

https://www.neb.com

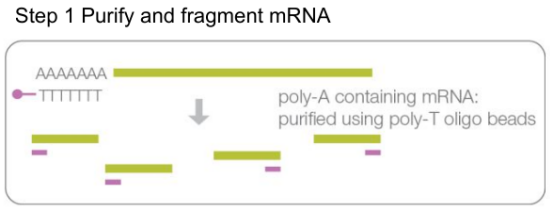


Takara SMARTer Pico kit

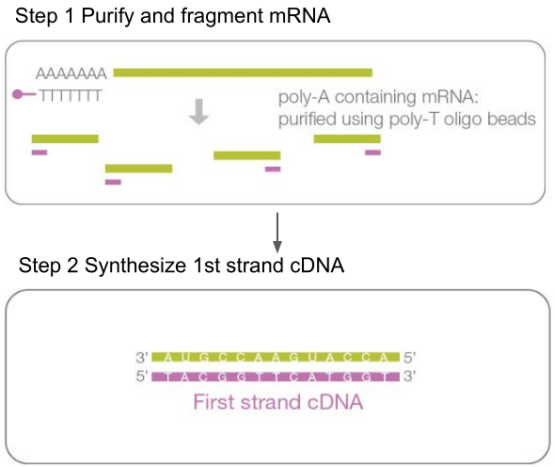
---

https://www.takarabio.com

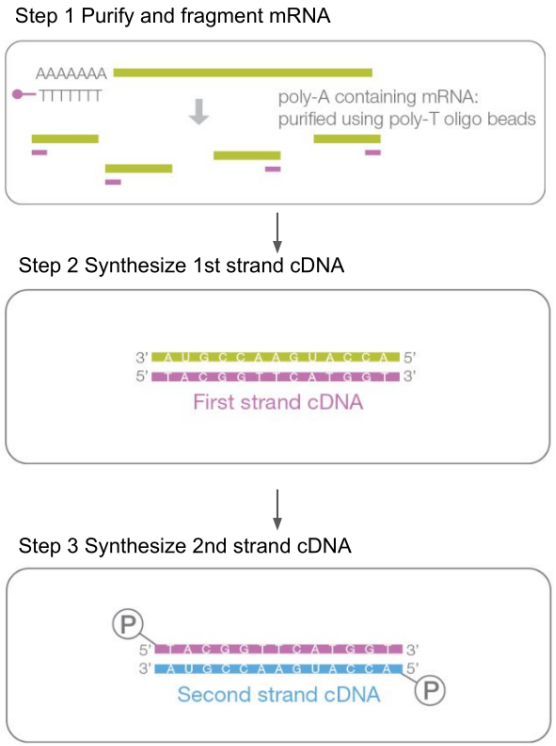
mRNA library



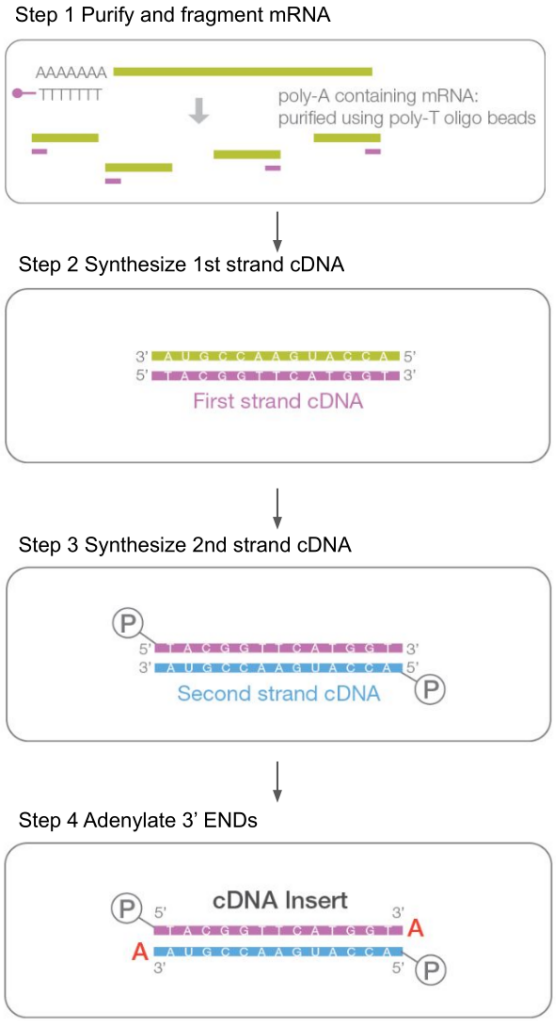
Step 1: The Poly-A containing mRNA molecules are purified using poly-T oligo attached magnetic beads. Following purification,the mRNA is fragmented into small pieces using divalent cations under elevated temperature.



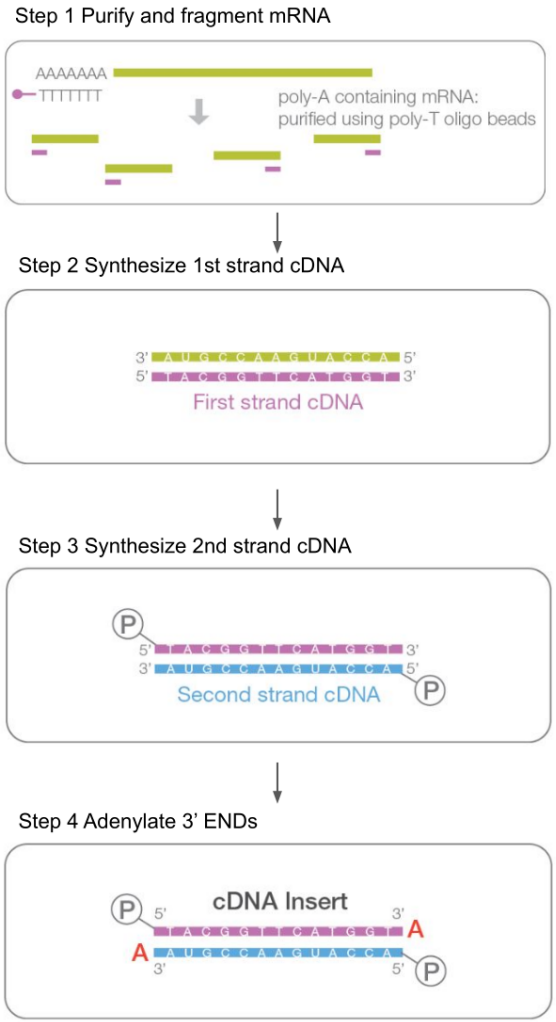
Step 2: Cleaved RNA fragments are copied into first strand cDNA using reverse transcriptase and random primers.

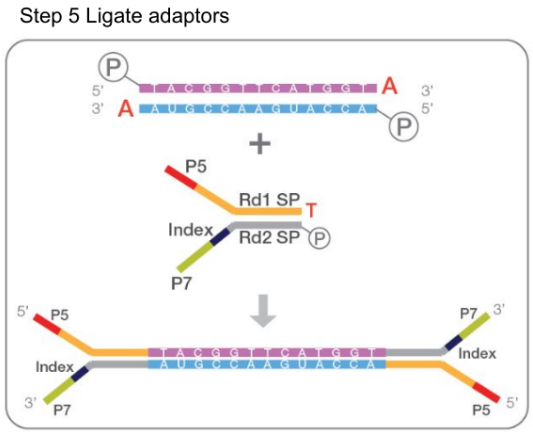


Step 3: Second strand cDNAs are synthesized using DNA PolymeraseI and RNase H.

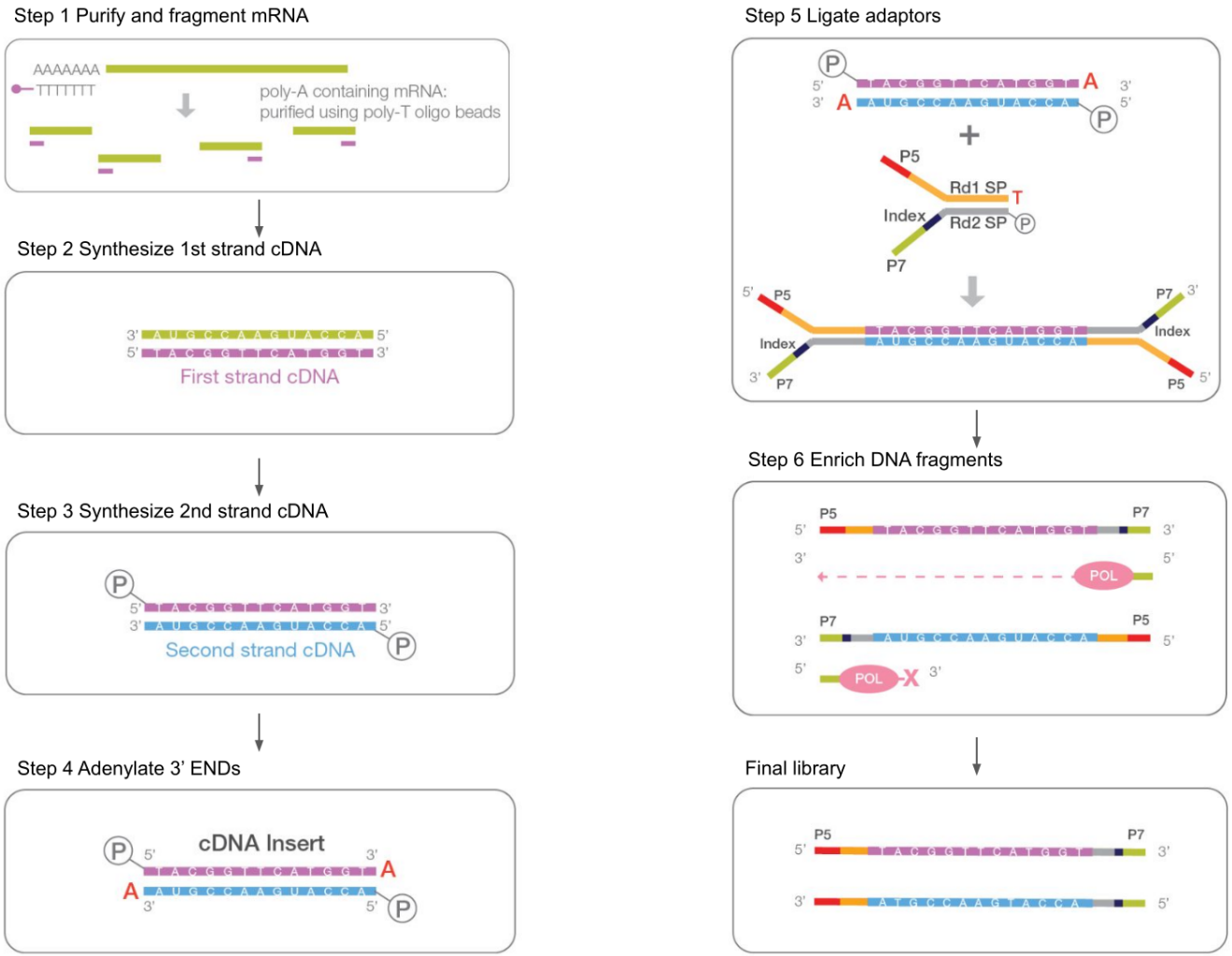


Step 4: A single 'A' nucleotide is added to the 3' ends of the blunt fragments to prevent them from ligating to each other during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragment.





Step 5: Adapter ligation prepares the dscDNA for hybridization onto a flow cell.

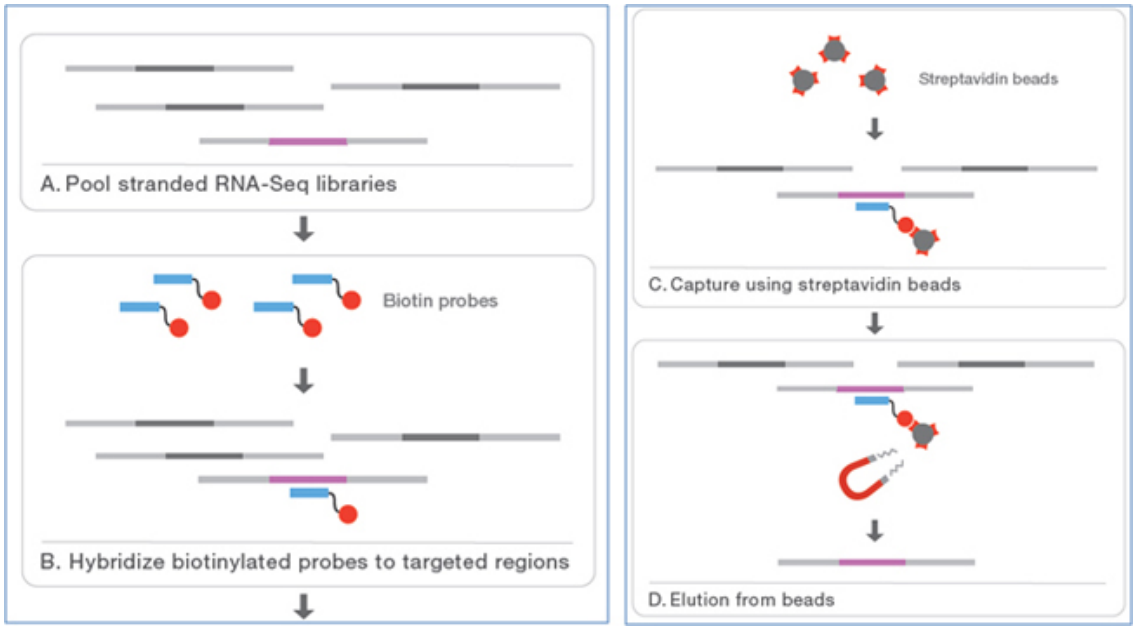


Step 6: DNA fragments are enriched with PCR and purified to create the final cDNA library.

---

https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-stranded-mrna.html

targeted RNA library

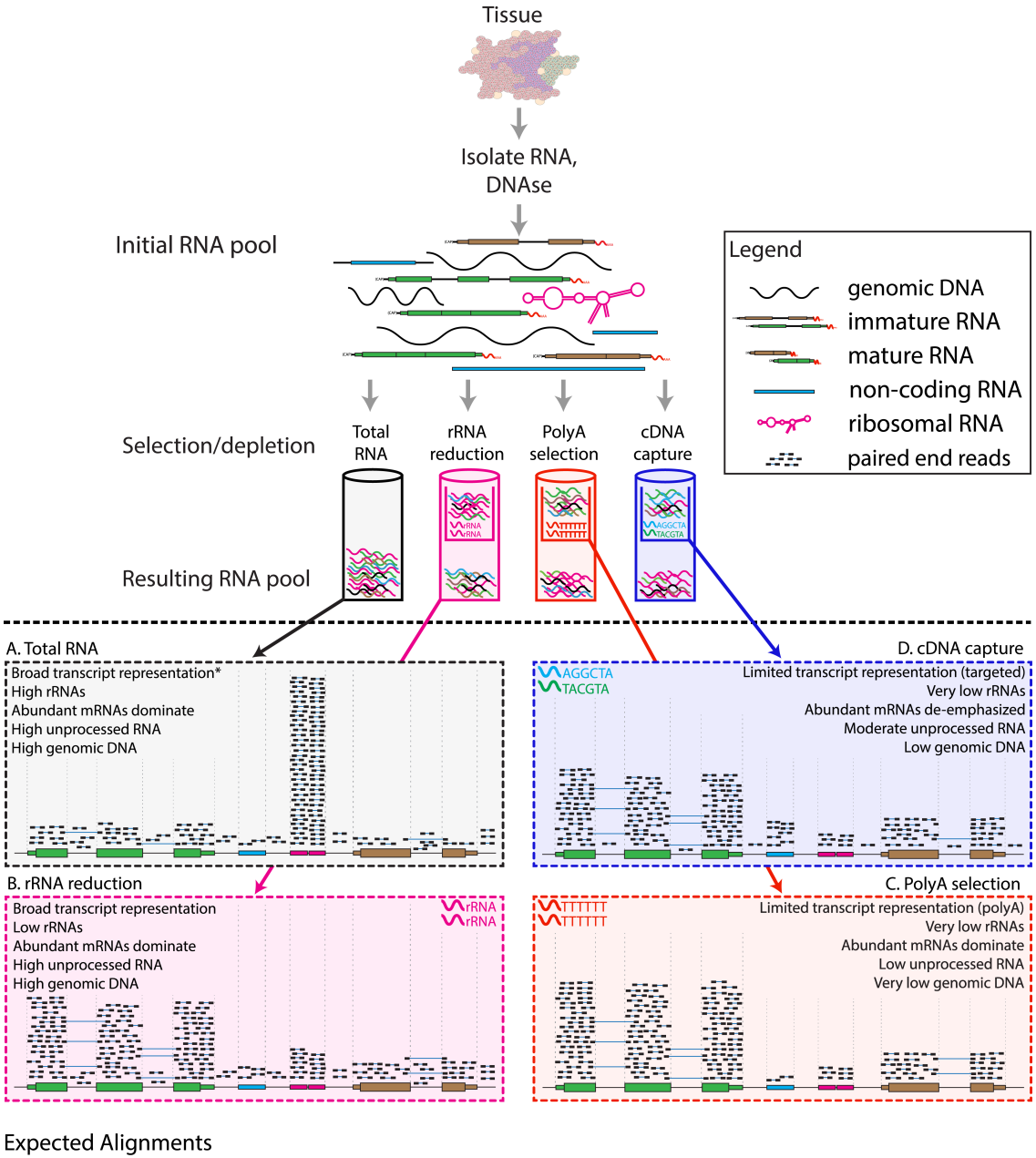


- saving costs
- improving coverage and sensitivity of detection of transcripts of interest
- simplifying analysis

---

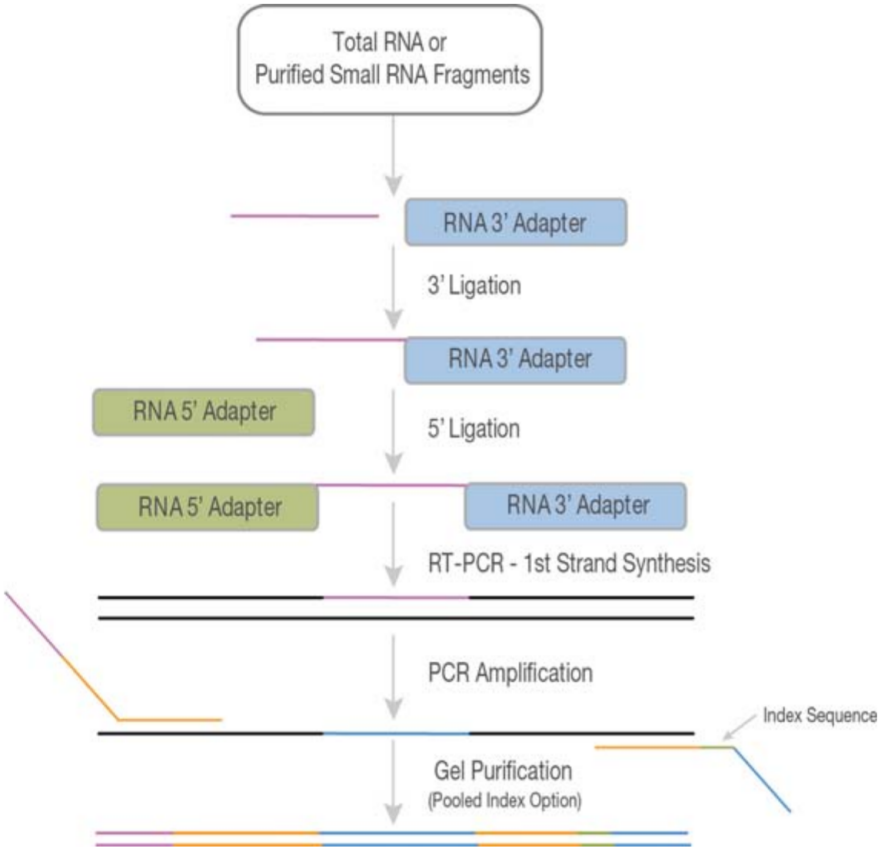
https://sapac.illumina.com/techniques/sequencing/rna-sequencing/targeted-rna-seq.html

Summary of long RNA libraries



Griffith M, Walker JR, Spies NC, Ainscough BJ, Griffith OL (2015) Informatics for RNA-Sequencing: A Web Resource for Analysis on the Cloud. PLOS Computational Biology 11(8): e1004393.

small RNA library



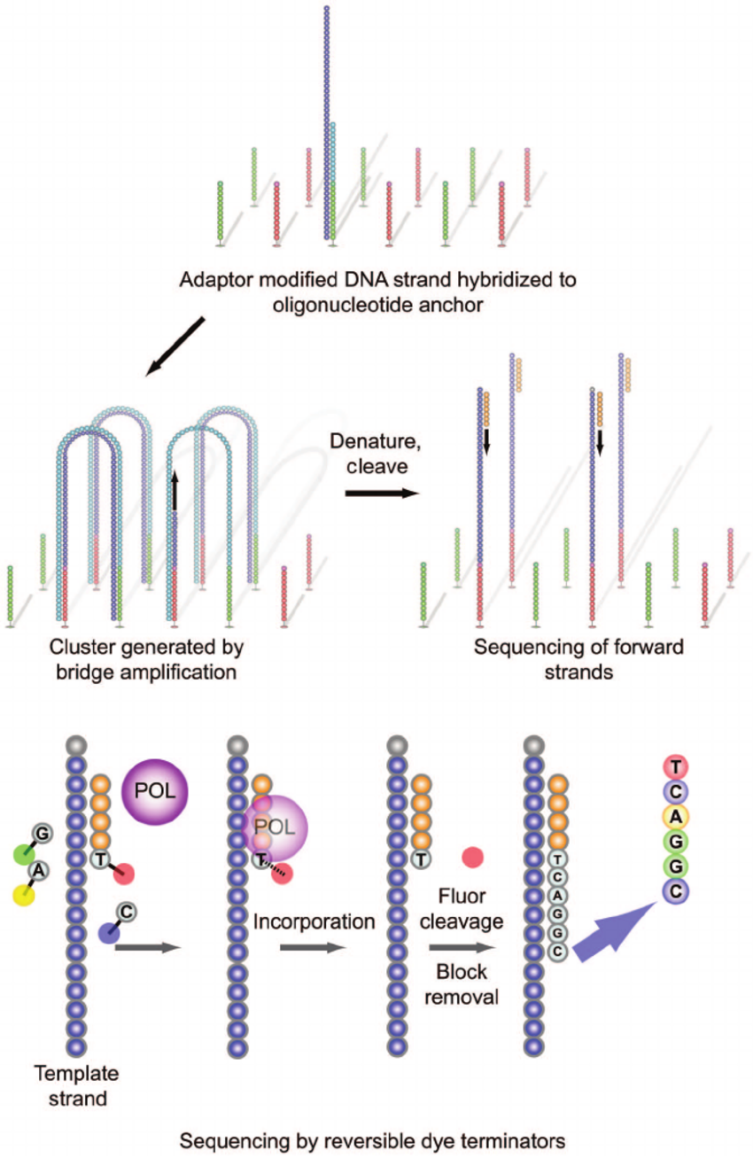
The RNA 3' adapter is specifically modified to target microRNAs and other small RNAs that have a 3' hydroxyl group resulting from enzymatic cleavage by Dicer or other RNA processing enzymes.

---

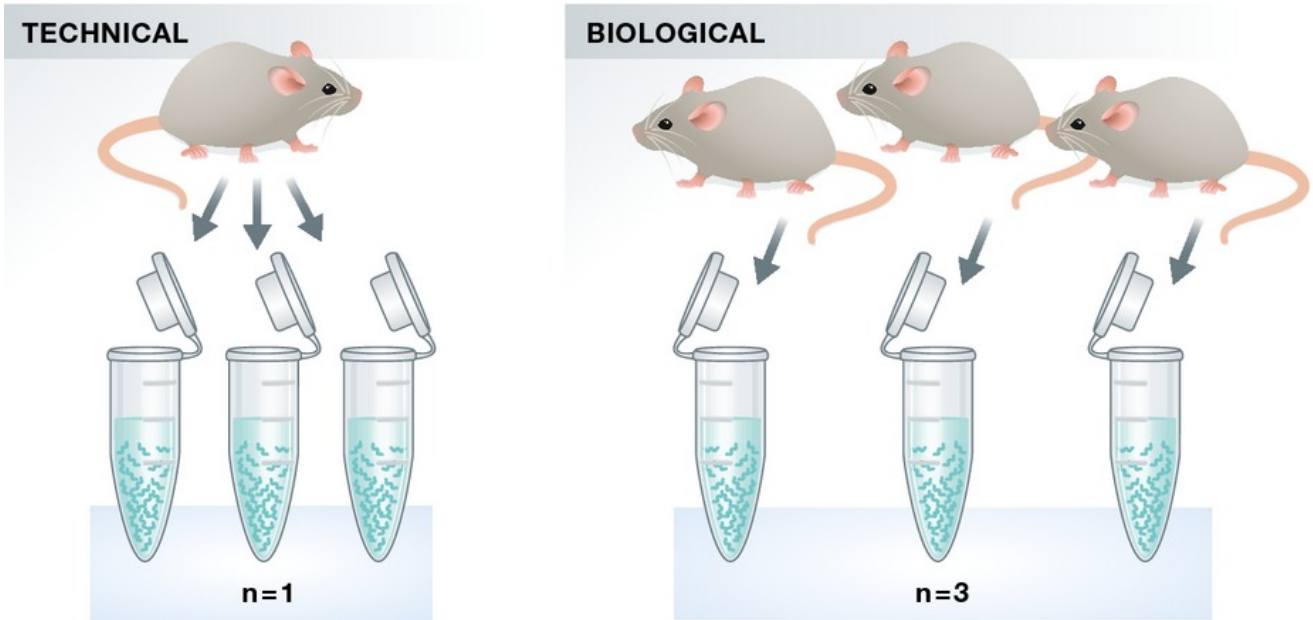
https://sapac.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-small-rna.html

Part 2, Next generation sequencing

- illumina sequencing
- technical replicates VS biological replicates
- paired end VS single end
- sequencing depth in RNA-Seq
- read length in RNA-Seq

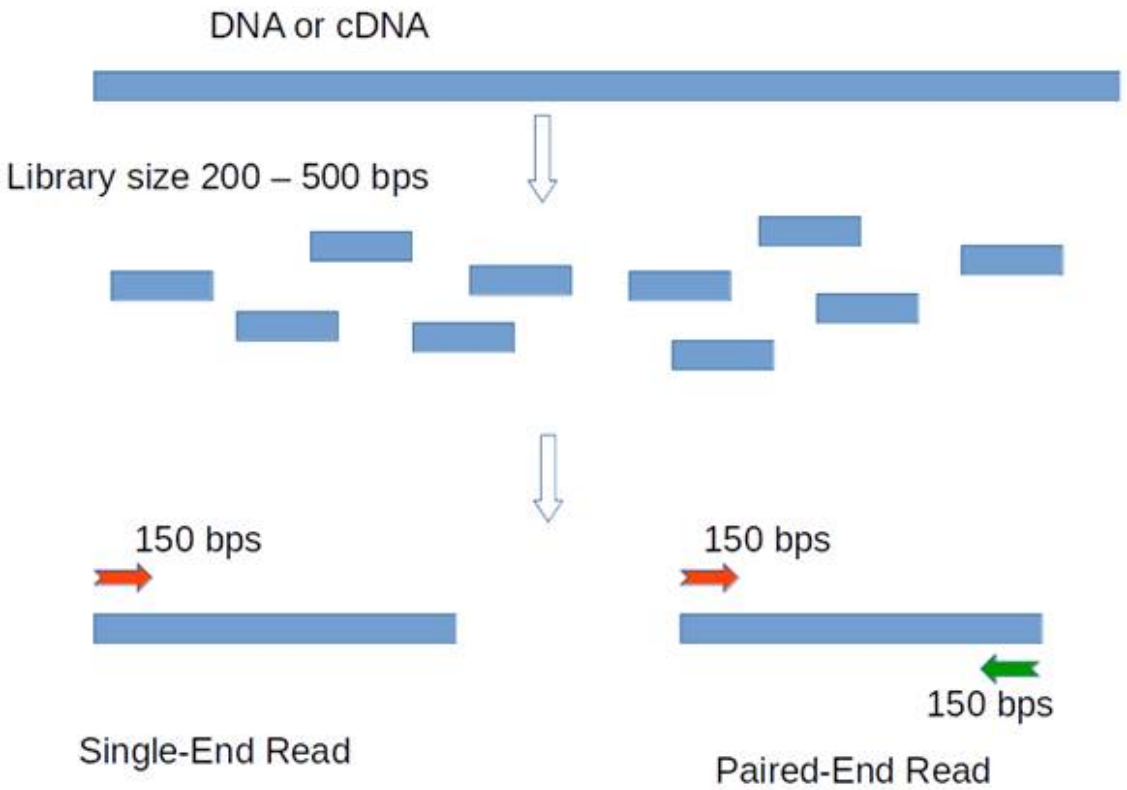


Karl V Voelkerding, Shale A Dames, Jacob D Durtschi, Next-Generation Sequencing: From Basic Research to Diagnostics, Clinical Chemistry, Volume 55, Issue 4, 1 April 2009.



- **Technical replicates** use the same biological sample to repeat the technical or experimental steps in order to accurately measure technical variation and remove it during analysis.
- **Biological replicates** use different biological samples of the same condition to measure the biological variation between samples.

Klaus, B. (2015), Statistical relevance–relevant statistics, part I. EMBO J, 34: 2727-2730.



- Single End (SE): only one end of each cDNA fragment is sequenced.
- Paired End (PE): both ends of each cDNA fragment are sequenced. Sequencing reads are labeled as pairs.

Sequencing depth in RNA-Seq is the number of reads in each sample. It varies depending on the goal of the study.

- Normal gene expression profiling experiments: 5–25 million reads per sample
- Experiments looking for a more global view of gene expression, or alternative splicing: 30–60 million reads per sample
- In-depth view of transcriptome, or assemble new transcripts: 100–200 million reads per sample
- Targeted RNA-Seq: 3 million reads per sample
- Small RNA-Seq: 1–5 million reads per sample

---

https://sapac.support.illumina.com/bulletins/2017/04/considerations-for-rna-seq-read-length-and-coverage-.html

Read length will depend on the application and final size of the library.

- Gene expression profiling: SE50–SE75
- Transcriptome annotation or assembly: longer, paired-end reads (such as 2 x 75 bp) to enable more complete coverage of the transcripts and identification of novel variants or splice sites
- small RNA-Seq: SE50

---

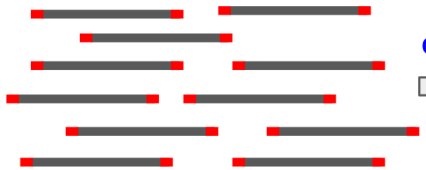
https://sapac.support.illumina.com/bulletins/2017/04/considerations-for-rna-seq-read-length-and-coverage-.html

Part 3, Bioinformatics analysis and Downstream analysis

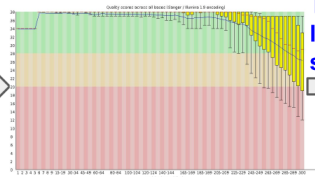
- differential gene expression analysis
- transcriptome assmbely
- alternative splicing
- non-coding RNAs (ncRNAs)

Base calling

Raw reads

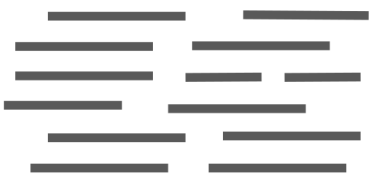


QC

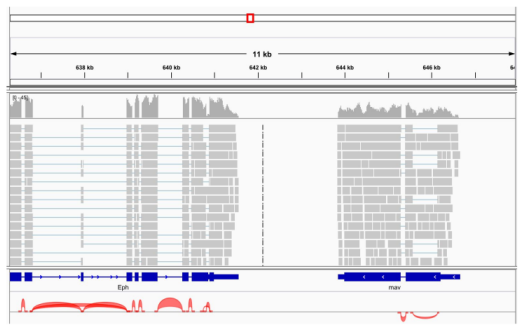


Trim adaptors and low quality sequences

Clean reads

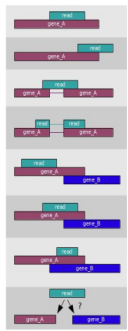


Genome mapping

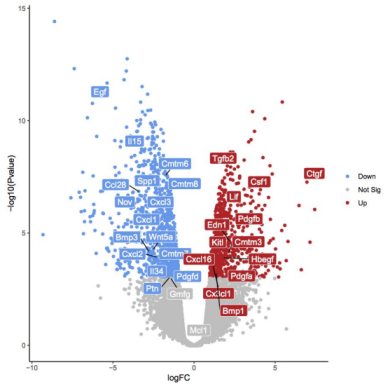


Feature count

	Sample1	Sample2	...
Gene 1	56	209	...
Gene 2	89	10	...
Gene 3	0	503	...
Gene 4	20	200	...
Gene 5	2	20	...
...	...	...	...

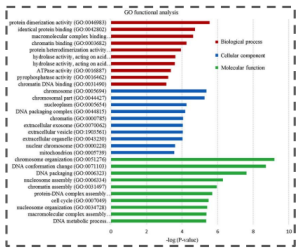


DE gene identification

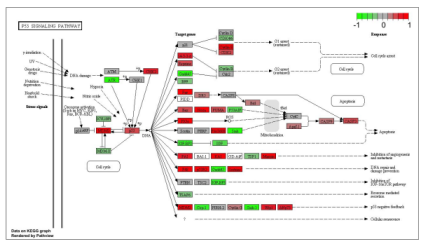


Functional analysis

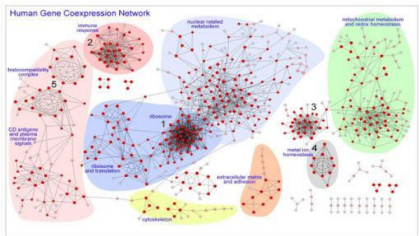
GO enrichment analysis



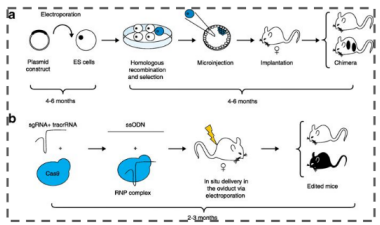
KEGG functional annotation

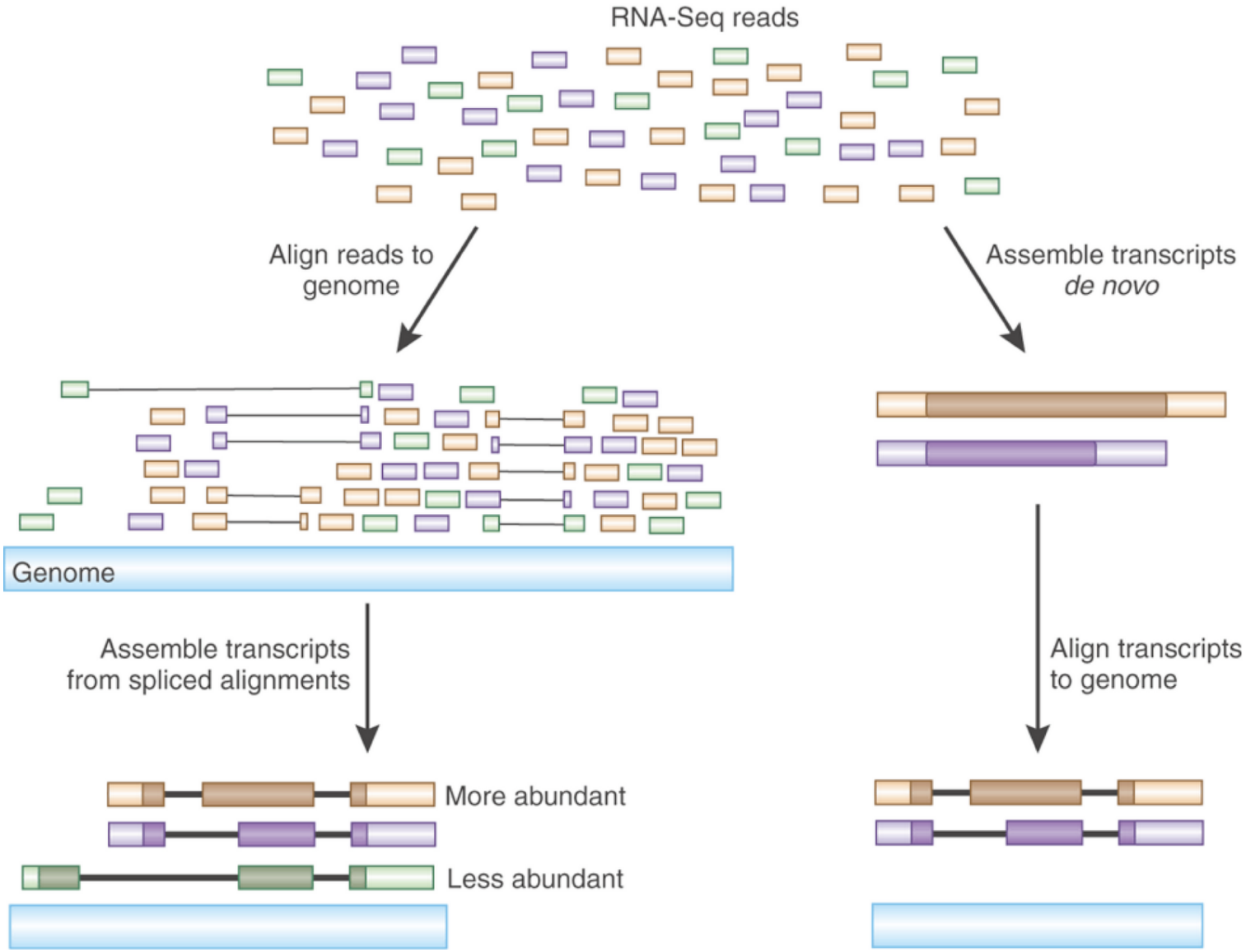


Co-expression networks

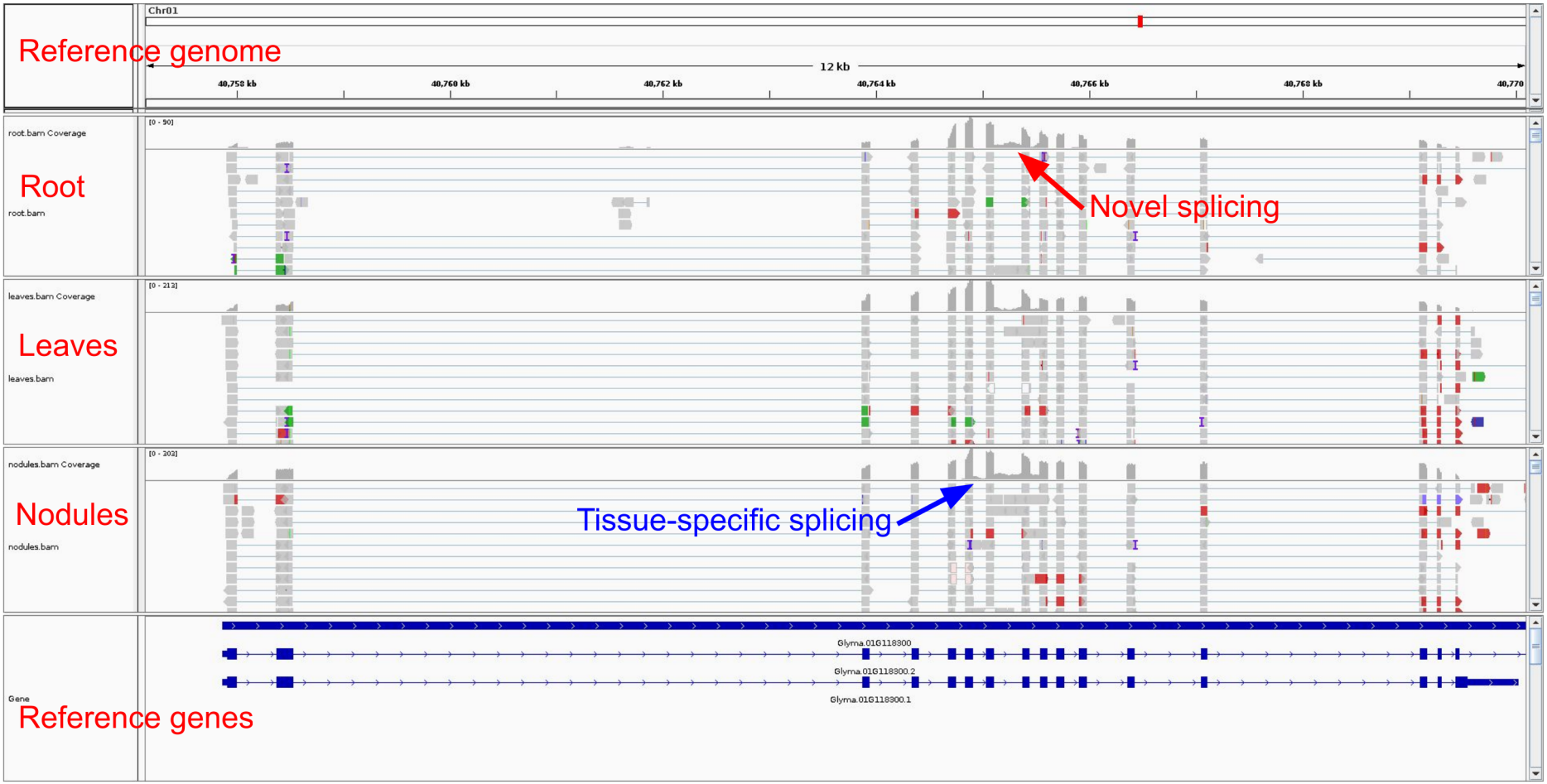


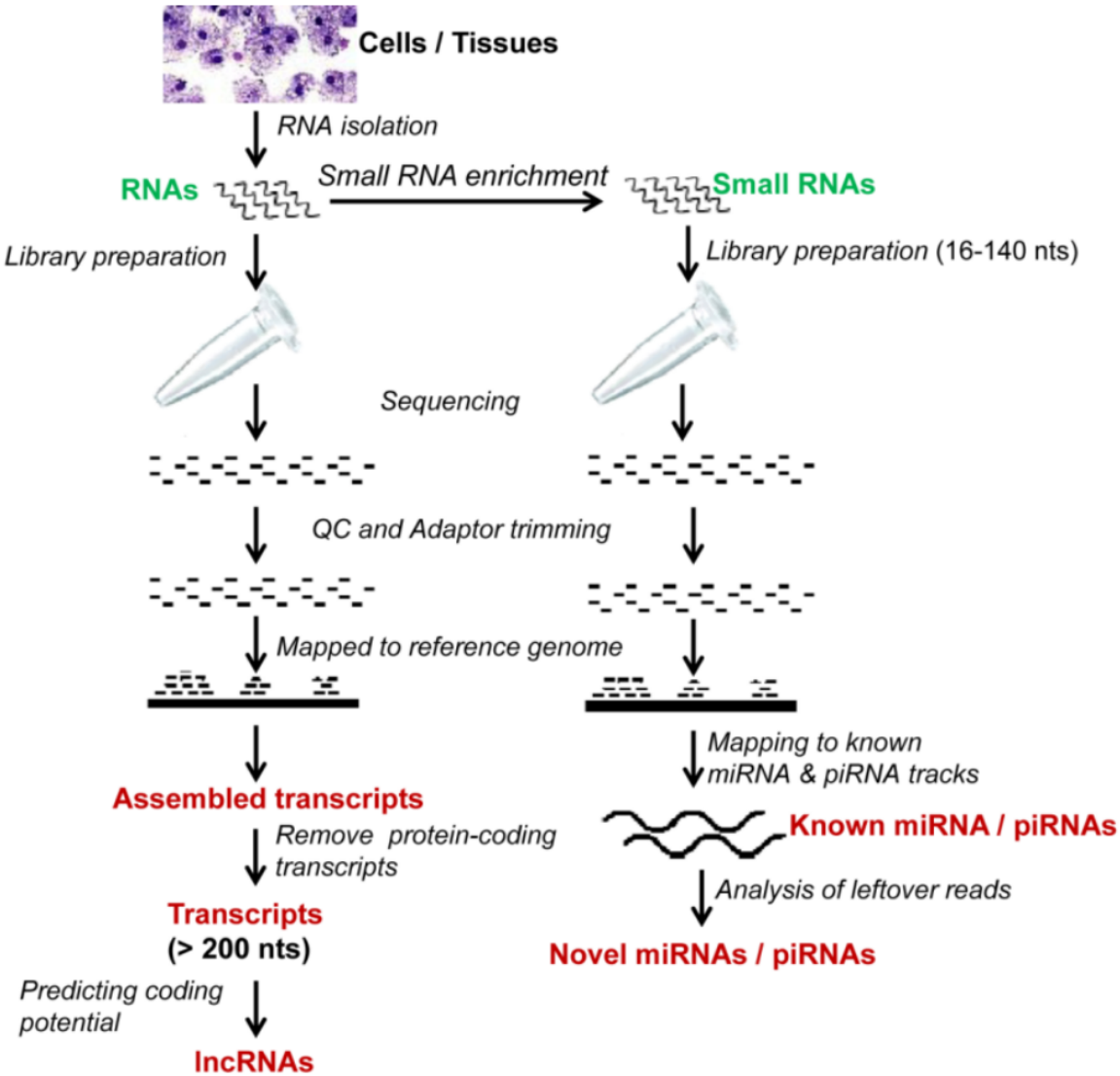
Functional validation





Haas, B., Zody, M. Advancing RNA-Seq analysis. Nat Biotechnol 28, 421–423 (2010).


Reference genome
Chr01
12 kb
40,758 kb
40,760 kb
40,762 kb
40,764 kb
40,766 kb
40,768 kb
40,770
root.bam Coverage
[0 - 90]
Root
root.bam
Novel splicing
leaves.bam Coverage
[0 - 213]
Leaves
leaves.bam
nodules.bam Coverage
[0 - 303]
Nodules
nodules.bam
Tissue-specific splicing
Gene
Glyma.01G118300
Glyma.01G118300.2
Glyma.01G118300.1
Reference genes



Mallick (2016) Decrypting the Treasures of Regulatory Non-coding RNAs in High-throughput Era. J Data Mining Genomics and Proteomics 7: e124.

RNA-Seq in a nutshell