



CEPLAS

Cluster of Excellence on Plant Sciences

# Leaf to library – Experimental considerations

CEPLAS RNA-Seq Workshop 2022



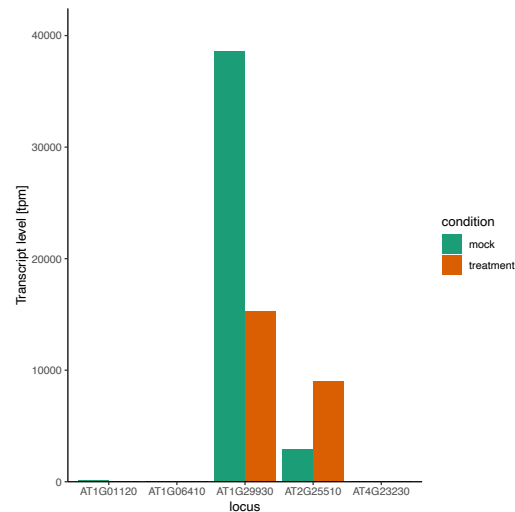
- Alternatives to RNA-Seq
- qPCR, microarray
  - Dynamic range (highly/lowly expressed genes)
  - need to know target (sequences)



# Main applications of RNA-Seq (~Transcriptomics)

## (m)RNA quantification

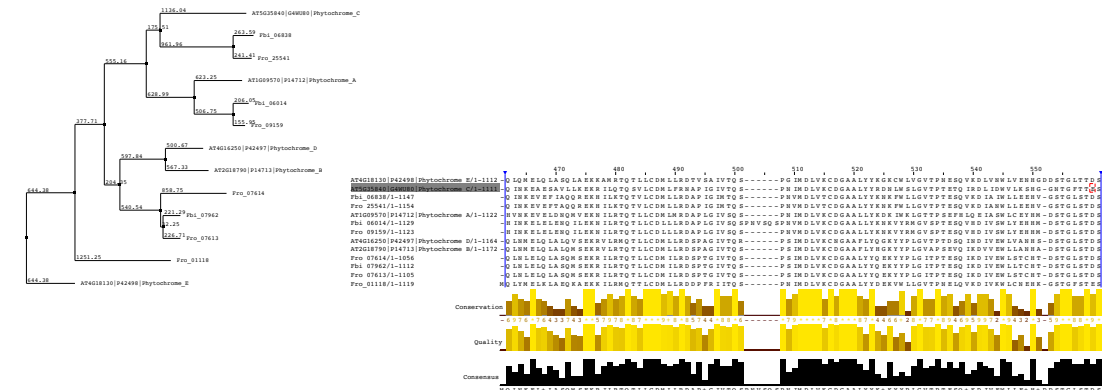
- Count occurrences of specific mRNAs
- “Gene Expression”
- Differential gene expression (between samples)
- ...



Days 1 & 2

## Transcript discovery

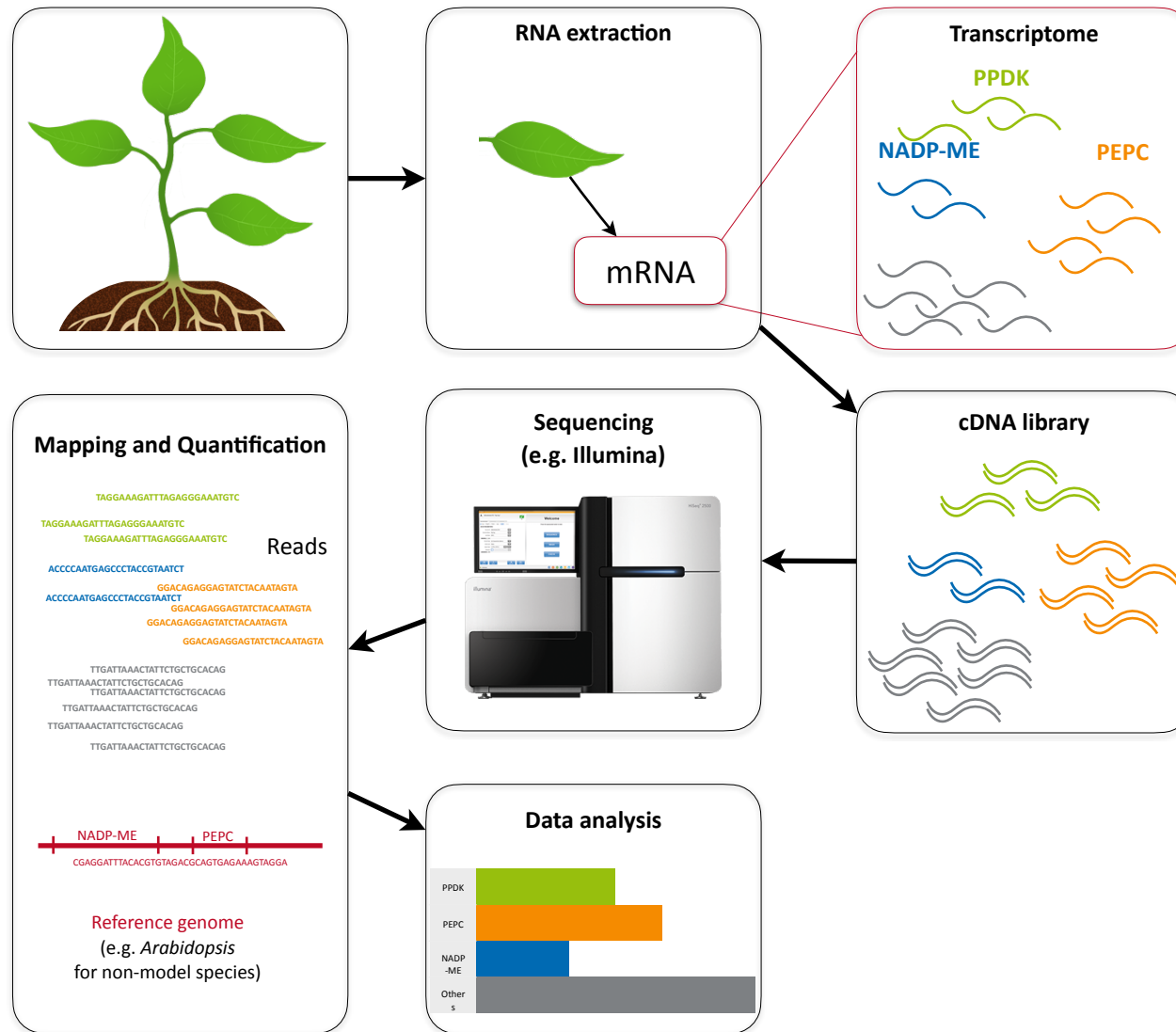
- Isoforms and alternative splicing
- Single nucleotide variations
- Non-coding RNAs
- ...



Day 3



# mRNA-Sequencing



Transcript levels ~ proxy for protein abundance

Photo: [www.illumina.com](http://www.illumina.com)



- (Still) expensive
- (Still) a lot of work
- You see virtually everything





- Clear hypothesis **before** starting the experiment
- Wet-lab and sampling
  - sampling in the wild, in the greenhouse, in a growth chamber (logistics!)
  - sampling conditions - time of day, age, development, tissue, light, etc.
  - storage of your samples (logistics!)
- Additional assays / experiments?
  - collect as much data as possible (ideally from the same sample set)
  - this will strengthen your study

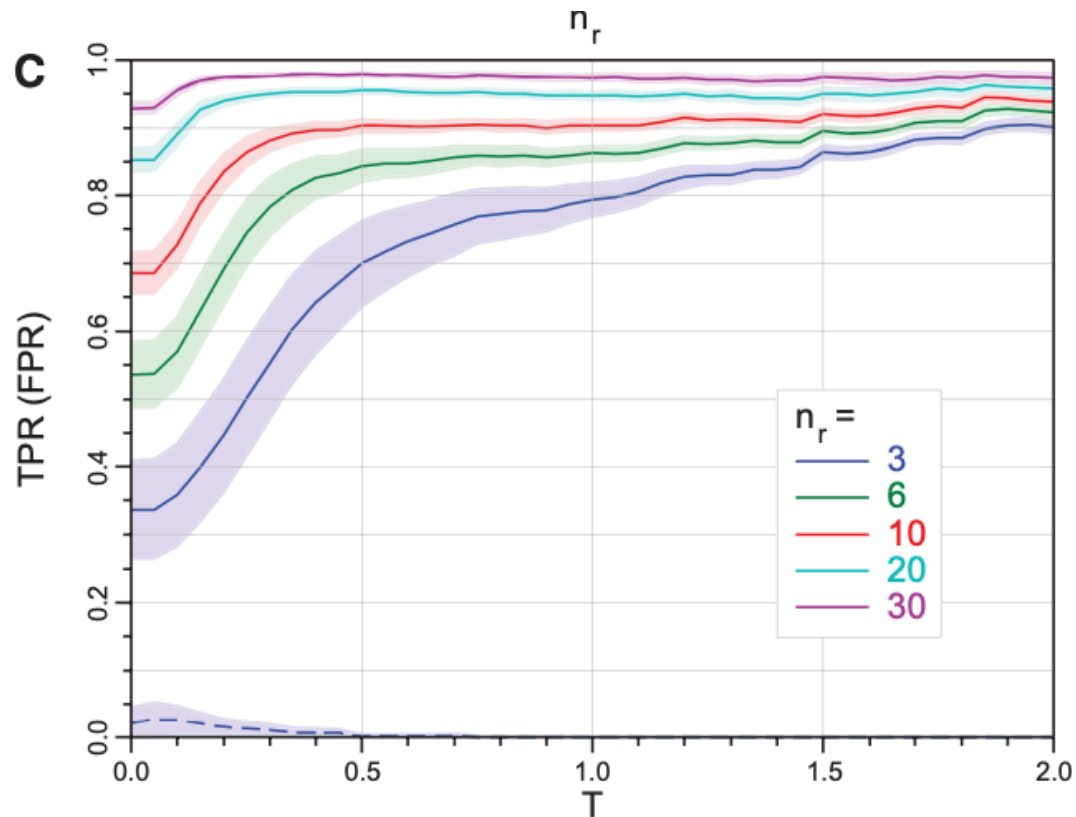


- For a “standard” approach, at least three biological replicates
- No need for technical replicates. Results seem very reproducible
- Multiplexing = multiple samples in one run
  - cuts the cost drastically





# How many replicates?



T:  $|\log_2(\text{FC})|$  threshold

TPR: True positive rate

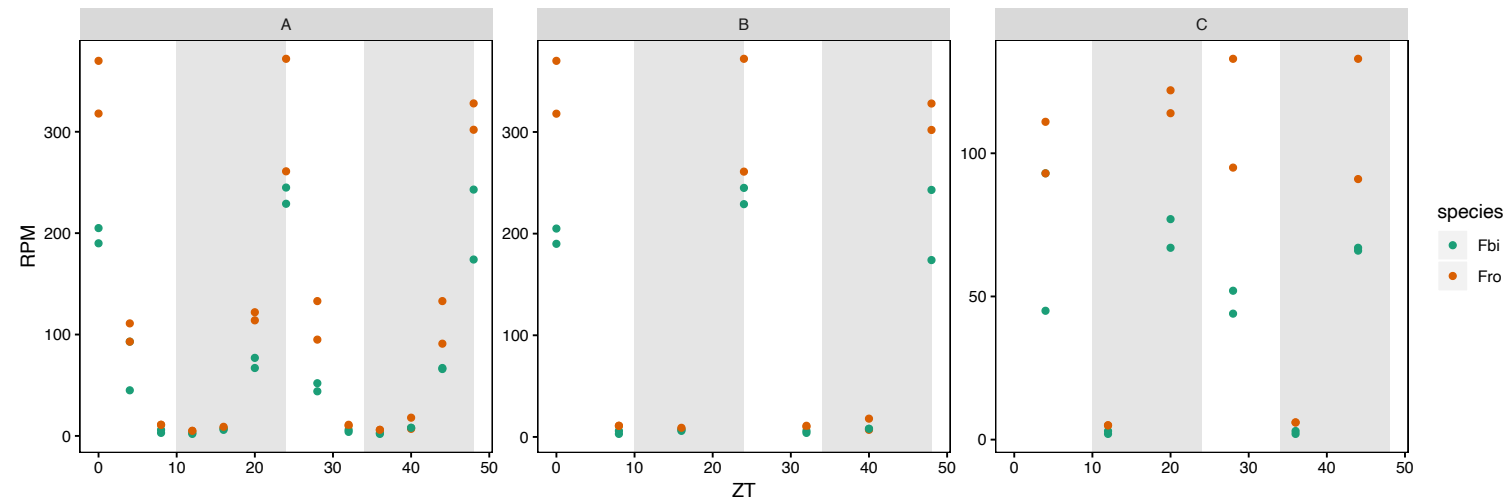
- At least six replicates per condition for all experiments
- Apply a fold-change threshold appropriate to the number of replicates per condition between  $0.1 \leq T \leq 0.5$

Schurch NJ et al. (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA Jun;22(6):839-51. doi: 10.1261/rna.053959.115



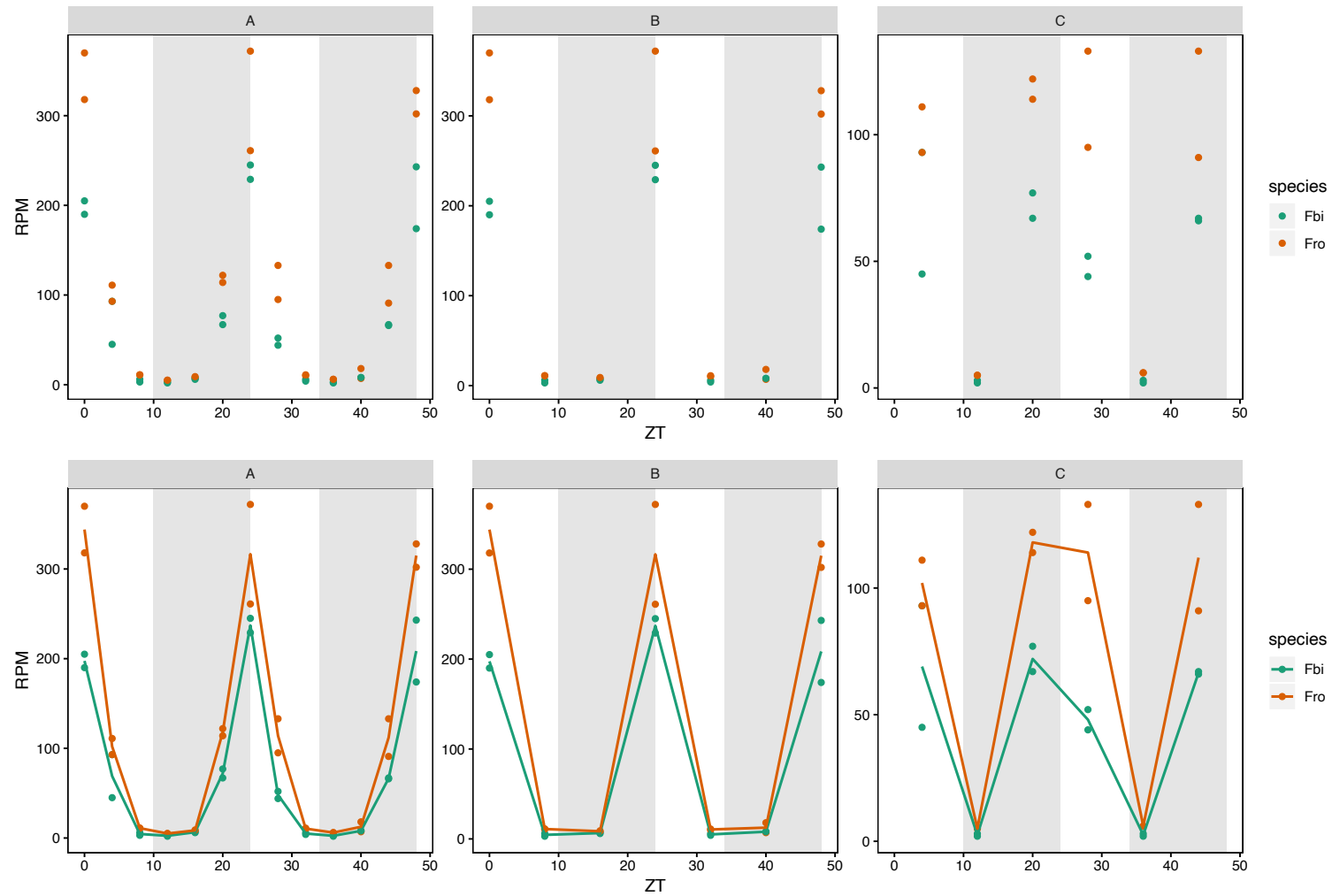


# Time-course: replication vs. resolution





# Time-course: replication vs. resolution





# Preparing high quality RNA

- RNA stabilisation
  - Keep RNA samples on ice
  - Avoid freeze thaw cycles
  - Shock-freeze and extract material in liquid nitrogen or commercially available RNA stabilization reagents (e.g. QIAGEN RNeasy)
- Prevent contamination with RNase
  - treat all surfaces and pipets with RNase EXITUS
  - wear lab coat and gloves
  - use pipet tips fresh from package

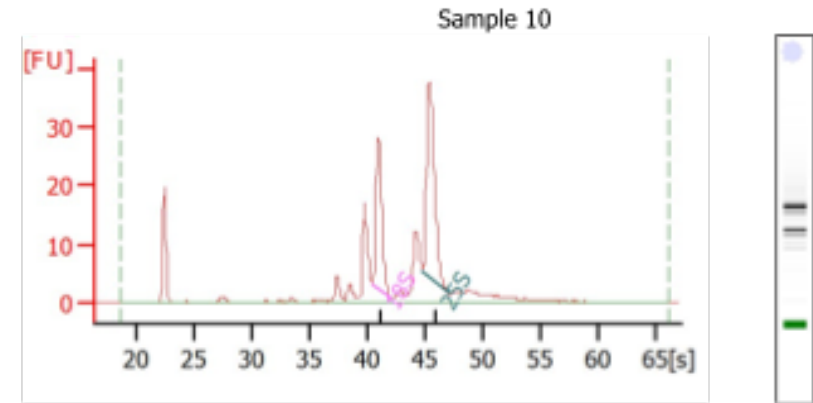
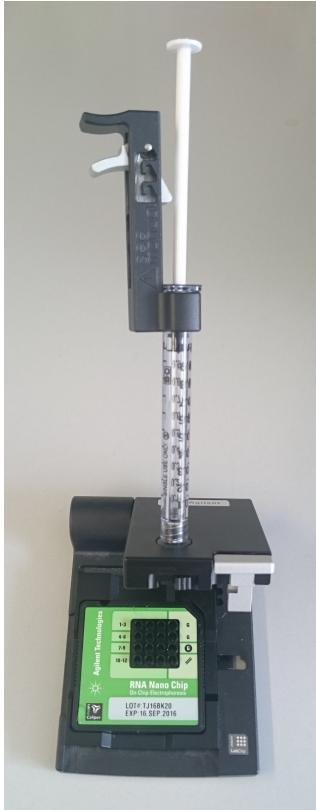




- Optimize the RNA isolation protocol for your organism and tissue of interest
  - cell and tissue disruption
  - compatibility of compounds in the tissue and your extraction buffer
- For a variety of plant tissues we have good experience with
  - Mortar and pestle, bead mill and motor driven pestle
  - SDS + phenol/chloroform extraction, Qiagen RNeasy plant mini Kit



# Quality control - Agilent Bioanalyzer



## Overall Results for sample 10 : Sample 10

RNA Area: 204.8  
RNA Concentration: 79 ng/ $\mu$ l  
rRNA Ratio [25s / 18s]: 1.6  
RNA Integrity Number (RIN): 8.7 (B.02.08)  
Result Flagging Color:    
Result Flagging Label: RIN: 8.70

## Fragment table for sample 10 : Sample 10

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	40.49	42.07	36.2	17.7
25S	44.86	46.97	59.1	28.9

<https://www.genomics.agilent.com/en/Bioanalyzer-System/2100-Bioanalyzer-Instruments/>

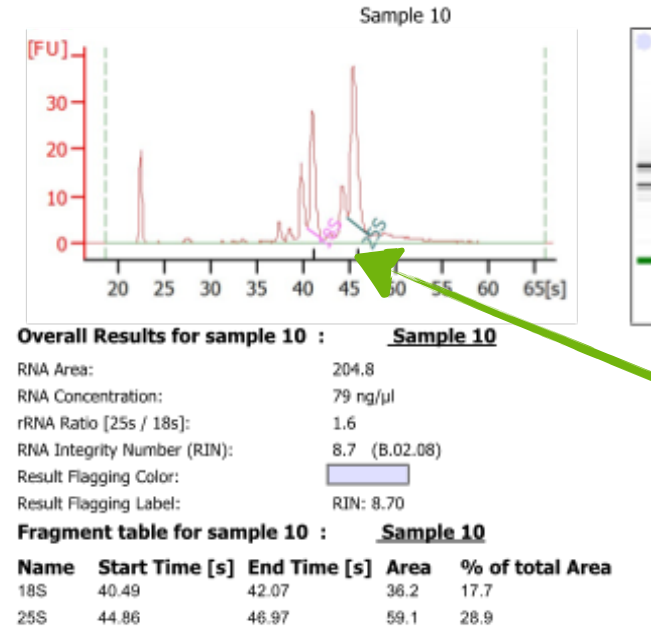
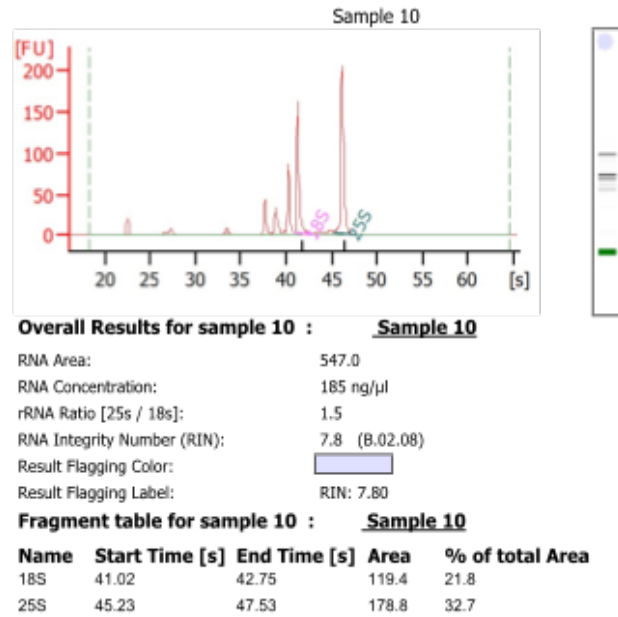


- Perform DNase digestion!!
  - Minimize incubation time to avoid RNA digestion
- gDNA contamination is a serious problem and you will have no chance to distinguish the reads after the sequencing
- gDNA oftentimes incompletely eliminated during library preparation
- Check on gels and the Bioanalyzer for DNA contamination
- When in doubt, use PCR to check for gDNA





# Quality control - Agilent Bioanalyzer



gDNA

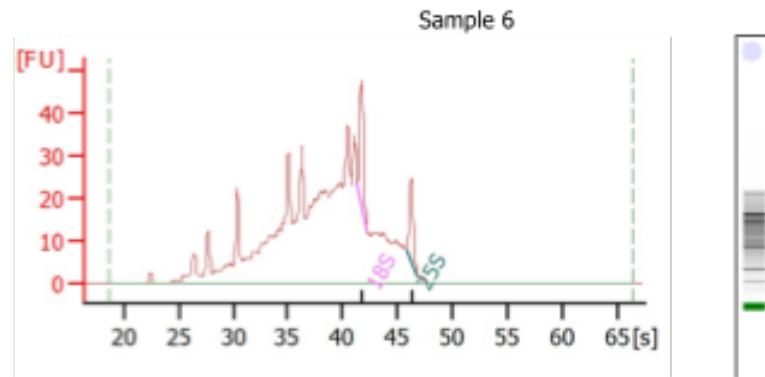
## Arabidopsis RNA

- very good quality
- DNase treated
- RIN number slightly decreased


- Good RIN
- But: gDNA contamination



# Quality control - Agilent Bioanalyzer

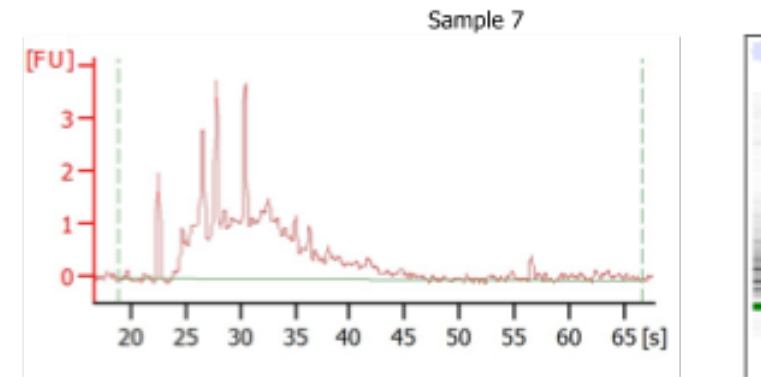


## Overall Results for sample 6 : Sample 6


RNA Area: 770.2  
RNA Concentration: 1,276 ng/μl  
rRNA Ratio [25s / 18s]: 0.6  
RNA Integrity Number (RIN): 3.6 (B.02.08)  
Result Flagging Color:   
Result Flagging Label: RIN: 3.60

## Fragment table for sample 6 : Sample 6

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.35	42.21	26.1	3.4
25S	45.77	46.83	16.7	2.2

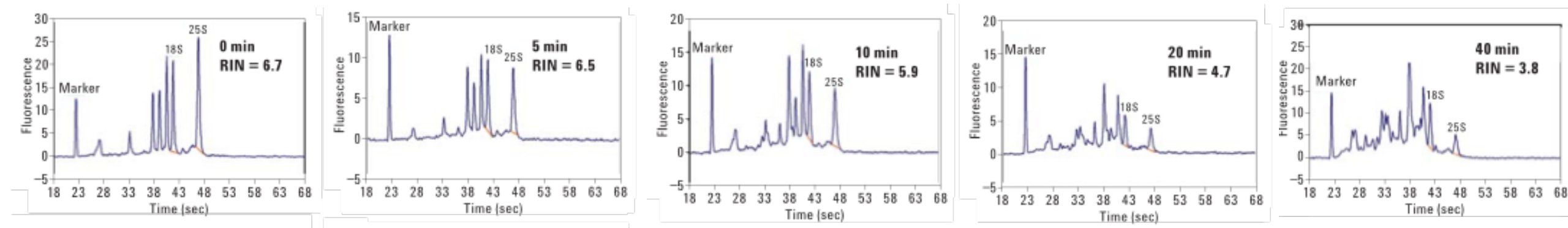


## Overall Results for sample 7 : Sample 7

RNA Area: 57.1  
RNA Concentration: 95 ng/μl  
rRNA Ratio [25s / 18s]: 0.0  
RNA Integrity Number (RIN): 1.9 (B.02.08)  
Result Flagging Color:   
Result Flagging Label: RIN: 1.90



# Time-dependent Degradation



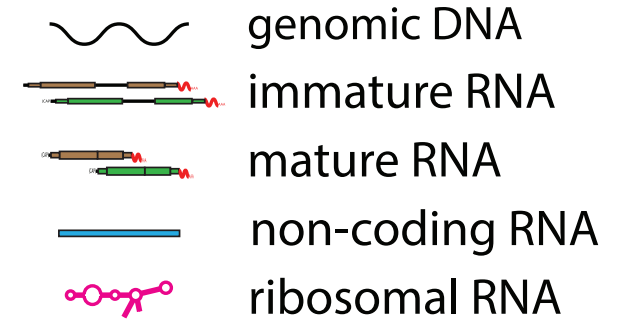
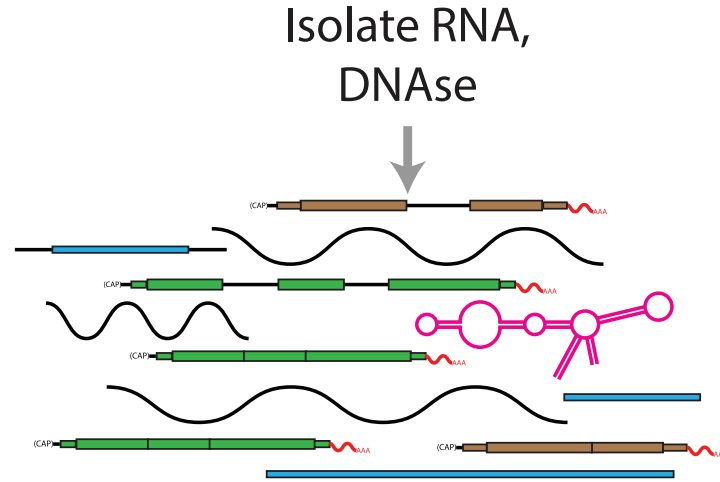
Time course degradation of tobacco RNA. Total RNA samples were incubated at room temperature at selected time points and separated with RNA Nano 6000 kit. Electropherograms indicate the position of 18S and 25S rRNA peaks. RIN values of individual samples are shown.

<https://www.genomics.agilent.com/en/Bioanalyzer-System/2100-Bioanalyzer-Instruments/>



# Which RNA do I actually Seq?

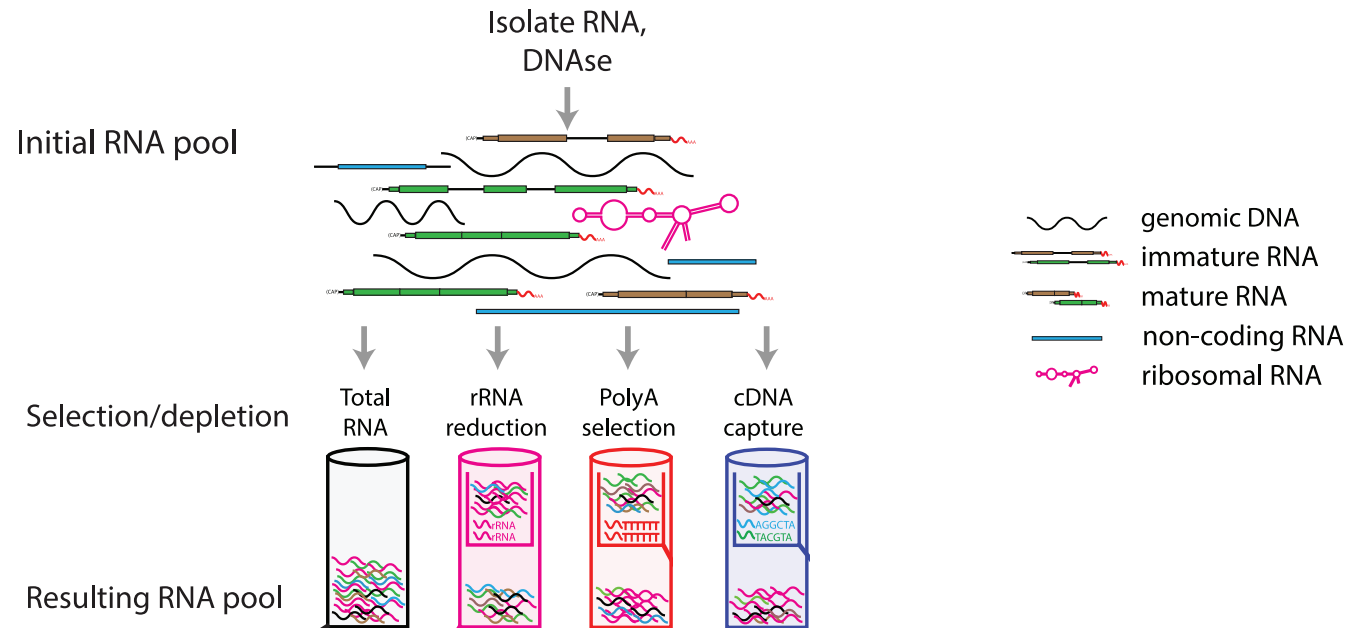
Initial RNA pool



Adapted from Griffith M (2015) PLoS Comput Biol. doi: 10.1371/journal.pcbi.1004393



# RNA selection / depletion



	Total RNA	rRNA reduction	PolyA selection	cDNA capture
<b>Transcript representation</b>	Broad	Broad	Limited	Limited (targeted)
<b>ribosomal RNA</b>	High	Low	Very low	Very low
<b>Abundant mRNAs</b>	dominate	dominate	dominate	de-emphasized
<b>Unprocessed RNA</b>	High	High	Low	Moderate
<b>genomic DNA</b>	High	High	Very low	Low

Adapted from Griffith M (2015) PLoS Comput Biol. doi: 10.1371/journal.pcbi.1004393



## “Default” for Plant mRNA-Seq

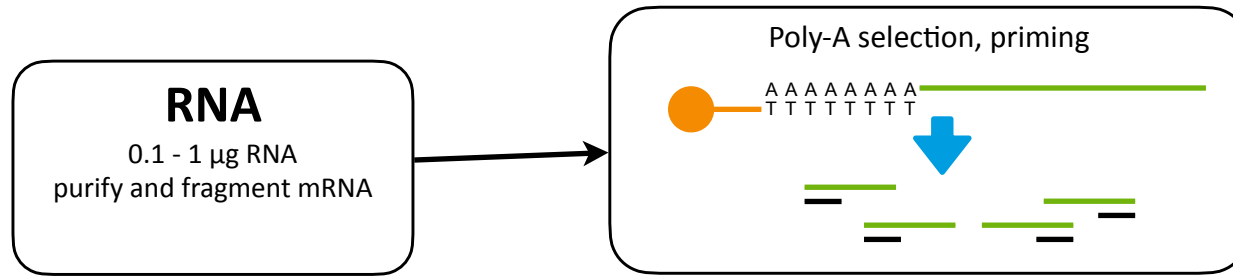


	Total RNA	rRNA reduction	PolyA selection	cDNA capture
<b>Transcript representation</b>	Broad	Broad	Limited	Limited (targeted)
<b>ribosomal RNA</b>	High	Low	Very low	Very low
<b>Abundant mRNAs</b>	dominate	dominate	dominate	de-emphasized
<b>Unprocessed RNA</b>	High	High	Low	Moderate
<b>genomic DNA</b>	High	High	Very low	Low

Adapted from Griffith M (2015) PLoS Comput Biol. doi: 10.1371/journal.pcbi.1004393



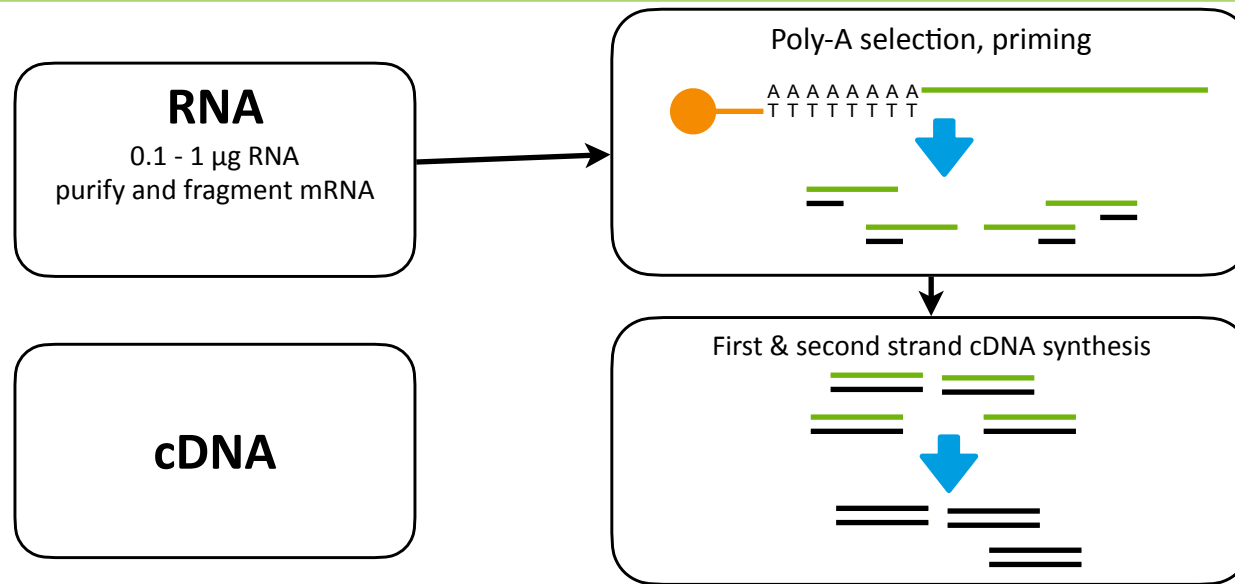
# Illumina workflow



adapted from [www.illumina.com](http://www.illumina.com)



# Illumina workflow

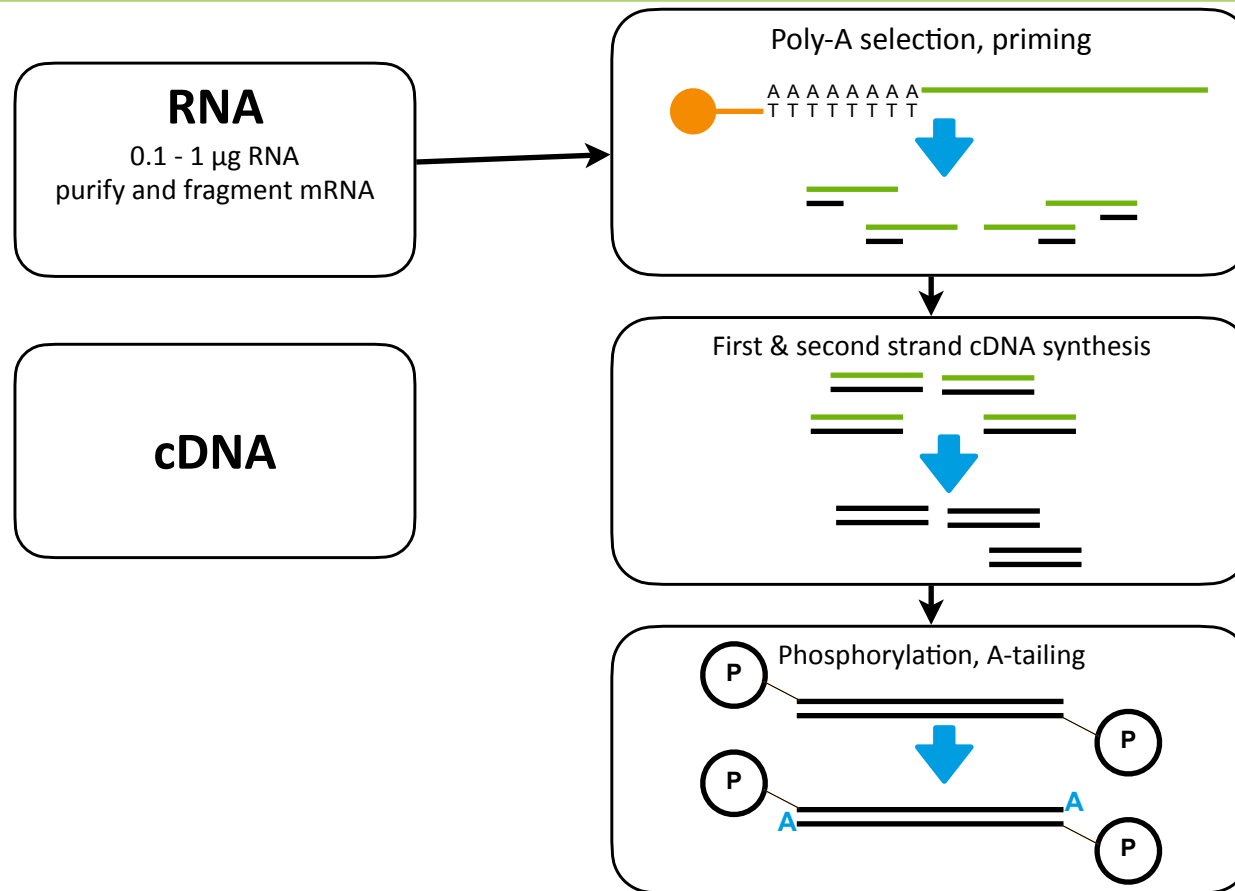


adapted from [www.illumina.com](http://www.illumina.com)





# Illumina workflow

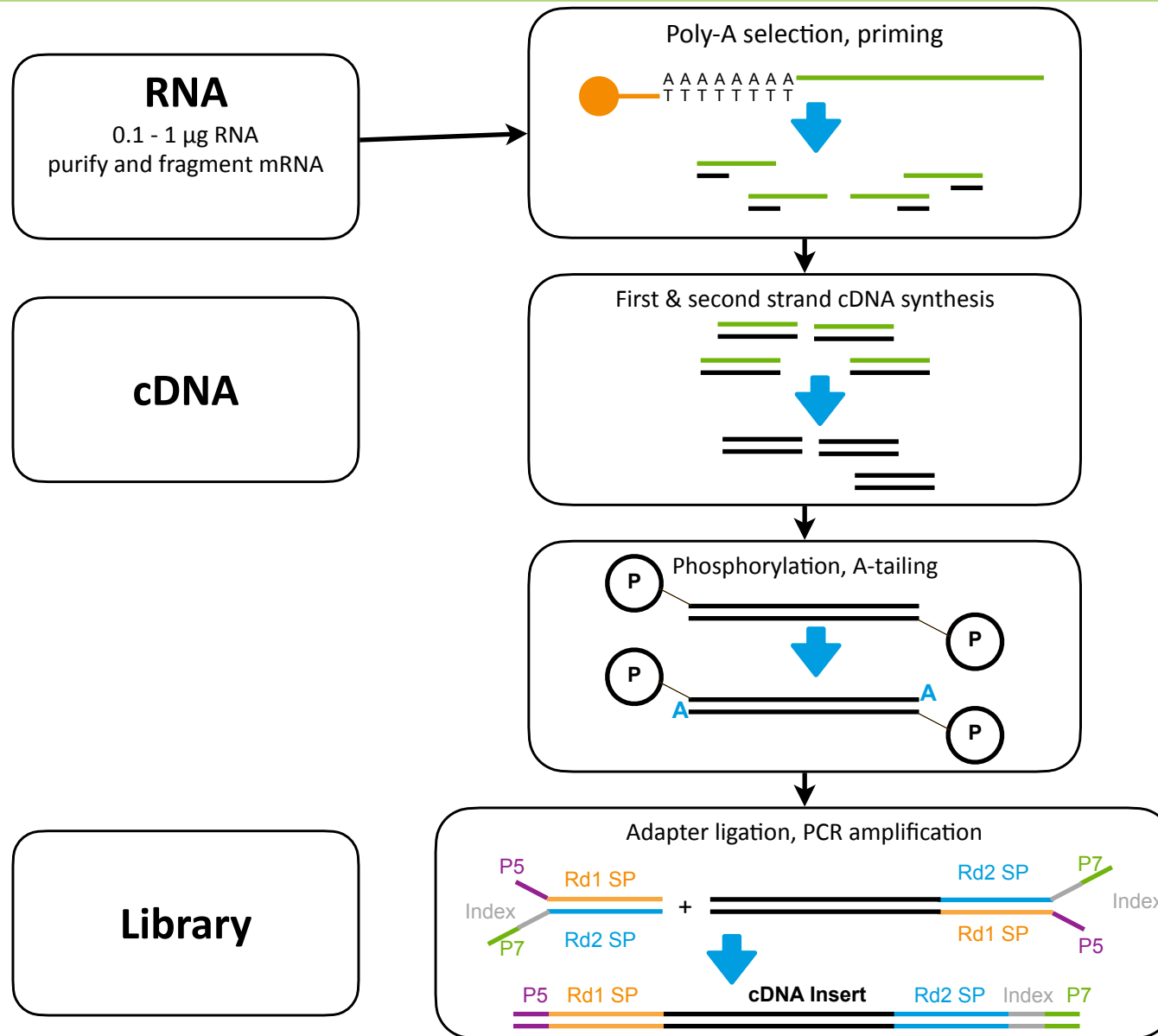


adapted from [www.illumina.com](http://www.illumina.com)





# Illumina workflow



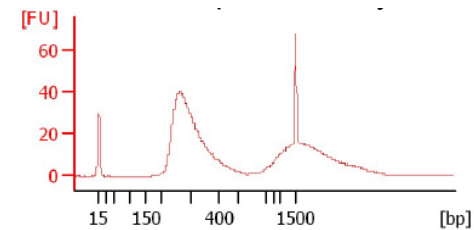
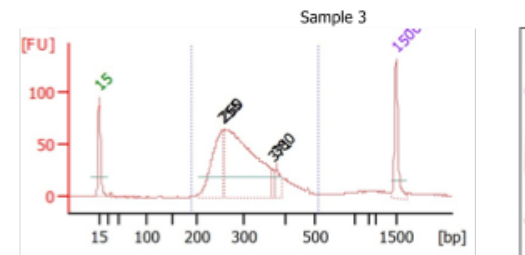
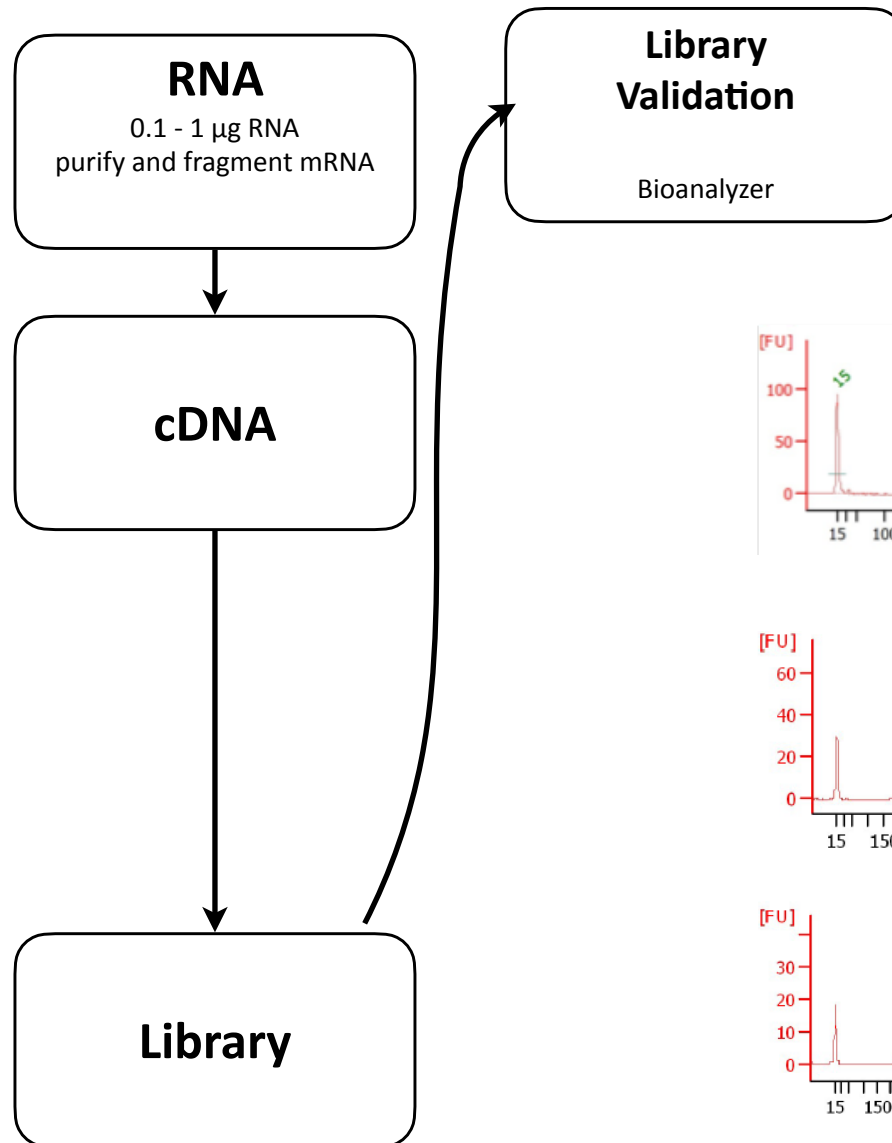
adapted from [www.illumina.com](http://www.illumina.com)



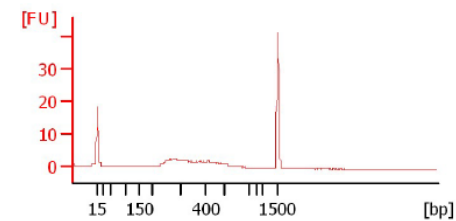




# Illumina workflow



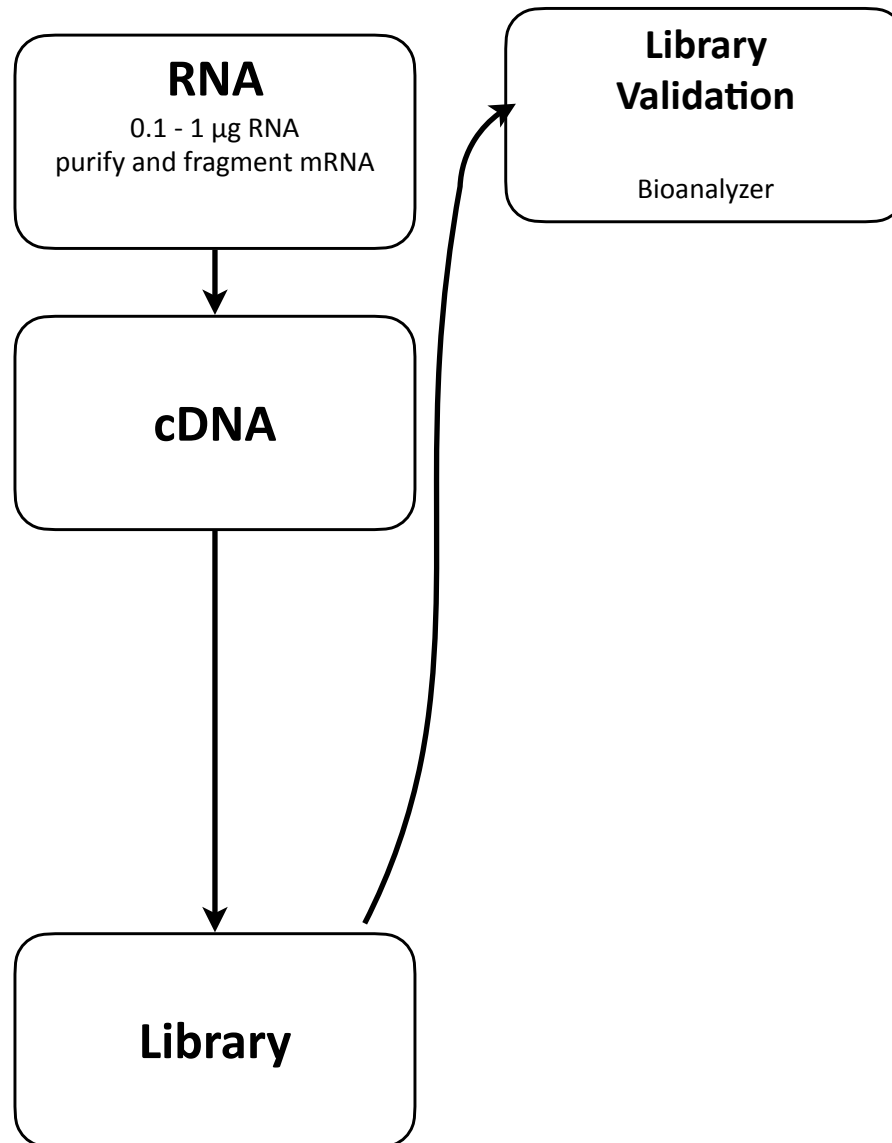
**Over-amplified**



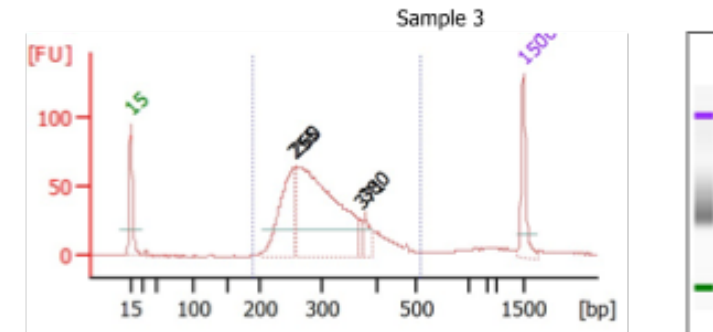
**Under-amplified**



# Illumina workflow



After running a DNA1000 chip set a region to determine your average library size



## Overall Results for sample 3 : Sample 3

Number of peaks found: 4

Area 1: 781.1

## Region table for sample 3 : Sample 3

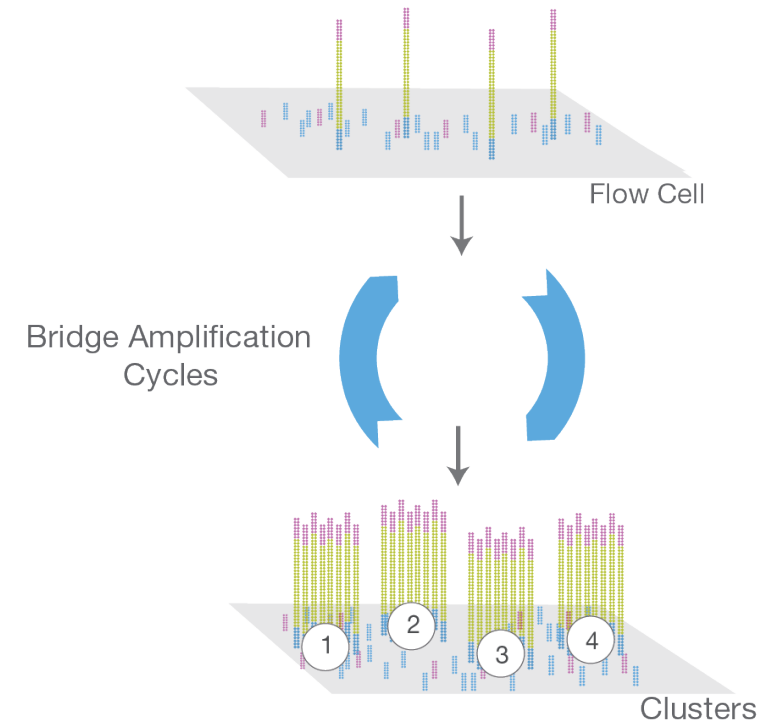
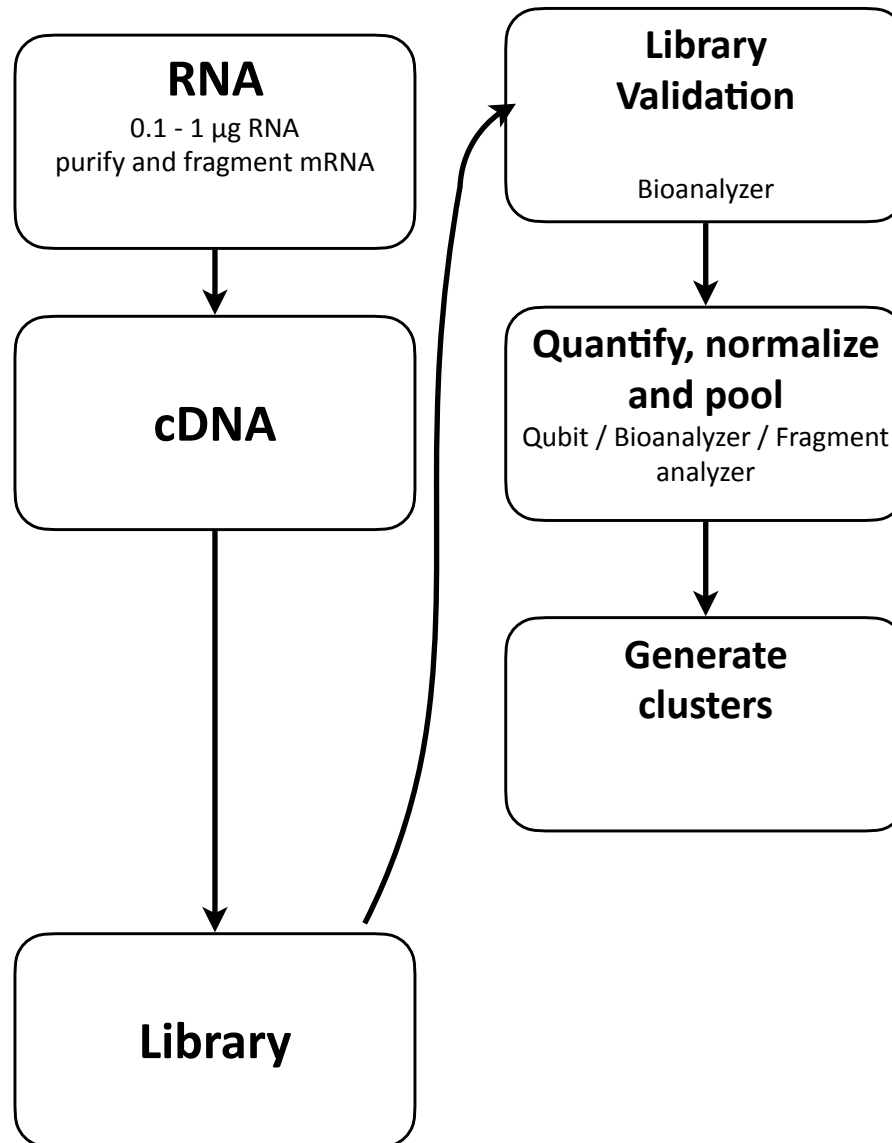
From [bp]	To [bp]	Area [bp]	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [ng/µl]	Co lor
192	522	781.1	85	306	20.0	22.50	

## Peak table for sample 3 : Sample 3

Peak	Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	15	4.20	424.2	Lower Marker
2	255	5.52	32.9	
3	260	14.74	86.0	
4	371	0.44	1.8	
5	380	0.79	3.2	
6	1,500	2.10	2.1	Upper Marker



# Illumina workflow

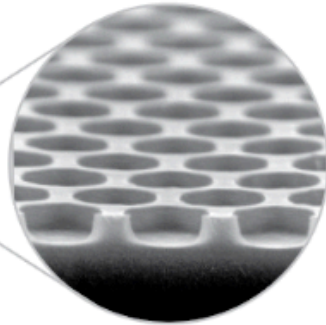


Library is loaded into a flow cell and the fragments are hybridized to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.





A.



[www.illumina.com](http://www.illumina.com)

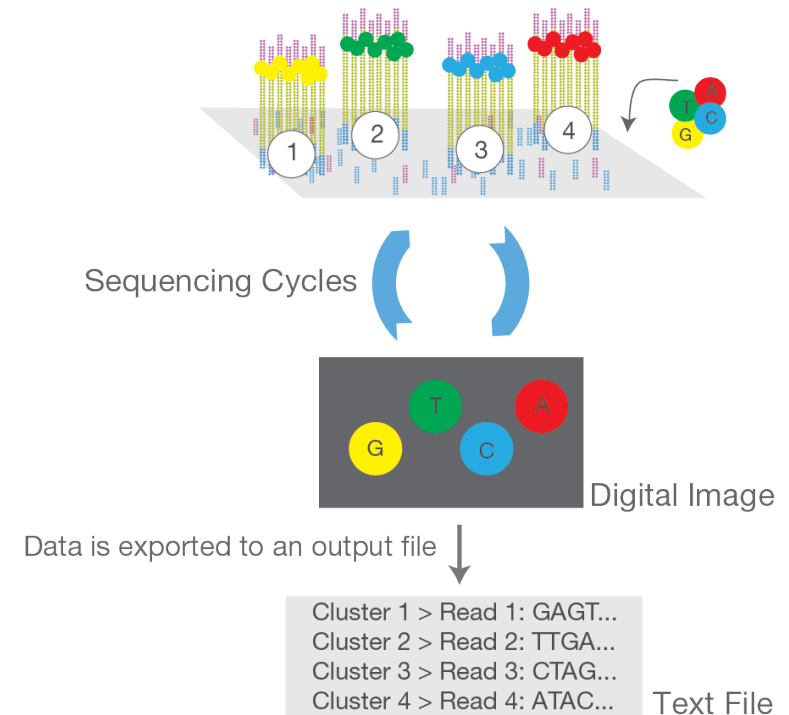
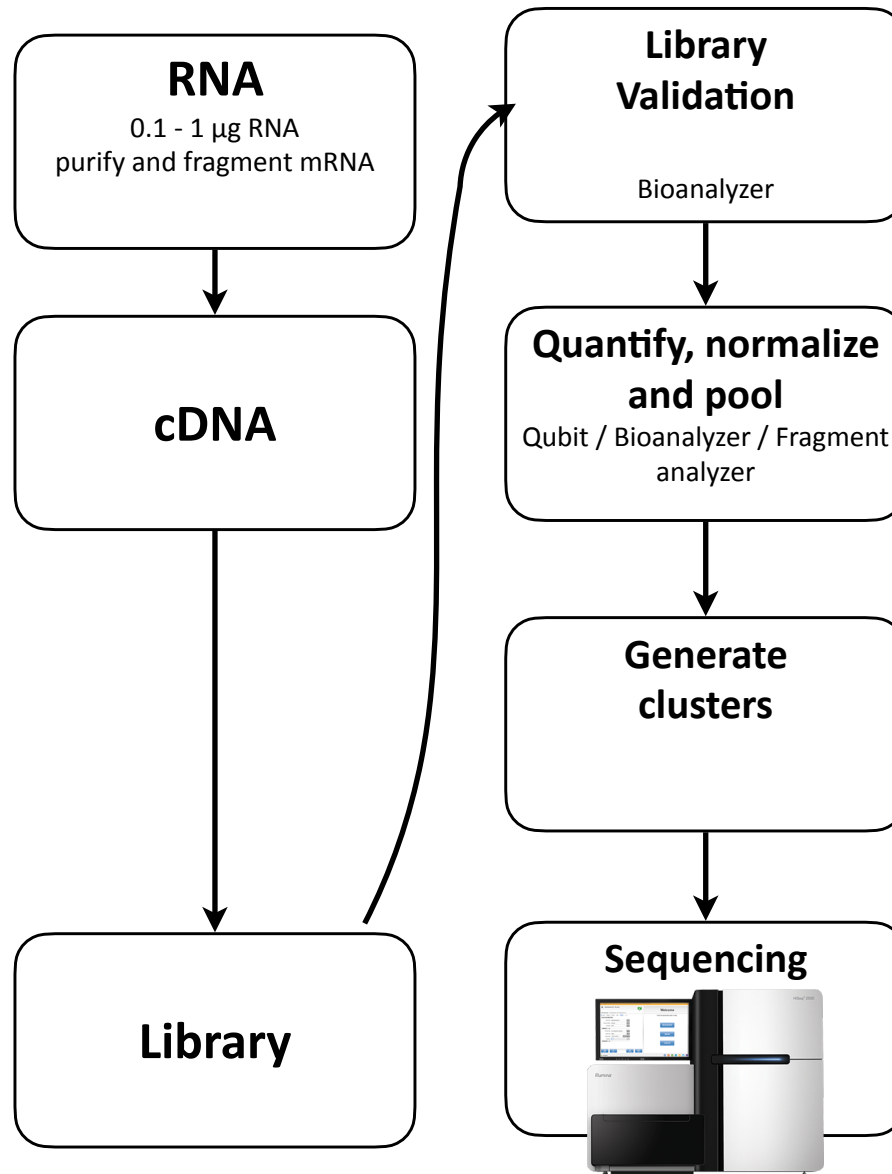
CEPLAS RNA-Seq Workshop 2022

[alisandra.denton@hhu.de](mailto:alisandra.denton@hhu.de) | [dominik.brilhaus@hhu.de](mailto:dominik.brilhaus@hhu.de)





# Illumina workflow

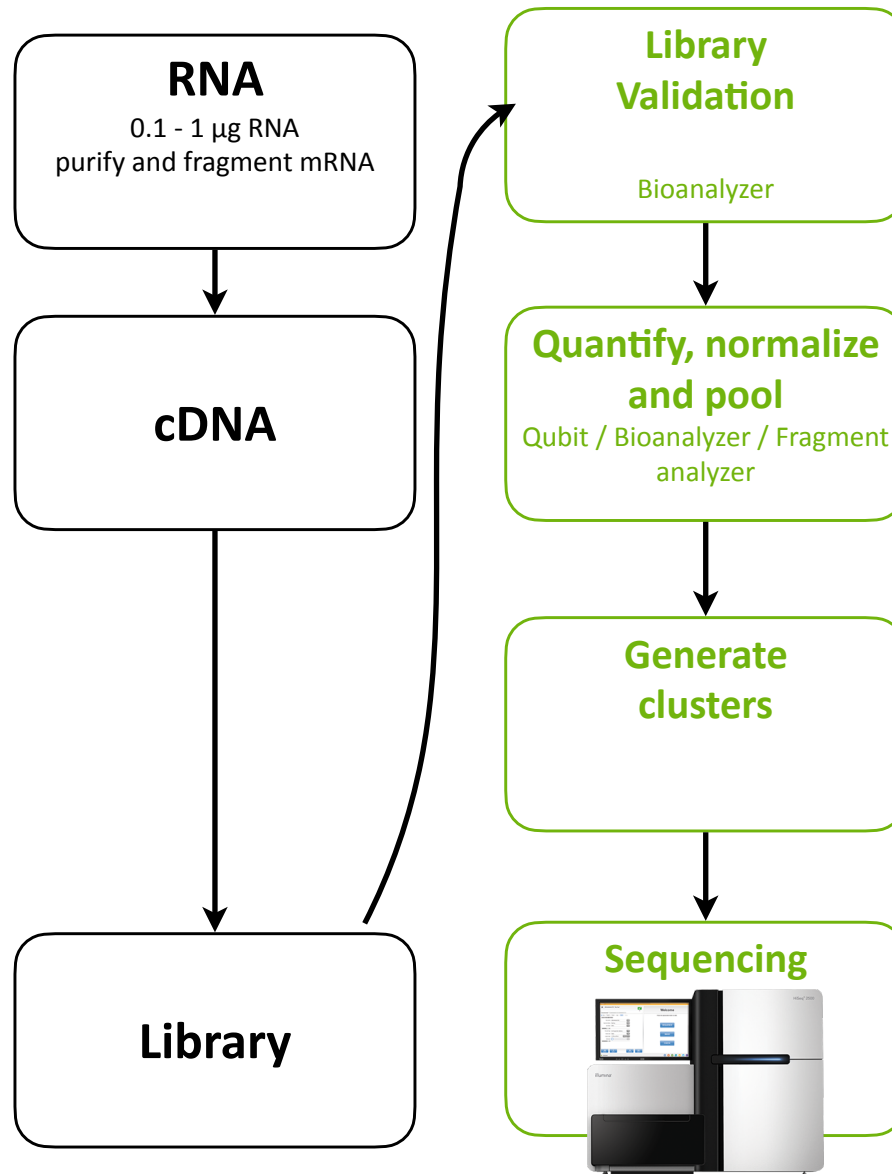


Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated “n” times to create a read length of “n” bases.

[www.illumina.com](http://www.illumina.com)



# Illumina workflow



**Typically outsourced  
to sequencing  
facility / company**

Photo: [www.illumina.com](http://www.illumina.com)





# Sequencing

- Usually outsourced
- Agree on what you receive
  - output format
  - quality control
  - filtering of data
- Think about data storage (backups!!)
  - One flow-cell 100 bp SE reads ~ 150 GB









# Current Illumina systems for mRNA-Seq

## STEP 2

Choose a sequencer

				
Product	MiniSeq System	MiSeq System	NextSeq 550 System	NovaSeq 6000 System
<b>Most important to me</b>	Instrument affordability, mRNA sequencing for small genomes, and onboard informatics	mRNA sequencing for small genomes, read length flexibility, and onboard informatics	RNA-Seq for broad range of genome sizes and instrument affordability	Lowest <sup>a</sup> cost/sample, and scalability for small- and large-scale research project and operations
<b>Max mRNA samples processed/flow cell (20 M reads/sample)<sup>b</sup></b>	1	1	2–8	384 <sup>c</sup>
<b>Max gene expression profiling/run (10 M reads/sample)</b>	2–3	2–3	12–36	384 <sup>c</sup>
<b>Run mode/kit type</b>	Mid-output/ High-output	Mid-output/ High-output	Mid-output/ High-output	SP, S1, S2, S4
<b>Flow cells processed/run</b>	1	1	1	1 or 2

a. Comparison among the Illumina portfolio based on individual lane indexing and S4 flow cells

b. Assumes a human-sized transcriptome

c. Limited by available indexes

[www.illumina.com](http://www.illumina.com) Illumina Methods Guide 2020 v. 6