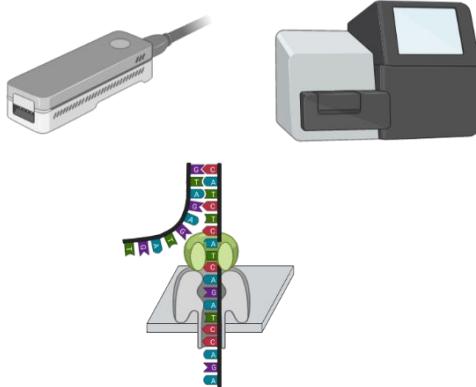
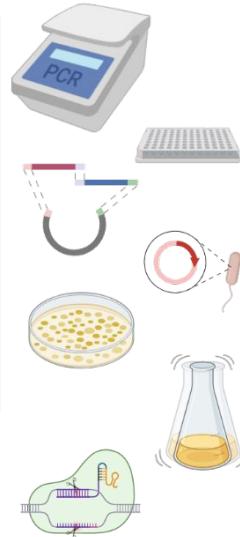
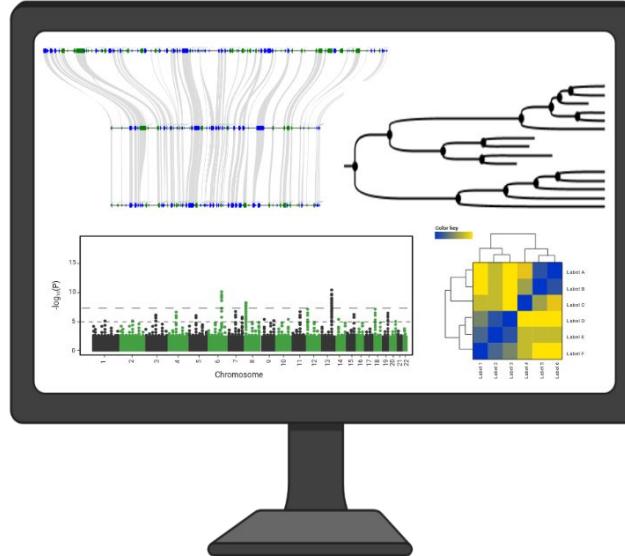




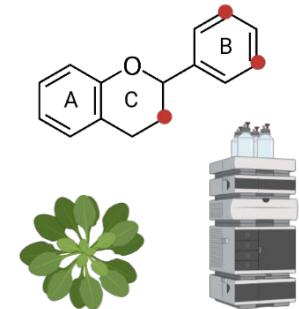
Technische
Universität
Braunschweig



Plant Biotechnology
and Bioinformatics



species biosynthesis proteins analysis
biosynthesis biosynthesis different conditions
functional variants DDA analysis
within genes variants H293-MYB
duplex site data functional Col 100% variant
sequencer ICGM divergently expressed non-canonical
single reference multiple transcripts identified
structure synthesis gene annotation level identified
sites synonymous mutations evolutionary pathway
plants plants accessions systems biology long distance
pigments model genome key against canonical Arabidopsis
Keywords genes across Canophylales
genomes sequencing evolution
across species identification free thaliana
flavonoid conservation sequencing Arabidopsis
gene read transcription synthetic MYB introns residues RNA-Seq
sequence sequence



Introduction to plant transcriptomics

Prof. Dr. Boas Pucker
(Plant Biotechnology and Bioinformatics)

Availability of slides

- All materials are freely available (CC BY) - after the lectures:
 - StudIP: **Applied Plant Transcriptomics**
 - GitHub: <https://github.com/bpucker/teaching>
- Questions: Feel free to ask at any time
- Feedback, comments, or questions: b.pucker[a]tu-bs.de

My figures and content can be re-used in accordance with CC BY 4.0, but this might not apply to all images/logos. Some figure were constructed using bioRender.com.

What causes these pigmentation patterns?

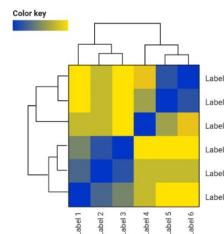
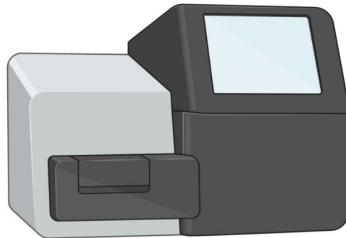


What is transcriptomics?



What is transcriptomics?

- Study of the transcriptome with high-throughput methods
- Characterized by technological progress and big data analyses



What is the transcriptome?



What is the transcriptome?

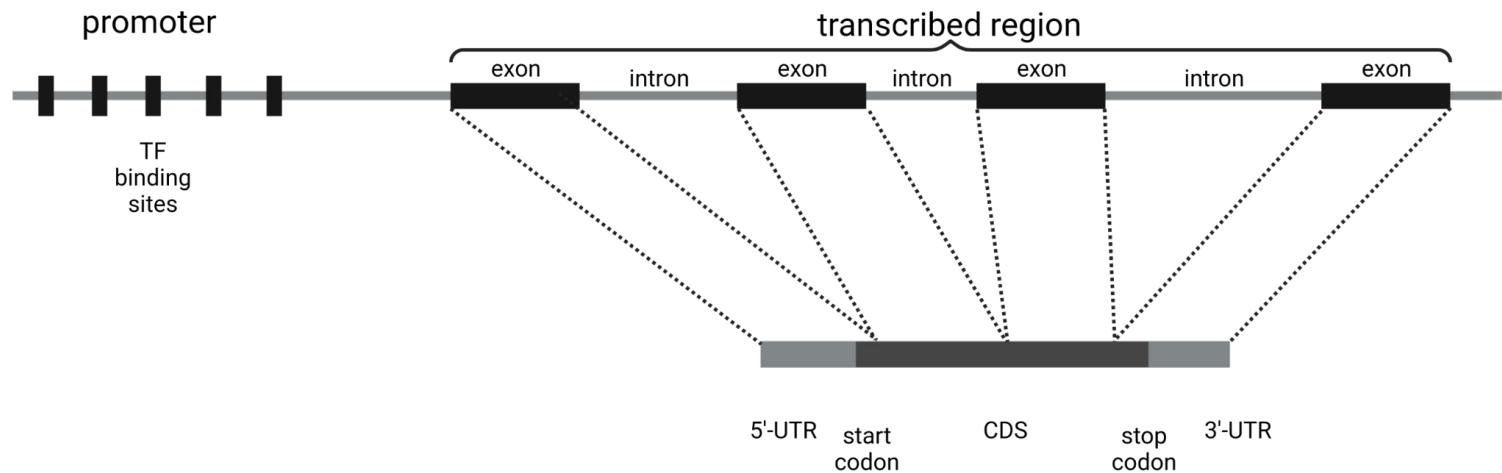
- “The set of all transcripts and their abundances in a defined cell/tissue/organism under defined conditions at a certain time point.”
- Highly variable over time
- Responding to various stimuli

What is a gene?

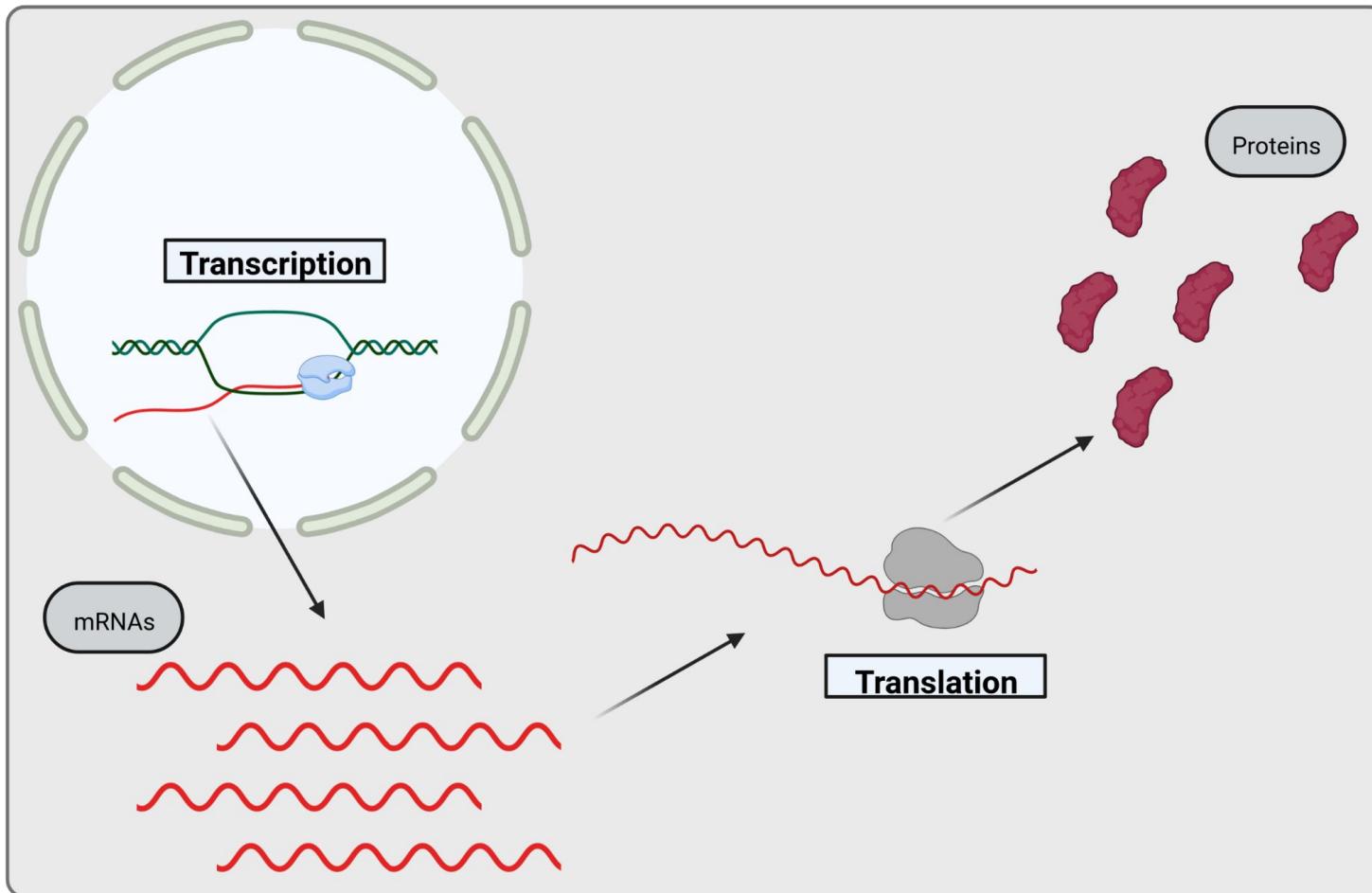


What is a gene?

- No perfect and universal definition
- Restrict to protein coding gene in plants:
 - promoter region
 - UTRs
 - coding sequence
 - introns



Protein biosynthesis (overview)

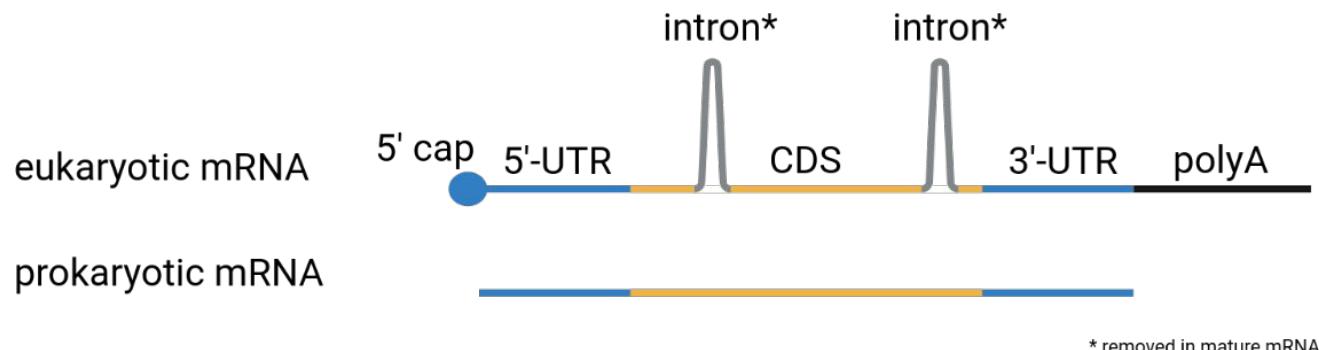


Types of RNA

- rRNA
- tRNA
- mRNA
- Various regulatory RNAs

Prokaryotes vs. eukaryotes - transcripts

- Eukaryotic genes can harbour introns
- Eukaryotic genes have a 5'-cap
- Eukaryotic genes have a poly-A tail



RNA extraction

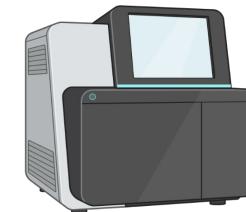
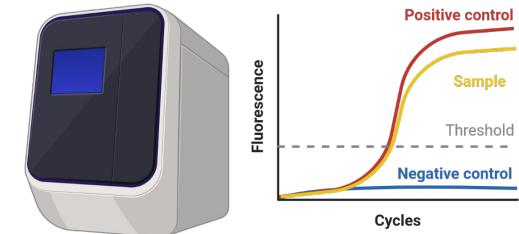
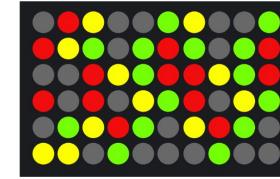
- Classical RNA extraction approaches:
 - Phenol-based methods: helpful to remove high polysaccharide contaminants
 - Trizol-based methods: commercial reagent combining phenol and guanidine isothiocyanate
 - CTAB-based methods: successive removal of polysaccharides, proteins, and specialized metabolites
- Kit-based RNA extraction methods

How to measure gene expression?



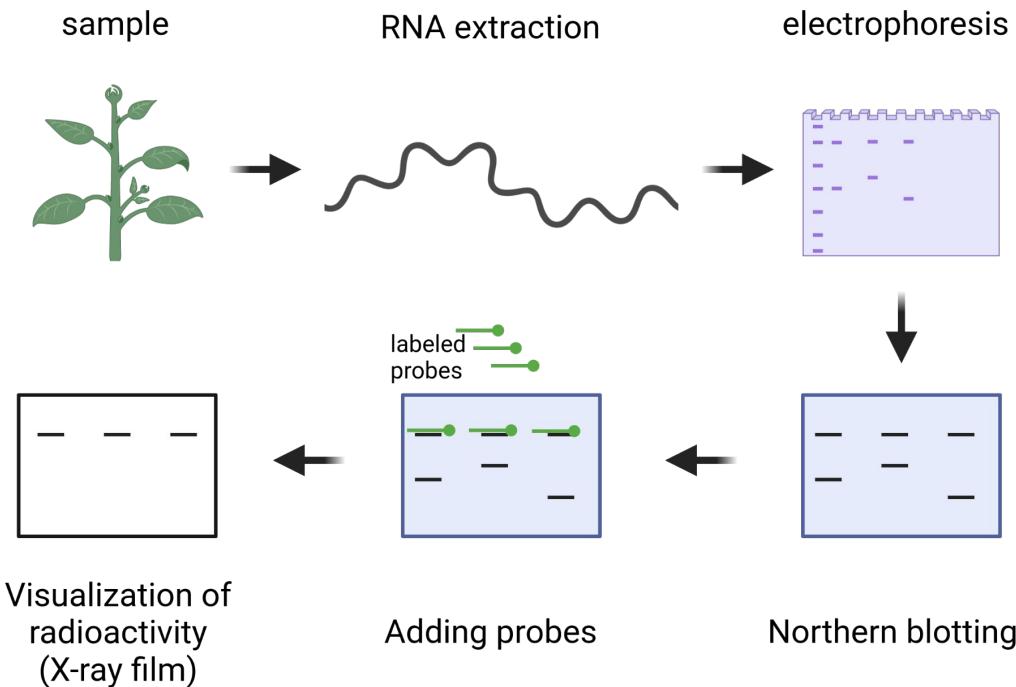
How to measure gene expression?

- Northern blot
- Reverse transcription quantitative PCR (RT-qPCR)
- Microarray
- Expressed Sequence Tags (ESTs)
- Serial Analysis of Gene Expression (SAGE)
- RNA-seq (not RNA sequencing!!)
- Droplet digital PCR (ddPCR)
- direct RNA sequencing



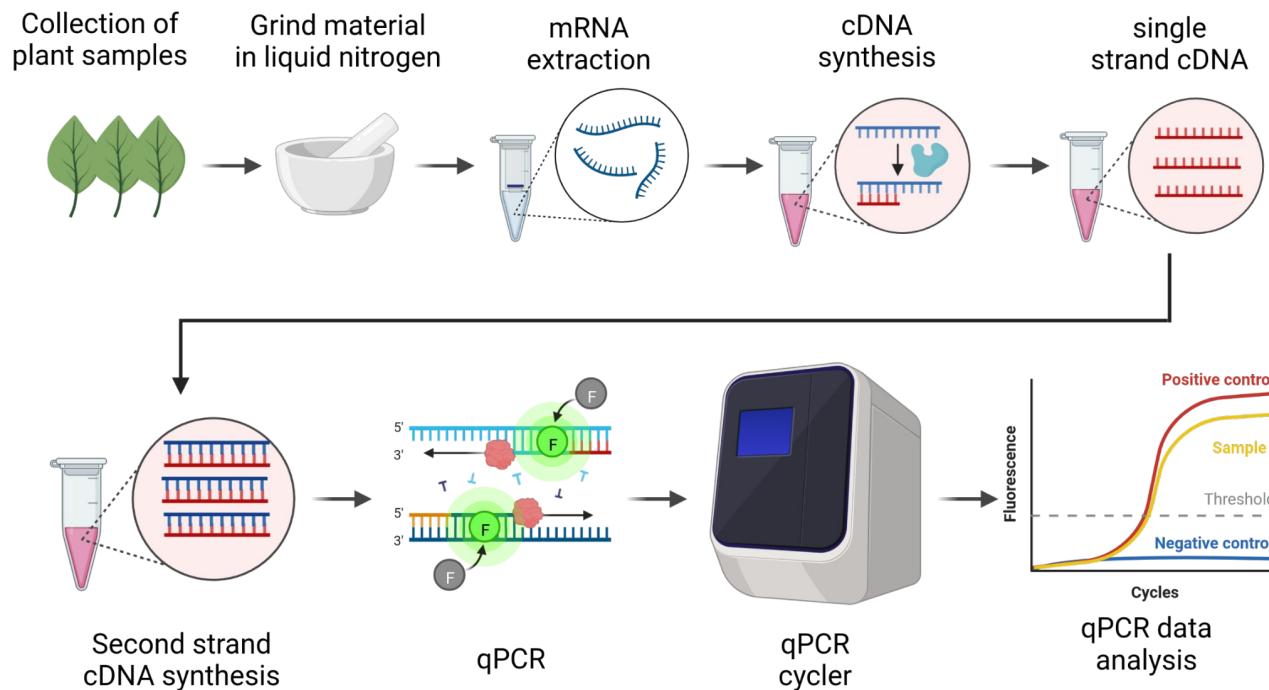
Northern blot

- Electrophoresis to separate mRNAs on a gel
- Visualization of RNA with radioactively labeled probes



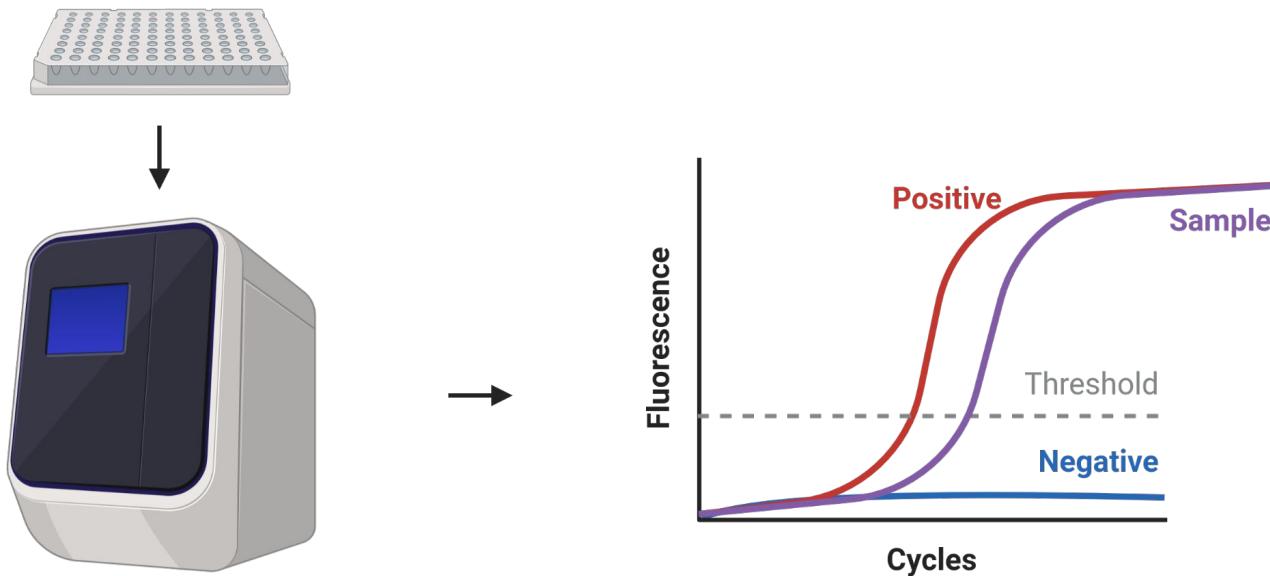
Reverse transcription quantitative PCR (RT-qPCR)

- Reverse transcription = conversion of mRNA into cDNA
- qPCR = quantitative PCR, real time detection of product formation

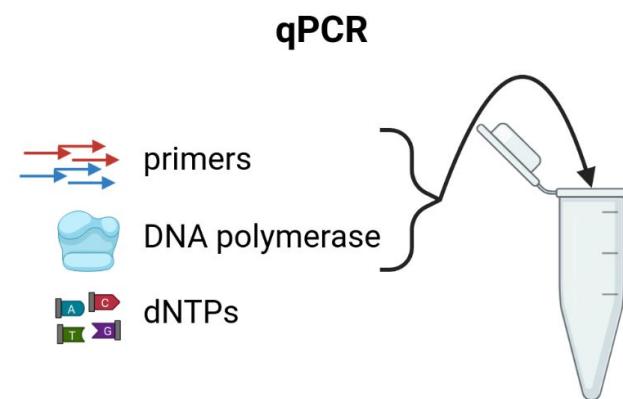
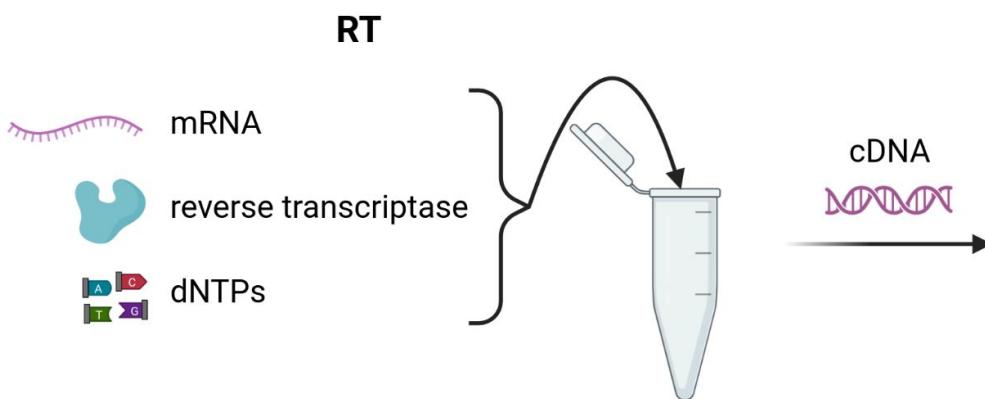
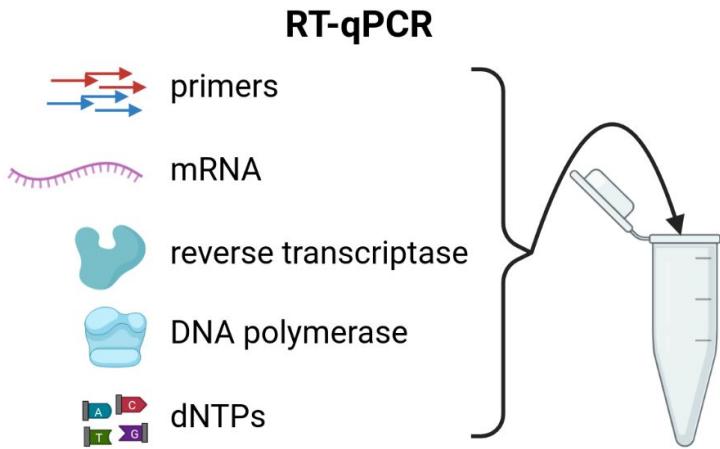


qPCR concept

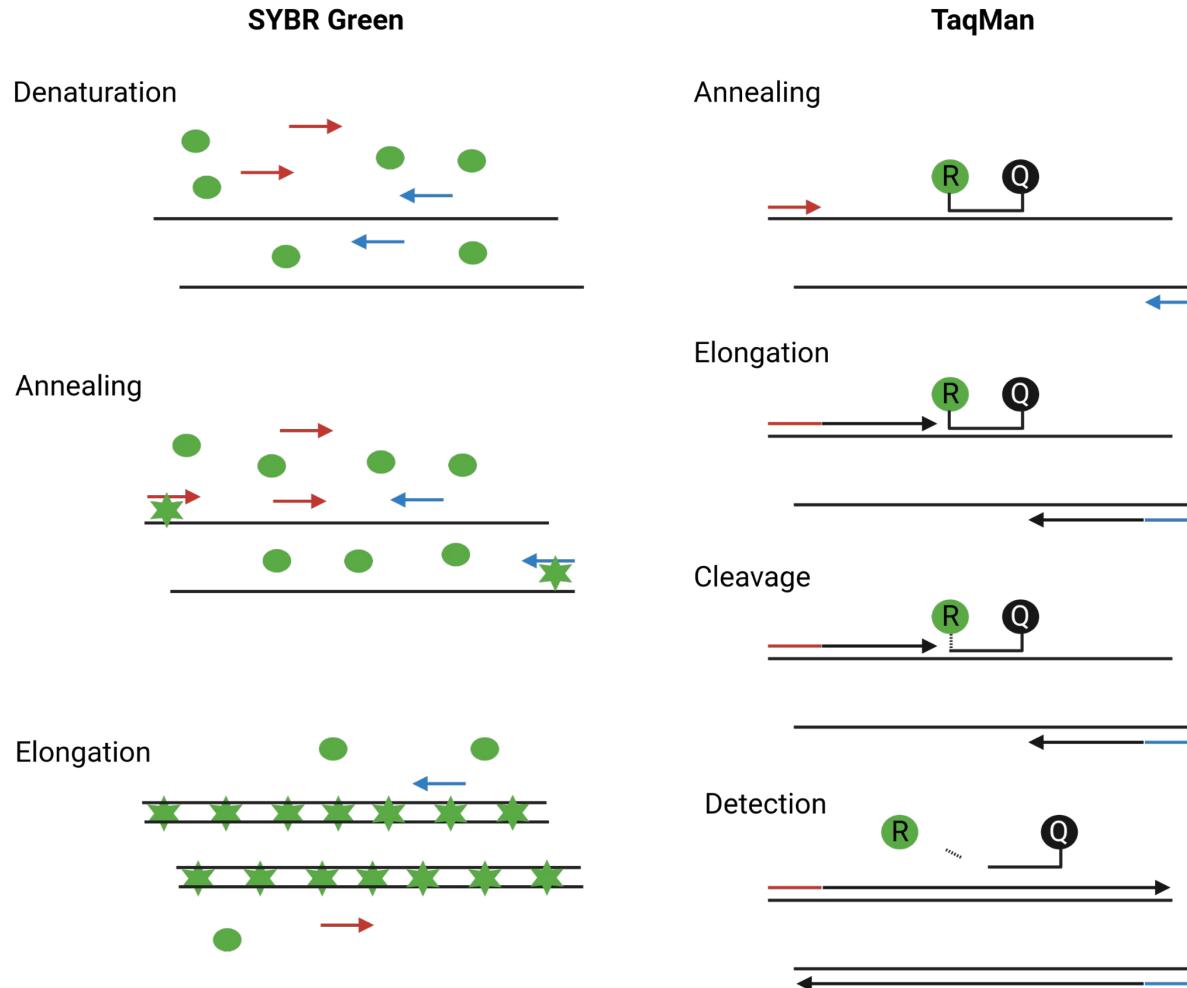
- Monitoring formation of PCR product in real time
- Comparison of sample against positive/negative controls
- Ct value = cycle in which the threshold is crossed



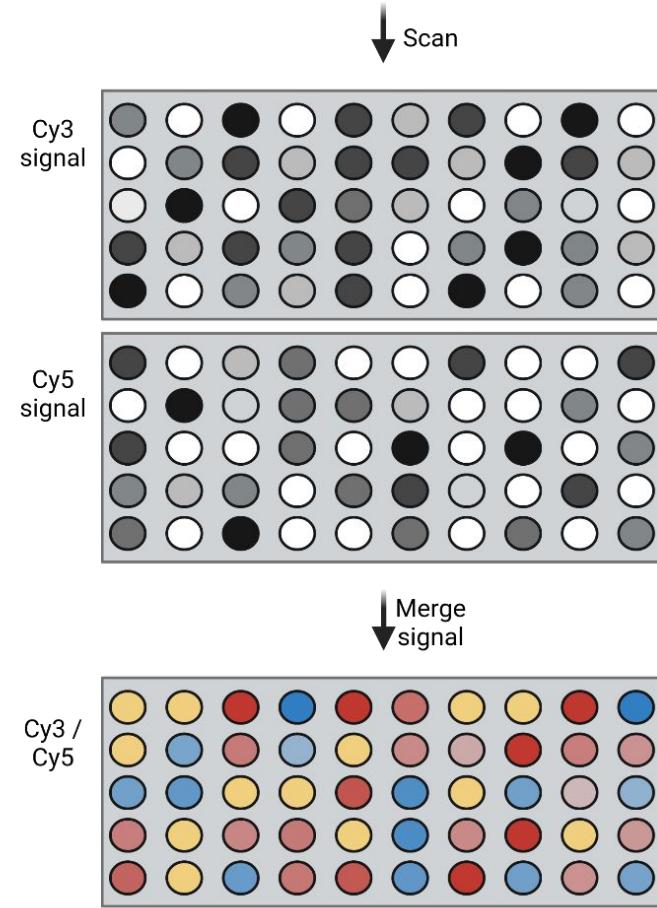
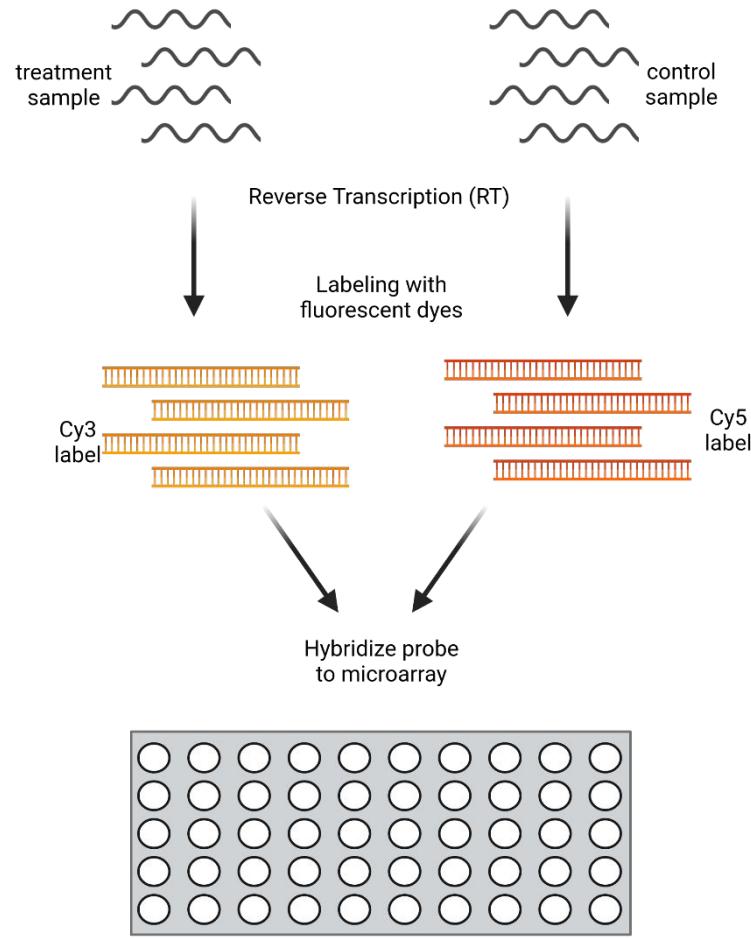
One-step vs. two-steps



qPCR: SYBR green vs. TaqMan



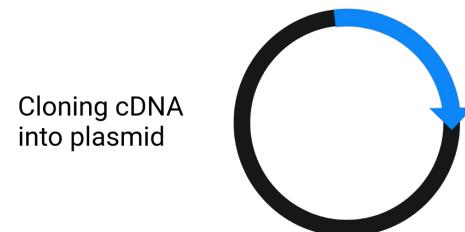
Microarray



Expressed Sequence Tags (ESTs)

- Fragment of a cDNA
- Indicates transcription of certain genomic segments
- Restriction to full length mRNAs/cDNAs possible
- Number of ESTs correlates with transcript abundance

mRNA _____
cDNA _____



Cloning cDNA
into plasmid

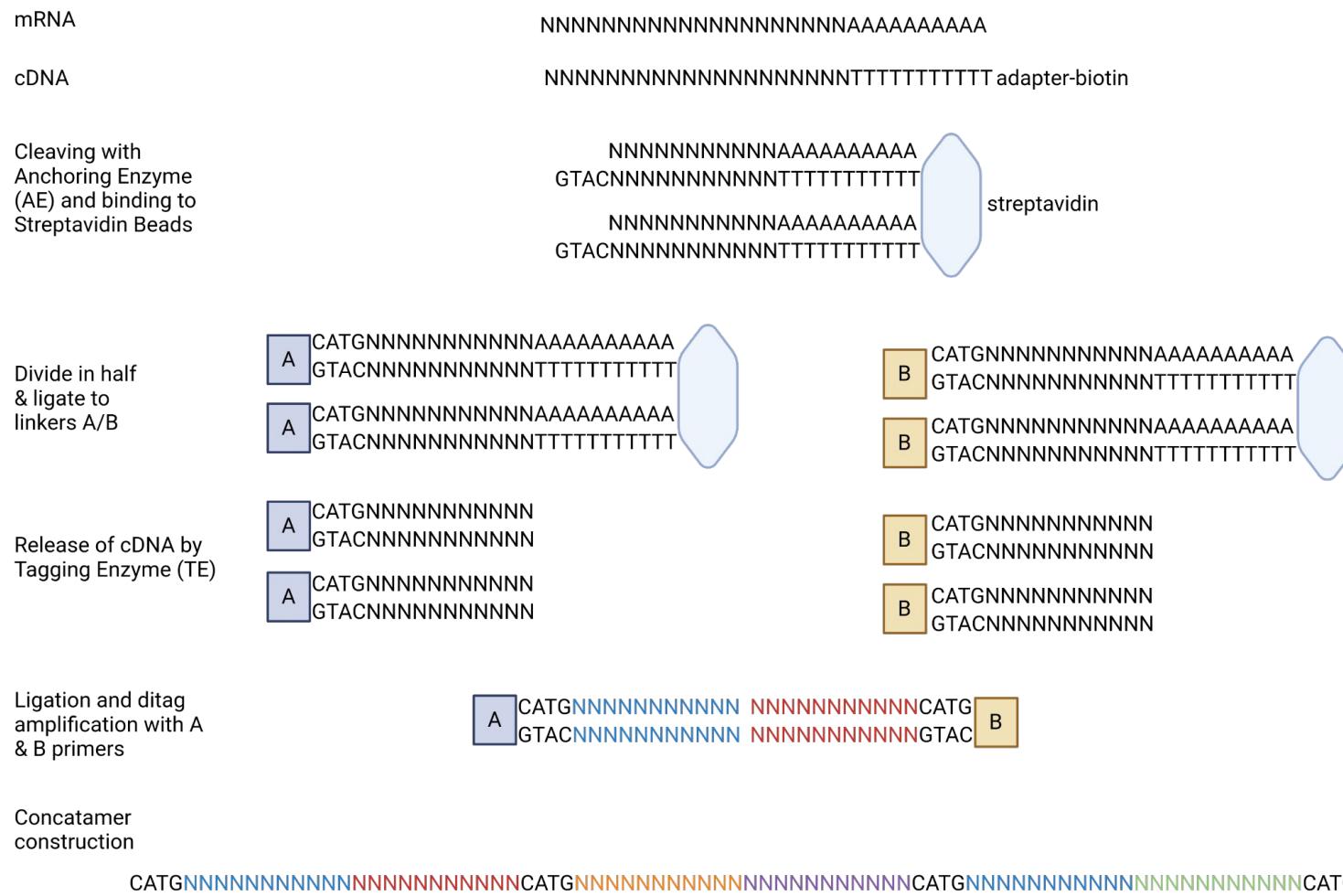
5' EST



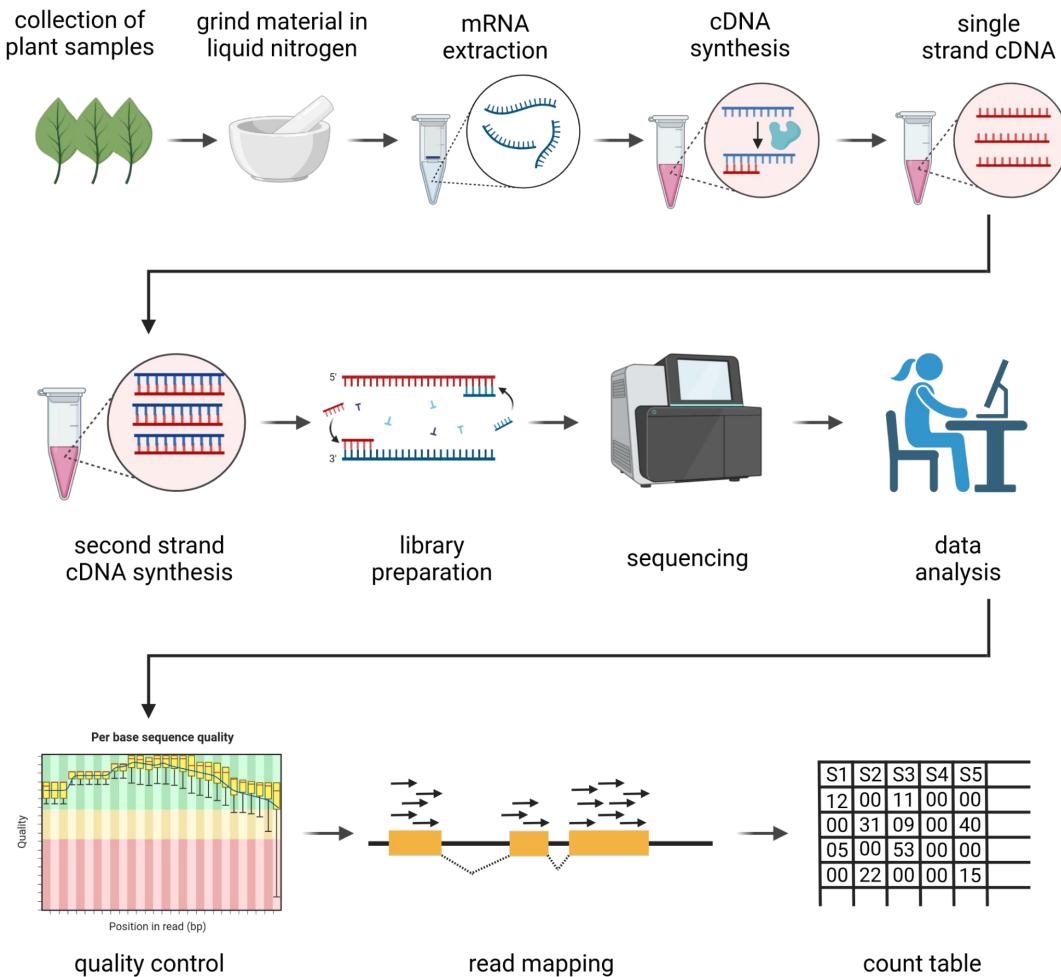
3' EST



Serial Analysis of Gene Expression (SAGE)

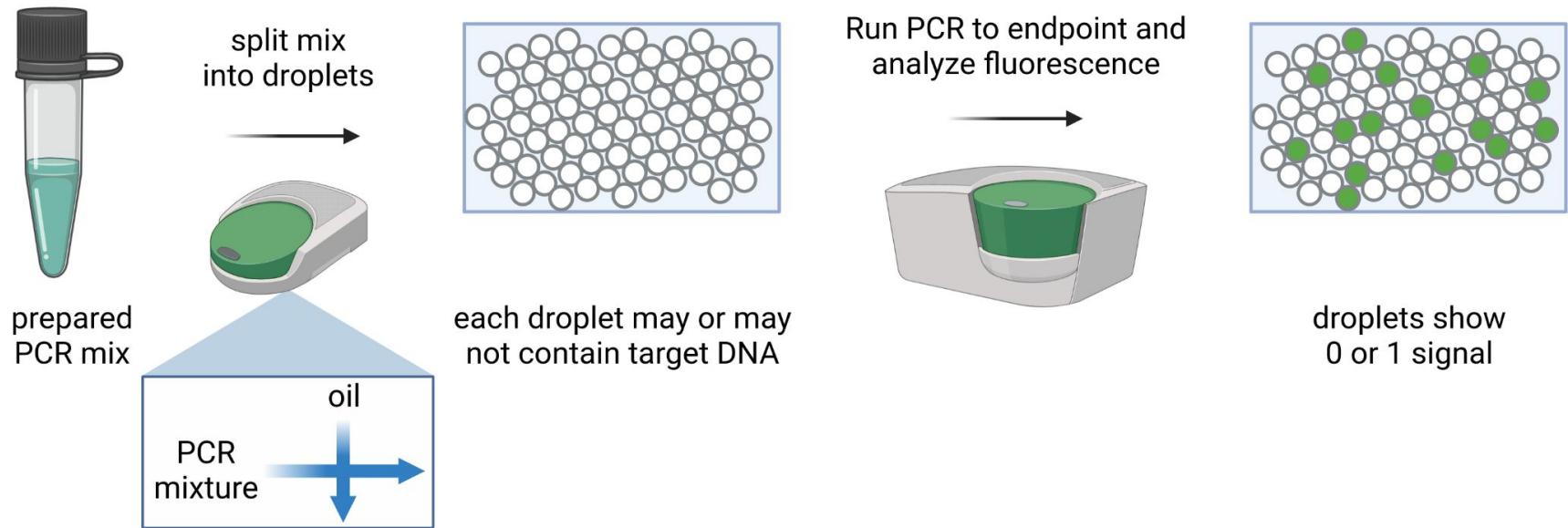


RNA-seq (not RNA sequencing!)

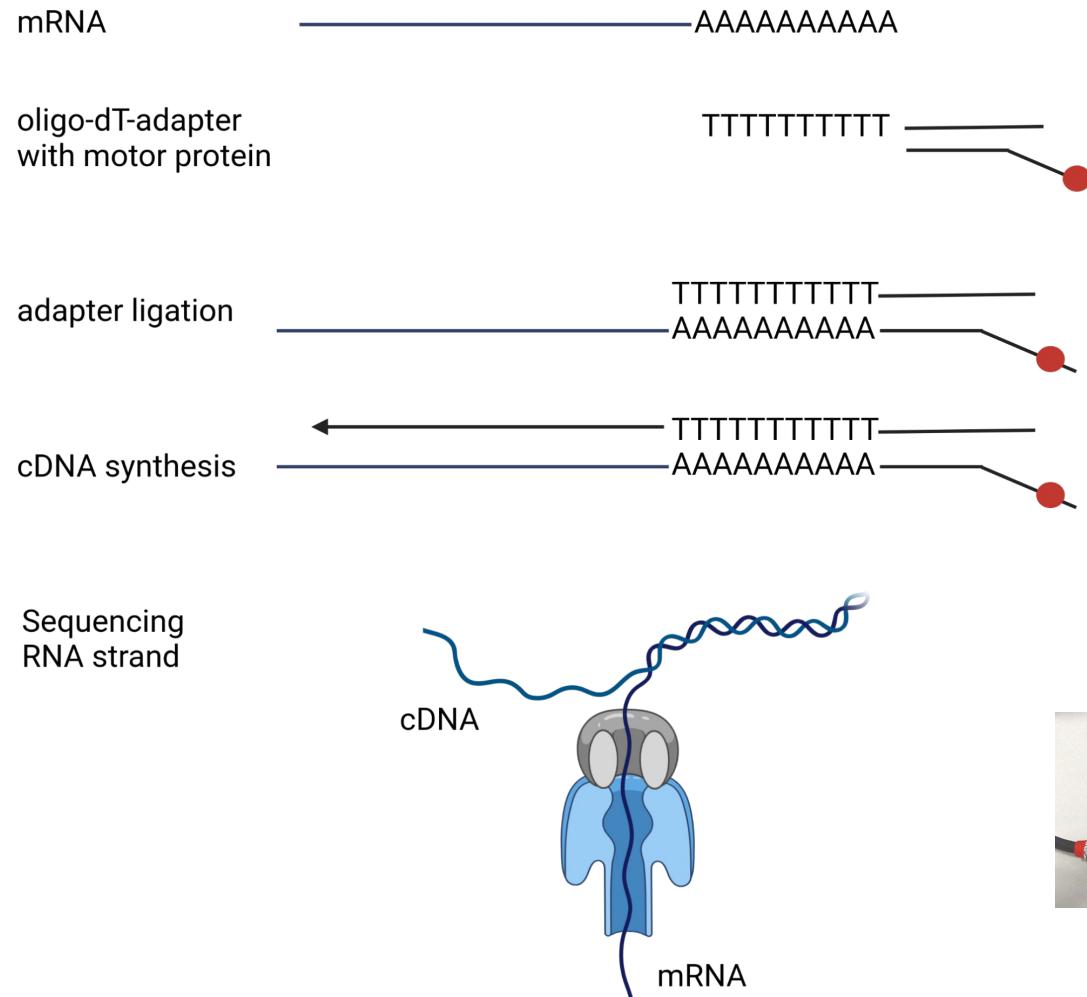


Droplet digital PCR (ddPCR)

- Digital = individual droplets show 0 or 1 signal
- Superior to qPCR in certain applications
- Requires specific hardware (droplet generation and analysis)



Direct RNA sequencing



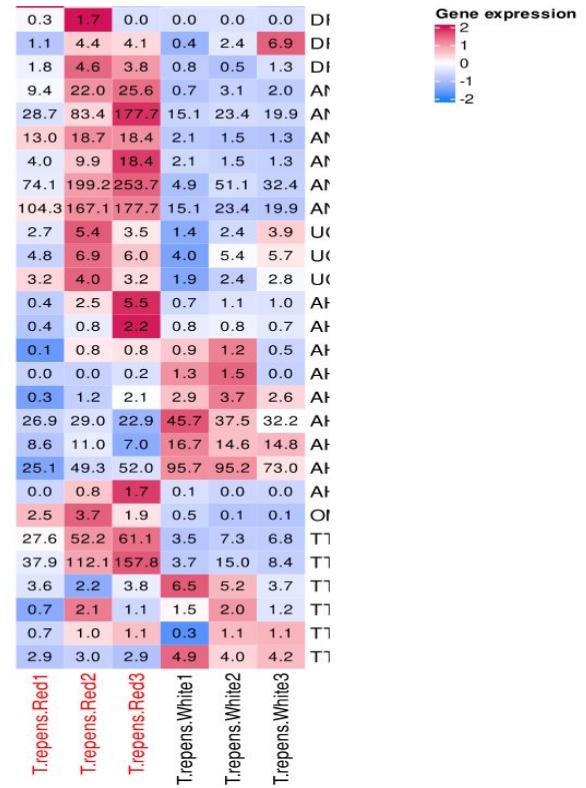
(photo credit: Melina Nowak)

It is all about comparison

- Gene expression is usually analyzed in a comparative way
 - Between samples: fold change differences
 - Between transcript of same samples: transcript per million transcripts
- Absolute measurement of transcript numbers is extremely tricky
- Differentially expressed genes (DEGs)

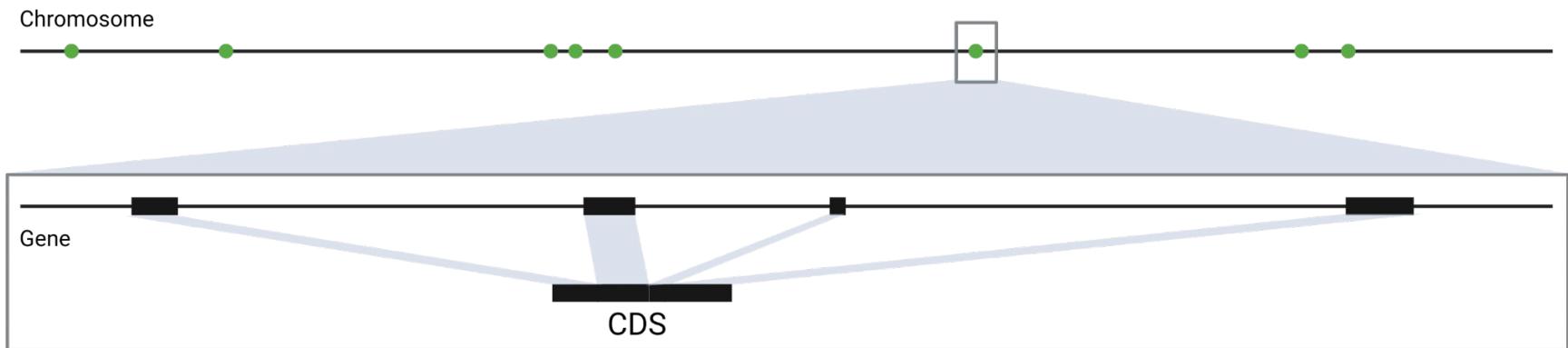
Heatmaps and visualization of gene expression

- Transcriptomic data sets are complex
- Simplified illustration required for interpretation
- Colors help to make data intuitive



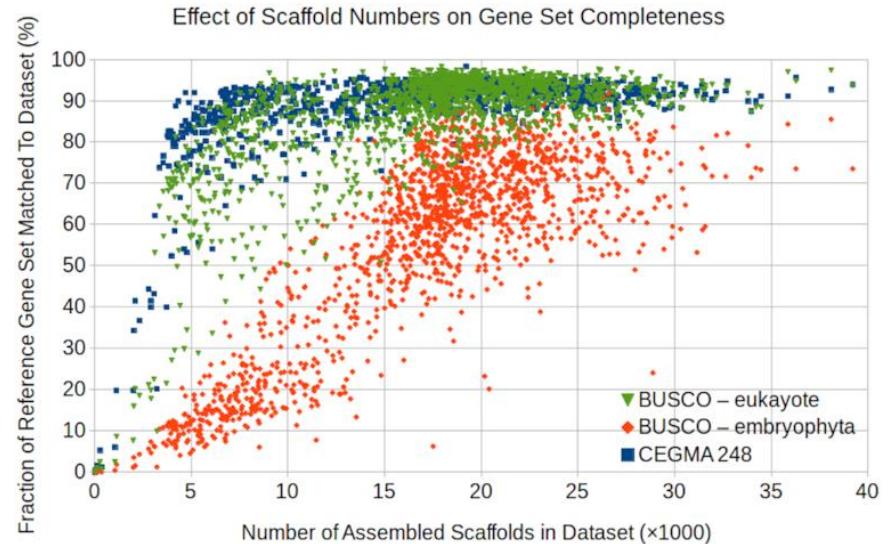
De novo transcriptome assembly

- Coding sequences account for tiny fraction of plant genome
- Transcriptome analysis can be cost-effective compared to genome sequencing
- Reveals genes relevant under certain conditions/at certain time point
- Analysis can be faster and straight forward



1KP

- Phylogenetics project to cover taxonomic width of plants
- Transcriptome assemblies for 1000 (1k) plants
- Large, international consortium for sequencing projects



<http://www.onekp.com/samples/list.php>
<https://doi.org/10.1093/gigascience/giz126>

Data sharing

- Gene Expression Omnibus (GEO): <https://www.ncbi.nlm.nih.gov/geo/>

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.



Keyword or GEO Accession Search

- Sequence Read Archive (SRA): <https://www.ncbi.nlm.nih.gov/sra>



SRA - Now available on the cloud

Sequence Read Archive (SRA) data, available through multiple cloud providers and NCBI servers, is the largest publicly available repository of high throughput sequencing data. The archive accepts data from all branches of life as well as metagenomic and environmental surveys. SRA stores raw sequencing data and alignment information to enhance reproducibility and facilitate new discoveries through data analysis.

- eFP browser: <http://bar.utoronto.ca/efp2/>



Select a plant



Arabidopsis



Tomato



Potato



Poplar

Time for questions!



Questions

1. What is a transcriptome?
2. Which elements form a eukaryotic gene?
3. Which are the three most abundant RNA types?
4. Which methods can be used to measure gene expression?
5. How does RT-qPCR work?
6. How does ddPCR work?