Lecture 4. Differential Gene Expression Profiling

Experimental, computational, and biological interpretation of differential gene expression (DGE).

Dep. Biotechnology Prof. Kenzhebayeva S.S.

Aim of the Lecture. Expected Learning Outcomes

Aim of the Lecture

To introduce the principles, experimental approaches, and bioinformatics methods used in differential gene expression profiling, with emphasis on identifying differentially expressed genes, interpreting their biological functions, and applying gene expression analysis in biotechnology and functional genomics.

Expected Learning Outcomes

- Define differential gene expression and explain its biological significance.
- Describe the major factors influencing gene expression under different physiological and environmental conditions.
- Explain the principles of experimental design, biological replication, and normalization of gene expression data.
- Compare RT-qPCR, microarrays, and RNA-seq approaches for differential expression analysis.
- Describe the main stages of differential gene expression analysis using RNA-seq data.
- Interpret fold change, $\log_2$ fold change, p-values, and adjusted p-values.
- Explain the functions of volcano plots, heatmaps, and clustering in gene expression analysis.
- Describe functional annotation, Gene Ontology enrichment, and pathway analysis.
- Explain approaches for validating differentially expressed genes using RT-qPCR.
- Discuss applications of differential gene expression profiling in plant, medical, microbial, and industrial biotechnology.
    - *Key learning outcome: Understanding how differential gene expression analysis reveals molecular responses to biological and environmental factors and supports candidate gene discovery.*

Lecture Plan

1. Introduction to Differential Gene Expression: Definition, biological significance, and research objectives.
2. Experimental Design and Gene Expression Measurement: Biological replicates, experimental conditions, RNA quality, and data normalization.
3. Technologies for Differential Expression Profiling: RT-qPCR, microarrays, and RNA sequencing.
4. Bioinformatics Analysis of Gene Expression Data: Read processing, quantification, normalization, and statistical methods.
5. Identification and Visualization of Differentially Expressed Genes: Fold change, statistical significance, volcano plots, and heatmaps.
6. Functional Interpretation of Gene Expression Changes: Gene Ontology, pathway enrichment, and regulatory networks.
7. Validation and Applications in Biotechnology: RT-qPCR validation, plant stress responses, disease research, and industrial applications.
8. Challenges and Future Perspectives: Biological variability, batch effects, single-cell differential expression, and multi-omics integration.

Introduction to Differential Gene Expression

Differential gene expression (DGE) refers to differences in the expression levels of genes between biological conditions, developmental stages, tissues, or genotypes.

Main characteristics:

- Gene expression varies depending on cellular functions and physiological conditions.
- Environmental factors, including drought, temperature, salinity, and pathogens, can alter gene expression.
- Changes in gene expression may reflect activation or repression of biological processes.
- Differential expression analysis identifies genes whose transcript abundance differs significantly between experimental groups.
- Biological significance:
- Understanding molecular responses to environmental changes.
- Identifying candidate genes involved in development and stress adaptation.
- Investigating regulatory mechanisms and metabolic pathways.
- Discovering molecular biomarkers associated with biological traits.

Key concept: Differential gene expression analysis reveals how gene activity changes across biological conditions.

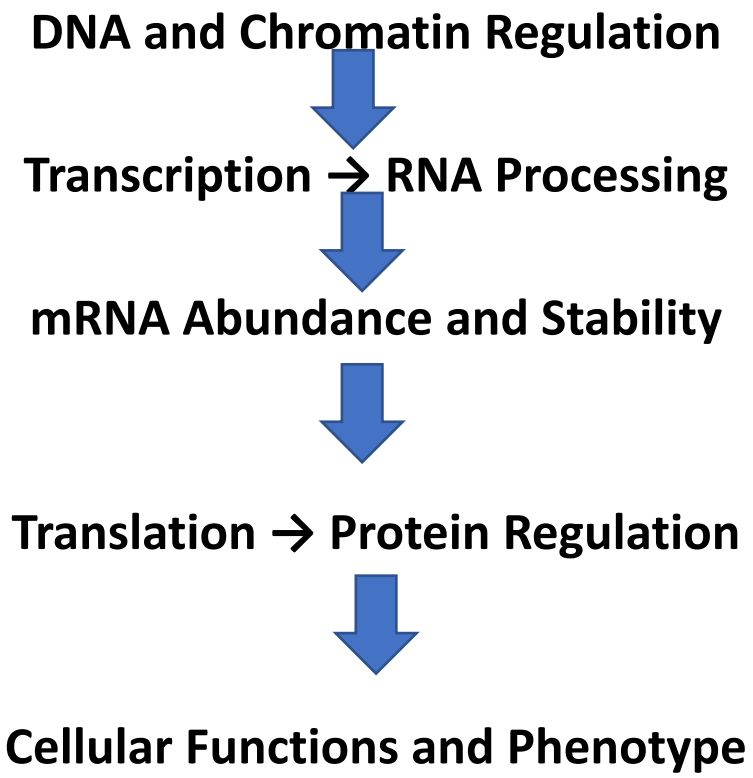
Biological Mechanisms Regulating Gene Expression

Gene expression is regulated at multiple molecular levels, enabling cells to respond to developmental signals and environmental conditions.

Major regulatory mechanisms:

1. Epigenetic regulation: DNA methylation, histone modifications, and chromatin accessibility influence transcriptional activity.
2. Transcriptional regulation: Transcription factors and regulatory DNA elements control transcription initiation and rate.
3. Post-transcriptional regulation: Alternative splicing, RNA processing, RNA stability, and regulatory RNAs influence transcript abundance.
4. Translational regulation: Ribosome recruitment and translation efficiency determine protein synthesis.
5. Post-translational regulation: Protein modification and degradation influence protein activity and stability.

Important distinction: RNA-seq measures RNA abundance; it does not directly measure protein abundance or activity.



Types of Differential Gene Expression

Differential expression can be investigated across different biological and experimental comparisons.

Comparison type	Example
Environmental response	Drought-stressed vs. well-watered wheat
Genotypic variation	Stress-tolerant vs. susceptible wheat genotypes
Tissue-specific expression	Roots vs. leaves
Developmental regulation	Seedling vs. flowering stage
Treatment response	Treated vs. untreated cells

Main expression patterns:

- Upregulation: Higher transcript abundance in one condition relative to another.
- Downregulation: Lower transcript abundance in one condition relative to another.
- No detectable differential expression: No statistically supported expression difference under the experimental conditions.

Key concept: Differential expression is always defined relative to a specified comparison and must be evaluated using appropriate statistical methods. Key concept: Differential expression is always defined relative to a specified comparison and must be evaluated using appropriate statistical methods.

Experimental Design for Differential Gene Expression Analysis

Experimental design is a critical stage of gene expression profiling that determines the reliability and biological interpretation of results.

Main components:

- Research hypothesis: Define the biological question and expected expression changes.
- Experimental groups: Establish control and treatment conditions.
- Biological replication: Include independent biological samples, preferably at least 3 per group, with more when variability is high.
- Sampling strategy: Standardize tissue type, developmental stage, collection time, and environmental conditions.
- Randomization: Minimize systematic differences between experimental groups.
- Technical consistency: Apply standardized RNA extraction, library preparation, and sequencing procedures.

Example in wheat biotechnology: Compare leaf transcriptomes of well-watered and drought-stressed plants at the same developmental stage using independent biological replicates.

Key concept: Appropriate experimental design is essential for distinguishing genuine biological responses from experimental variability.

RNA Isolation and Quality Assessment

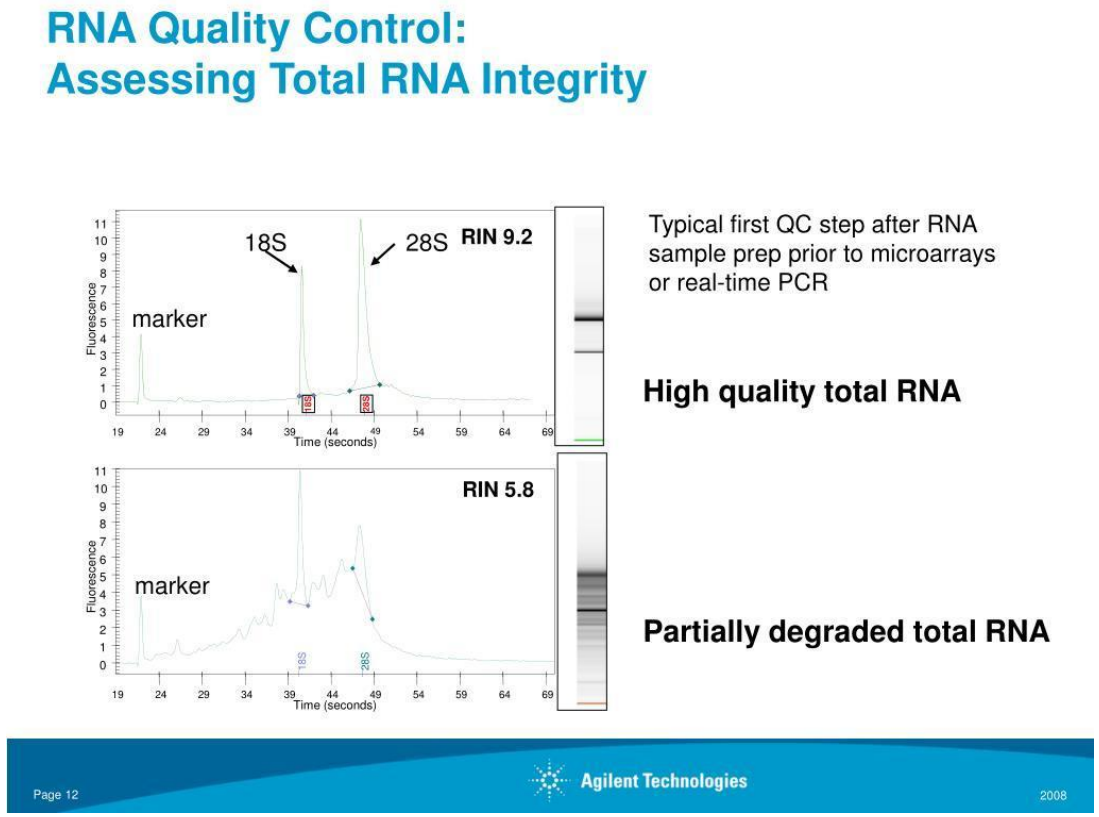
High-quality RNA is essential for reliable gene expression profiling using RNA-seq, RT-qPCR, and microarrays.

Main stages:

1. Sample collection and preservation: Rapid freezing or stabilization to minimize RNA degradation.
2. RNA extraction: Isolation of total RNA using suitable extraction reagents or commercial kits.
3. Removal of contaminants: Elimination of genomic DNA, proteins, phenolic compounds, and polysaccharides.
4. RNA concentration and purity: Evaluation using spectrophotometric and fluorometric methods.
5. RNA integrity assessment: Analysis using electrophoresis or automated capillary systems.

- Important quality indicators:
- A260/A280 ratio: approximately 2.0 for relatively pure RNA.
- A260/A230 ratio: commonly around 2.0–2.2 for low contamination, although sample chemistry affects interpretation.
- RNA Integrity Number (RIN): higher values indicate better integrity; required thresholds depend on the assay and library preparation method.

Key concept: RNA integrity, purity, and consistency directly influence the quality of downstream gene expression measurements.



Technologies for Gene Expression Profiling

Several molecular technologies are available for measuring gene expression, differing in throughput, sensitivity, and analytical requirements.

Main considerations:

- RT-qPCR provides sensitive quantification of selected transcripts.
- Microarrays require prior sequence information for probe design.
- RNA-seq offers a broad dynamic range and can detect previously unannotated transcripts, depending on experimental design and analysis.
- Technology selection depends on research objectives, sample number, budget, and available infrastructure.

  - *Key concept: RNA-seq is widely used for transcriptome-wide discovery, while RT-qPCR remains valuable for targeted gene expression analysis.*

Technology	Principle	Main application
RT-qPCR	Quantification of specific RNA targets following reverse transcription	Targeted expression analysis and validation
DNA microarrays	Hybridization of labeled nucleic acids to predefined probes	Profiling known transcripts
RNA-seq	High-throughput sequencing of RNA-derived libraries	Transcriptome-wide expression profiling

RNA Sequencing (RNA-seq) for Differential Expression Profiling

RNA-seq is a high-throughput sequencing approach used to identify and quantify RNA transcripts across biological samples.

Main workflow:

1. RNA extraction and quality control.
2. mRNA enrichment or ribosomal RNA depletion.
3. RNA fragmentation and sequencing-library preparation.
4. High-throughput sequencing.
5. Read quality control and alignment or transcript-level quantification.
6. Statistical analysis of differential gene expression.
7. Functional annotation and biological interpretation.



RNA Sequencing (RNA-seq) for Differential Expression Profiling

Applications: Identification of stress-responsive genes, analysis of developmental regulation, characterization of mutant lines, and discovery of candidate molecular pathways.

- Example: RNA-seq can be used to investigate transcriptomic differences between drought-tolerant and drought-sensitive wheat genotypes under water-deficit conditions.

Key concept: RNA-seq provides genome-wide expression measurements, while appropriate normalization and statistical analysis are required to identify reliable DEGs.



RNA-seq Data Processing and Quality Control

RNA-seq data processing involves evaluating and preparing sequencing reads for accurate transcript quantification and downstream statistical analysis.

Main stages:

1. Raw data generation: Sequencing reads are typically stored in FASTQ format.
2. Read quality assessment: Evaluation of base quality scores, GC content, sequence duplication, and adapter contamination.
3. Read trimming and filtering: Removal of adapter sequences and low-quality bases when necessary.
4. Post-processing quality control: Verification of sequencing-read quality after filtering.
5. Data organization: Preparation of sample metadata and sequencing files for further analysis.

Common bioinformatics tools:

- FastQC — assessment of sequencing-read quality.
- MultiQC — integration of quality-control reports.
- fastp — read filtering, trimming, and quality assessment.

Key concept: Quality control helps identify technical problems that may compromise gene expression analysis.

Read Alignment and Transcript Quantification

Read alignment determines the genomic or transcriptomic origin of sequencing reads, while quantification estimates transcript or gene abundance.

Major approaches:

1. Reference genome alignment: Mapping reads to a reference genome using splice-aware aligners such as STAR or HISAT2.
2. Transcriptome-based quantification: Estimation of transcript abundance using tools such as Salmon.
3. Gene-level counting: Assignment of aligned reads to annotated genes using featureCounts or similar tools.
4. Expression matrix construction: Organization of gene expression measurements across biological samples.

Expression measurements:

- Raw read counts — commonly used for differential expression analysis.
- TPM (transcripts per million) — useful for describing relative transcript abundance within samples, accounting for transcript length and sequencing depth.

Key concept: Reliable quantification is essential for comparing gene expression between experimental conditions.

Normalization of Gene Expression Data

Normalization adjusts gene expression measurements to reduce technical differences between samples and improve statistical comparability.

Why normalization is necessary:

- Sequencing depth may differ between samples.
- RNA composition can vary between biological conditions.
- Highly expressed genes may influence library composition.
- Technical variation can obscure genuine biological differences.

Common normalization approaches:

1. DESeq2 median-of-ratios: Estimates sample-specific size factors from gene count distributions.
2. TMM (trimmed mean of M-values): Adjusts for differences in RNA composition between samples.
3. TPM: Accounts for transcript length and library size but is not generally used directly as input to standard DESeq2 differential expression testing.

Important distinction: Normalization does not automatically eliminate batch effects or biological confounding.

Key concept: Appropriate normalization improves the reliability of differential expression comparisons.

Statistical Analysis of Differential Gene Expression

Differential expression analysis identifies genes showing statistically supported changes in RNA abundance between experimental conditions.

Main analytical parameters:

1. Fold change (FC): Ratio of expression levels between two conditions.
2. Log_2 fold change (log_2FC): Logarithmic measure of the direction and magnitude of expression differences.
3. p-value: Statistical evidence against the null hypothesis of no expression difference.
4. Adjusted p-value (padj): p-value corrected for multiple testing, commonly using the Benjamini–Hochberg procedure.
5. False discovery rate (FDR): Expected proportion of false discoveries among the rejected null hypotheses, under the relevant statistical assumptions.

Common tools:

1. DESeq2 — negative binomial statistical modeling.
2. edgeR — count-based differential expression analysis.
3. limma-voom — linear modeling with precision weights.
4. Illustrative thresholds: padj < 0.05 and $|log_2FC| \geq 1$, selected according to study objectives.

Key concept: Statistical significance and biological effect size should be evaluated together.

Identification of Upregulated and Downregulated Genes

Differentially expressed genes (DEGs) are classified according to the direction and magnitude of their expression changes.

Upregulated

+2.0

log_2FCHigher expression in treatment

Unchanged

0.0

log_2FCNo estimated change

Downregulated

–2.0

log_2FCLower expression in treatment

Illustrative log_2 fold changes relative to a control condition. Statistical significance must be assessed separately.

Interpretation:

1. $log_2FC = +1$ indicates a twofold increase.
2. $log_2FC = -1$ indicates a twofold decrease.
3. $log_2FC = 0$ indicates no estimated expression change.

Example in wheat: Drought exposure may increase the expression of selected dehydration-responsive genes while reducing the expression of some photosynthesis-related genes.

Key concept: The direction of expression change depends on how the experimental contrast is defined.

Differential Expression Analysis Using DESeq2

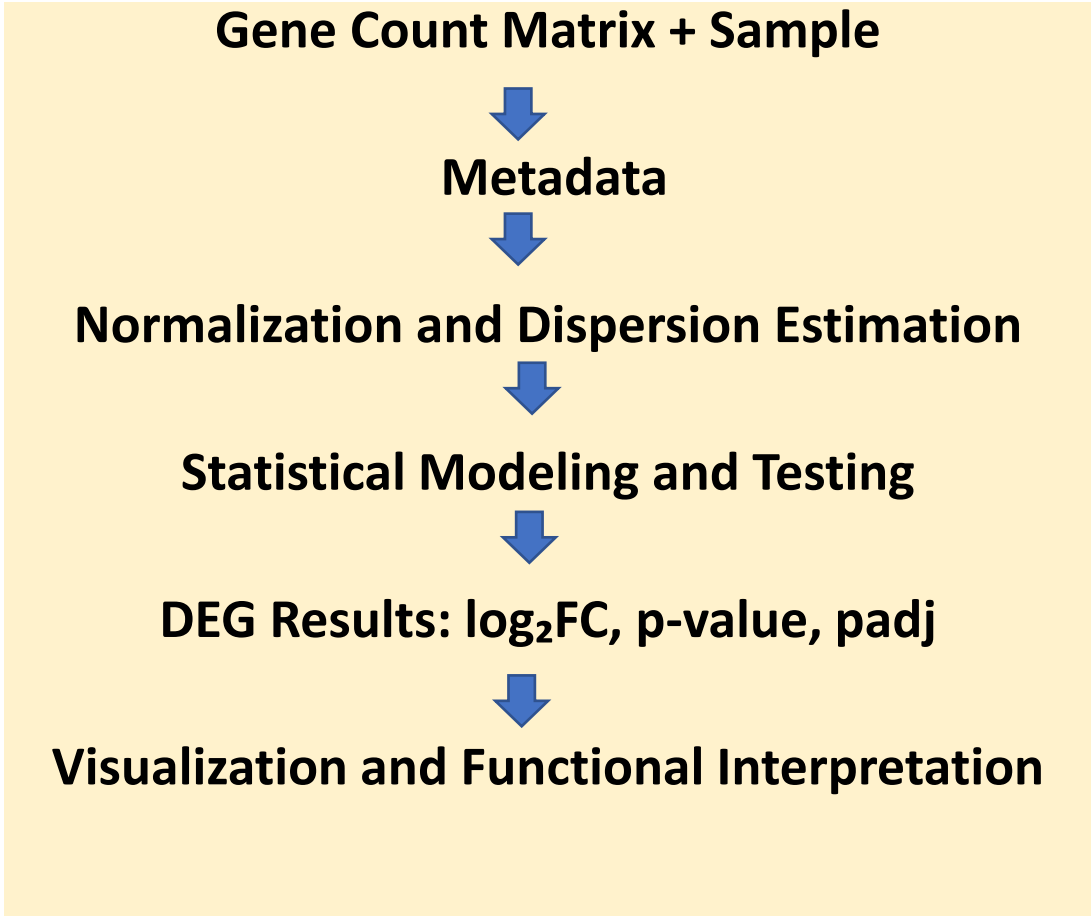
DESeq2 is a widely used R/Bioconductor package for identifying differentially expressed genes from RNA-seq count data.

Main analytical workflow:

1. Import a gene-level count matrix and sample metadata.
2. Define the experimental design and comparison groups.
3. Estimate normalization size factors.
4. Estimate gene-wise dispersion parameters.
5. Fit negative binomial generalized linear models.
6. Perform statistical hypothesis testing.
7. Adjust p-values for multiple comparisons.
8. Generate tables of DEGs and diagnostic visualizations.

Main outputs:

1. Normalized gene counts.
2. Log_2 fold changes.
3. Statistical significance values.
4. Differentially expressed gene lists.
5. Results for downstream functional analysis.



Volcano Plot: Visualization of Differentially Expressed Genes

A volcano plot is a graphical method used to visualize both the magnitude and statistical significance of gene expression changes between experimental conditions.

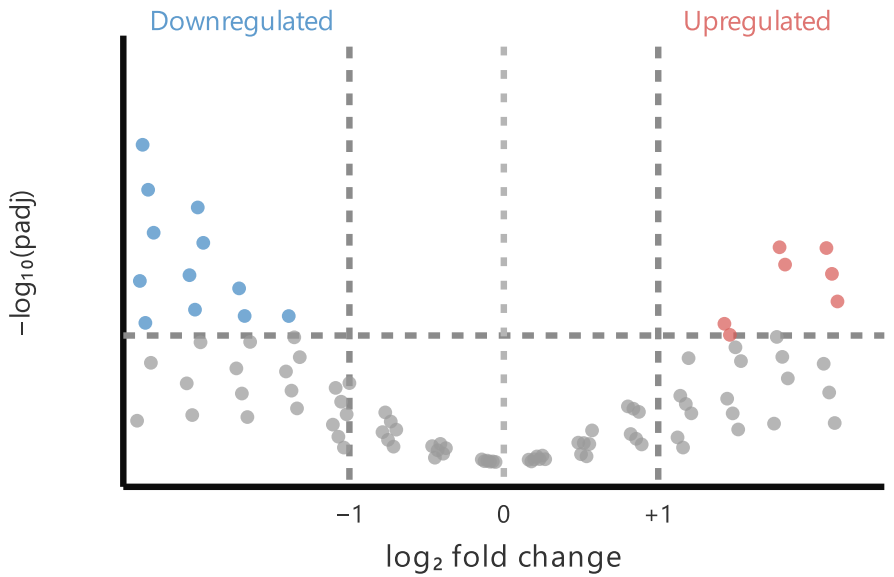
Main components:

- X-axis: $\log_2$ fold change ($\log_2$FC), indicating the direction and magnitude of expression changes.
- Y-axis: $-\log_{10}$(adjusted p-value), indicating statistical significance.
- Upregulated genes: Located on the right side of the plot.
- Downregulated genes: Located on the left side.
- Non-significant genes: Generally located below the significance threshold.

Applications:

- Identification of significantly differentially expressed genes (DEGs).
- Comparison of gene expression responses between experimental conditions.
- Selection of candidate genes for further functional analysis.

Key concept: Volcano plots combine statistical significance and expression effect size in a single visualization.



Simulated data. Dashed lines indicate illustrative fold-change and significance thresholds.

Heatmaps: Visualization of Gene Expression Patterns

A heatmap is a graphical representation of gene expression data in which numerical values are displayed using a color scale.

Main characteristics:

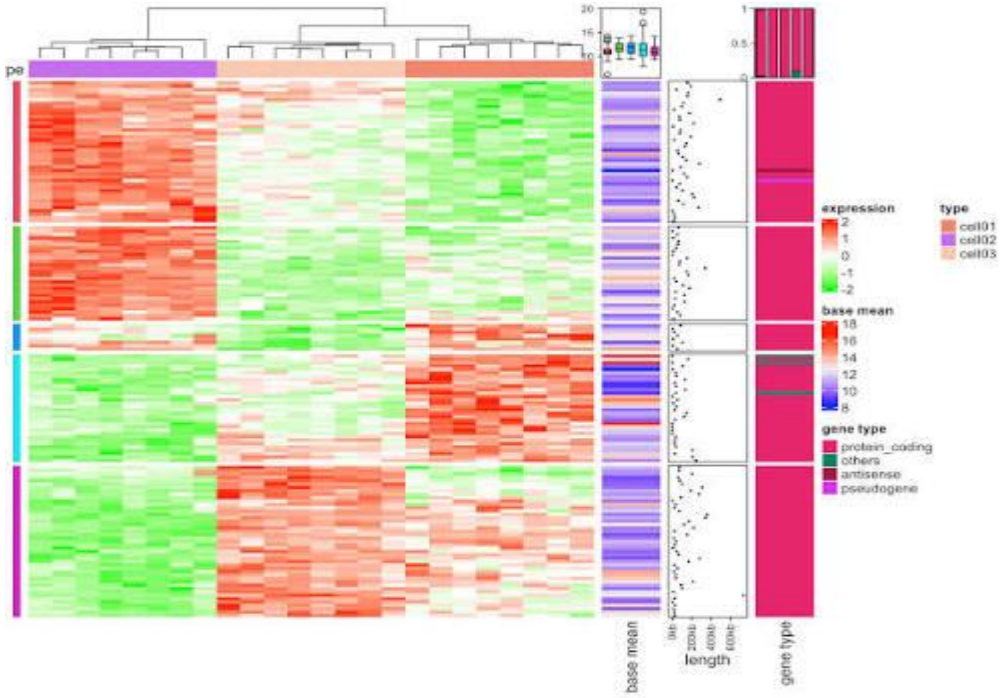
1. Rows typically represent genes.
2. Columns represent biological samples or experimental conditions.
3. Colors indicate relative expression levels.
4. Gene and sample clustering can reveal similarities and differences in expression profiles.

Applications:

- Visualization of differentially expressed genes.
- Identification of co-expression patterns.
- Comparison of control and treated samples.
- Investigation of tissue-specific or genotype-dependent expression.

Important consideration: Heatmaps often display normalized, log-transformed, or row-standardized expression values. The color scale must clearly identify the measurement used.

Illustrative gene expression heatmap



Clustering Analysis of Gene Expression Data

Clustering analysis is an unsupervised statistical approach used to group genes or samples according to similarities in their expression profiles.

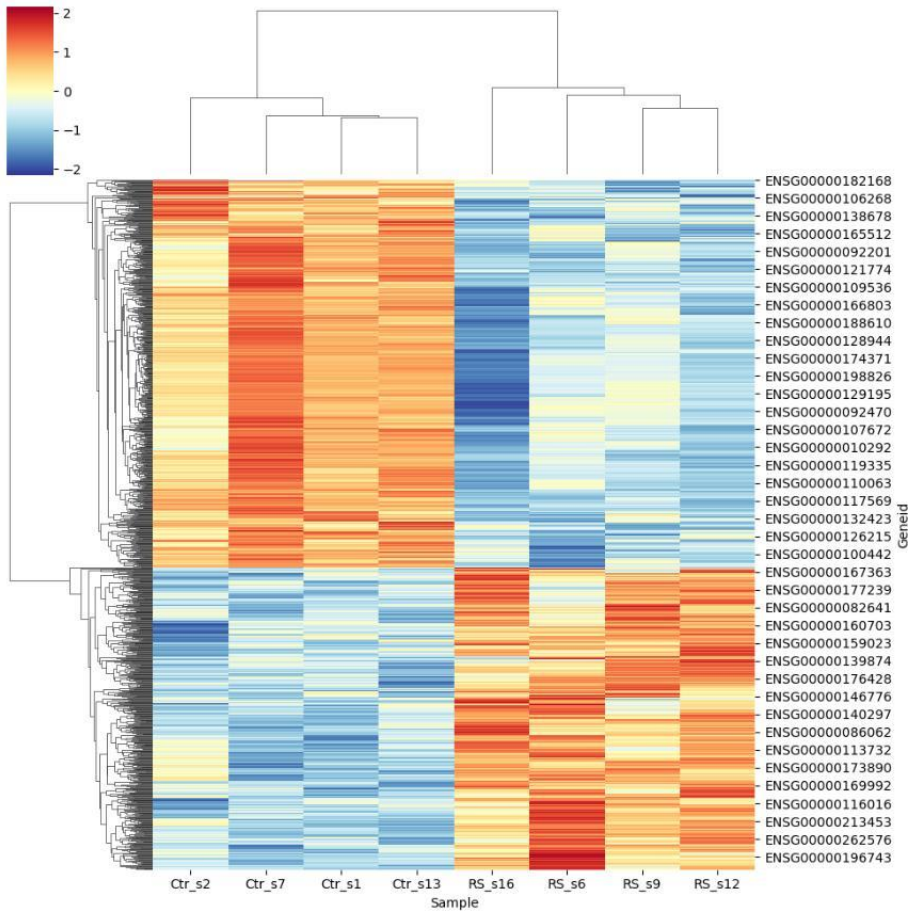
Major clustering methods:

1. Hierarchical clustering: Organizes genes or samples into a tree-like structure called a dendrogram.
2. K-means clustering: Divides expression profiles into a predefined number of clusters.
3. Consensus clustering: Evaluates the stability of clusters across repeated analyses or resampling.

Applications:

- Identification of co-expressed gene groups.
- Classification of biological samples.
- Detection of treatment-specific expression patterns.
- Exploration of potential regulatory relationships.
- Important considerations:
- Results depend on the distance metric, data transformation, and clustering method.
- Genes in the same cluster are not necessarily regulated by the same mechanism.
- Clustering reveals patterns but does not establish causality.

Key concept: Clustering identifies similarities among expression profiles and facilitates the discovery of biologically meaningful patterns.



Biological Interpretation of Gene Expression Patterns

Biological interpretation connects differential expression results with molecular functions, regulatory processes, and physiological responses.

Main interpretation approaches:

1. Gene annotation: Identification of gene functions using reference databases.
2. Gene Ontology (GO): Classification of genes according to biological processes, molecular functions, and cellular components.
3. Pathway enrichment: Identification of biological pathways overrepresented among DEGs.
4. Co-expression analysis: Investigation of genes exhibiting correlated expression patterns.
5. Candidate gene prioritization: Selection of genes for experimental validation.

Key concept: Differential expression provides evidence of transcriptional responses, but additional physiological, biochemical, and functional experiments are needed to establish biological mechanisms.

Example: Wheat responses to drought stress

Expression pattern	Possible biological interpretation
Increased expression of selected P5CS genes	Potential activation of proline biosynthesis
Altered expression of CAT and APX genes	Changes in antioxidant defense regulation
Increased expression of some LEA genes	Activation of dehydration-associated protective responses
Reduced expression of selected photosynthesis-related genes	Possible adjustment of photosynthetic processes

Functional Enrichment Analysis and Gene Ontology (GO)

Functional enrichment analysis is a bioinformatics approach used to identify biological functions and processes statistically overrepresented among differentially expressed genes.

Gene Ontology (GO) classifies gene functions into three major categories:

1. **Biological Process (BP):** Biological activities such as photosynthesis, stress responses, and cellular metabolism.
2. **Molecular Function (MF):** Molecular activities such as enzyme catalysis, DNA binding, and transporter activity.
3. **Cellular Component (CC):** Cellular locations and structures, including the nucleus, chloroplasts, and mitochondria.

Main analytical stages:

1. Identify statistically significant DEGs.
2. Associate genes with GO annotations.
3. Perform enrichment analysis using an appropriate background gene set.
4. Apply multiple-testing correction.
5. Interpret significantly enriched GO terms.

Common tools:

clusterProfiler, g:Profiler, and agriGO.

Key concept: GO enrichment reveals biological functions associated with gene expression changes but does not directly demonstrate changes in biochemical activity.

Three categories of Gene Ontology

Biological Process

Stress response
Photosynthesis
Metabolism

Molecular Function

Catalytic activity
DNA binding
Transport

Cellular Component

Nucleus
Chloroplast
Mitochondrion

Biological Pathway Enrichment Analysis

Pathway enrichment analysis identifies metabolic and signaling pathways that are statistically associated with a set of differentially expressed genes.

Major pathway databases:

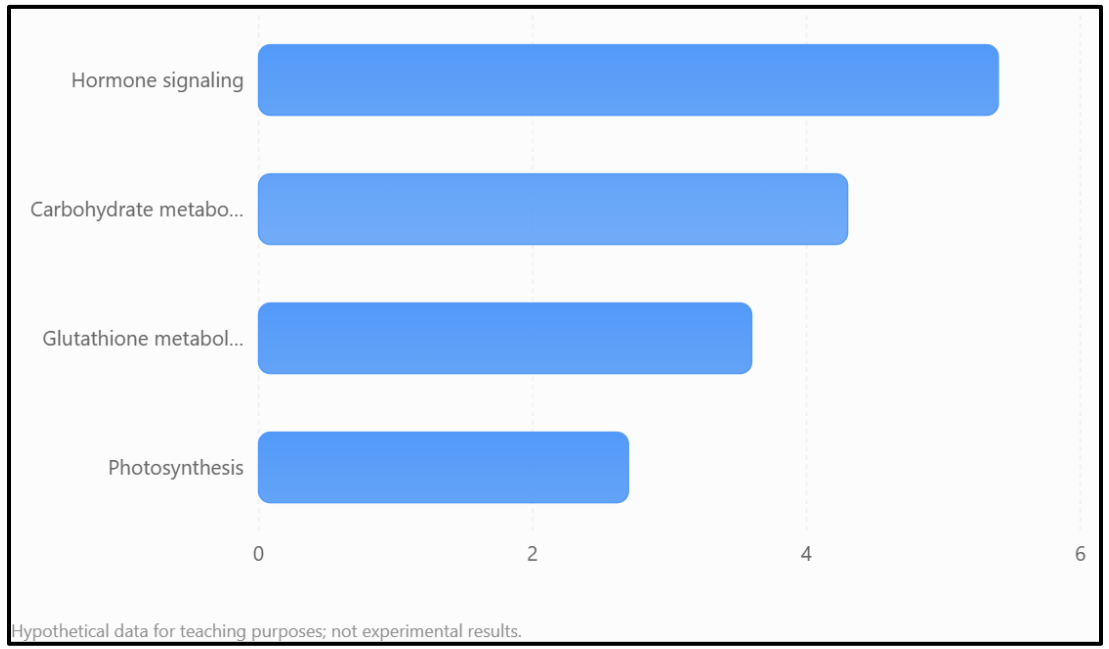
- KEGG: Metabolic pathways, signaling networks, and molecular interactions.
- Reactome: Curated biological pathways and molecular reactions.
- Plant Reactome: Plant-focused pathway information and comparative analysis.
- MapMan/Mercator: Functional classification and visualization of plant gene expression data.

Main applications:

- Identification of metabolic pathways affected by experimental treatments.
- Investigation of stress signaling and defense-related processes.
- Analysis of photosynthesis, respiration, and secondary metabolism.
- Prioritization of pathways for further molecular and biochemical investigations.

Example in wheat: Differential expression analysis under drought conditions may identify enrichment of pathways related to hormone signaling, carbohydrate metabolism, and antioxidant-associated processes.

Key concept: Pathway enrichment helps interpret coordinated transcriptional responses, although pathway activity cannot be established from transcript abundance alone.



Illustrative pathway enrichment results

Gene Regulatory Networks and Co-expression Analysis

Gene regulatory networks (GRNs) represent relationships between regulatory genes, transcription factors, and their potential target genes.

Main components:

1. Nodes: Genes, transcription factors, or regulatory molecules.
2. Edges: Regulatory interactions or inferred relationships between network components.
3. Hub genes: Highly connected genes that may have important regulatory roles.
4. Co-expression modules: Groups of genes exhibiting correlated expression patterns.

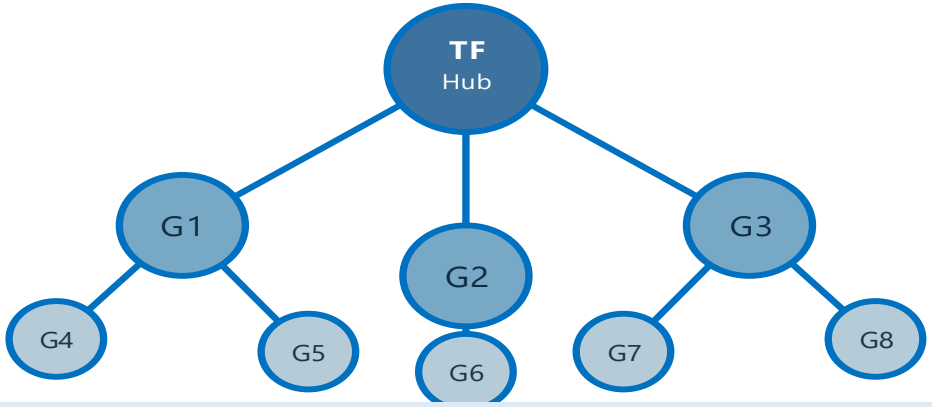
- **Major analytical approaches:**
- Weighted Gene Co-expression Network Analysis (WGCNA).
- Transcription factor binding motif analysis.
- Regulatory network inference.
- Integration of transcriptomic and other omics data.

Important distinction: Co-expression indicates statistical association between gene expression profiles; it does not establish direct transcriptional regulation.

Key concept: Network analysis supports the identification of candidate regulatory relationships and genes potentially involved in coordinated biological respons

Applications in biotechnology:

1. Identification of candidate regulatory genes.
2. Investigation of molecular responses to environmental stress.
3. Discovery of co-expression modules associated with physiological traits.
4. Prioritization of candidate genes for functional validation.



Schematic network showing a candidate transcription factor (TF), connected genes, and potential regulatory relationships. Connections are illustrative, not experimentally validated.

Validation of Differential Gene Expression Using RT-qPCR

Reverse transcription quantitative PCR (RT-qPCR) is a sensitive molecular technique used to quantify selected RNA transcripts and independently assess expression patterns identified by RNA-seq.

Main stages of RT-qPCR validation:

1. Candidate gene selection: Choose representative upregulated and downregulated genes identified by RNA-seq.
2. RNA isolation and cDNA synthesis: Extract high-quality RNA and convert it into complementary DNA.
3. Primer design: Develop gene-specific primers and verify amplification specificity and efficiency.
4. Reference gene selection: Use experimentally validated, stably expressed reference genes for normalization.
5. Quantitative analysis: Determine Cq values and calculate relative gene expression using an efficiency-appropriate method, such as $2^{-\Delta\Delta Cq}$ when its assumptions are satisfied.
6. Comparison with RNA-seq: Evaluate agreement in the direction and magnitude of expression changes.

Applications:

- Confirmation of selected transcriptomic expression patterns.
- Analysis of candidate genes across additional biological samples.
- Investigation of tissue-specific and stress-responsive gene expression.

RT-qPCR validation workflow



RNA-seq: Identification of DEGs



Selection of candidate genes



RNA extraction and cDNA synthesis



RT-qPCR amplification and normalization



Comparison of RNA-seq and RT-qPCR results

Key concept: RT-qPCR provides targeted, complementary evidence for RNA-seq findings but does not independently establish gene function.

Applications of Differential Gene Expression in Plant Biotechnology

Differential gene expression profiling is widely applied to investigate molecular mechanisms controlling plant growth, development, environmental adaptation, and crop productivity.

Major applications:

- Abiotic stress tolerance: Identification of genes responding to drought, salinity, heat, and nutrient deficiency.
- Disease resistance: Analysis of defense-associated genes activated or repressed during pathogen infection.
- Plant development: Investigation of genes controlling germination, flowering, root growth, and grain development.
- Nutritional quality: Identification of candidate genes involved in Fe, Zn, protein, and carbohydrate metabolism.
- Mutation breeding: Characterization of transcriptional differences between mutant lines and parental genotypes.
- Molecular breeding: Prioritization of candidate genes and regulatory pathways for functional studies.

Key concept: Differential expression analysis helps identify candidate genes associated with stress adaptation, but physiological and functional validation is required to confirm their roles.

Example: Drought responses in wheat

Candidate gene group	Potential biological role
P5CS	Proline biosynthesis and osmotic adjustment
CAT, APX	Antioxidant defense
DREB, NAC	Stress-responsive transcriptional regulation
LEA	Cellular protection during dehydration
TaVIT-related genes	Intracellular iron transport and homeostasis

Applications of Differential Gene Expression in Other Areas of Biotechnology

Differential gene expression profiling provides important information about molecular responses and regulatory mechanisms across biological systems.

Major applications:

- Medical biotechnology: *Identification of disease-associated expression signatures, diagnostic biomarkers, and potential therapeutic targets.*
- *Pharmaceutical biotechnology*: *Investigation of cellular responses to drugs and mechanisms of drug resistance.*
- Microbial biotechnology: *Analysis of genes controlling microbial growth, stress adaptation, and metabolite biosynthesis.*
- Industrial biotechnology: *Optimization of microbial strains for enzyme production, fermentation, and bioprocess efficiency.*
- Environmental biotechnology: *Investigation of gene expression responses to pollutants, heavy metals, and environmental stress.*
- Animal biotechnology: *Analysis of genes associated with development, immunity, productivity, and disease susceptibility.*

Illustrative examples:

- Comparing gene expression in antibiotic-resistant and susceptible bacterial strains.
- Investigating transcriptional responses of microorganisms during industrial fermentation.
- Identifying disease-associated expression patterns in human tissues.

Key concept: Differential gene expression profiling connects molecular responses with biological functions and supports innovation across diverse areas of biotechnology.

Challenges and Future Perspectives of Differential Gene Expression Profiling

Major challenges:

- Biological variability: Differences between individuals, tissues, developmental stages, and environmental conditions.
- Technical limitations: RNA degradation, sequencing biases, batch effects, and insufficient biological replication.
- Statistical challenges: Multiple testing, low-abundance transcripts, and interpretation of small expression changes.
- Functional interpretation: Differential expression does not necessarily indicate changes in protein abundance, enzyme activity, or biological function.
- Data integration: Combining transcriptomic results with other molecular and physiological datasets remains computationally challenging.

Future perspectives:

- Single-cell and spatial transcriptomics for cell-specific gene expression analysis.
- Long-read RNA sequencing for improved transcript and alternative splicing characterization.
- Integration of transcriptomics with proteomics, metabolomics, and phenomics.
- Artificial intelligence and machine learning for gene expression pattern analysis.
- Functional validation of candidate genes using genome editing and other molecular approaches.

Key message: Advances in transcriptomic technologies and integrated data analysis will improve our understanding of gene regulation and biological complexity.

Conclusion

1. Differential gene expression profiling is a fundamental approach for investigating transcriptional responses to biological and environmental factors.
2. RNA-seq enables transcriptome-wide identification and quantification of gene expression differences.
3. Appropriate experimental design, RNA quality assessment, normalization, and statistical analysis are essential for reliable results.
4. Volcano plots, heatmaps, and clustering facilitate visualization and interpretation of gene expression patterns.
5. Functional enrichment and regulatory network analyses help identify biological processes and candidate regulatory mechanisms.
6. RT-qPCR provides complementary validation of selected RNA-seq expression patterns.
7. Differential expression profiling supports gene discovery and research in plant, medical, microbial, industrial, and environmental biotechnology.

Key message: Differential gene expression analysis connects transcriptomic changes with biological processes and provides a foundation for functional genomics and biotechnology applications.

Control Questions

1. What is differential gene expression, and why is it important?
2. What molecular mechanisms regulate gene expression?
3. Why are biological replicates essential for RNA-seq experiments?
4. What are the main stages of RNA-seq data processing?
5. What is the purpose of normalization in differential expression analysis?
6. What do $\log_2$ fold change, p-value, and adjusted p-value indicate?
7. How does DESeq2 identify differentially expressed genes?
8. What information can be obtained from volcano plots and heatmaps?
9. What is the purpose of hierarchical clustering in gene expression analysis?
10. How are Gene Ontology and pathway enrichment analyses used to interpret DEGs?
11. Why is RT-qPCR used to validate selected RNA-seq results?
12. How can differential expression profiling contribute to crop improvement and other biotechnology applications?

Recommended Literature

- 1. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15, 550. https://doi.org/10.1186/s13059-014-0550-8
- 2. Conesa, A., et al. (2016). A survey of best practices for RNA-seq data analysis. Genome Biology, 17, 13. https://doi.org/10.1186/s13059-016-0881-8
- 3. Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. F1000Research, 4, 1521. https://doi.org/10.12688/f1000research.7563.2
- 4. Zhu, A., Ibrahim, J. G., & Love, M. I. (2019). Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences. Bioinformatics, 35, 2084–2092. https://doi.org/10.1093/bioinformatics/bty895
- 5. Bustin, S. A., et al. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry, 55, 611–622. https://doi.org/10.1373/clinchem.2008.112797
- 6. Wu, T., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation, 2, 100141. https://doi.org/10.1016/j.xinn.2021.100141