

“Lecture” 10

Assignment 2 working period

Today

- Assignment 2
 - Tips
 - Working period

Assignment 2

The goal of this assignment is to learn how to analyze differential expression in a bulk transcriptomics experiment, perform functional annotations, and communicate results. You will choose to either use data from a study that examined yeast biofilm (velum) development during wine aging, or you will find and source a dataset of your choice (transcriptomics data, including at least 3 groups).

The yeast dataset is available here: <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2020.00538/full>

Velum is a membrane that grows on top of the fermenting juice. The study sampled three stages in velum development and had three replicates for each stage. We can assume batches were assigned for stage sampling using a completely randomized experimental design.

Your report must include the following:

- A statistical comparison of your experimental groups, including lists of differentially expressed genes with magnitude of change and multiple comparison-adjusted p values.
- A functional analysis, categorizing your genes by functional annotation and comparing your groups with ORA or GSEA

Grade breakdown (/30)

1. **Introduction (5):** Brief background of the biological topic. Justification for the planned tools and methods.
2. **Methods (5):** Detailed description of methods, including citations for tools used. Includes the code used, which may be provided in a separate file.
3. **Results (10):** Results section in standard paper format, including plots/figures/tables as necessary. Should include at least one figure summarizing overall data structure, at least one figure showing top differentially expressed genes, and at least one figure showing the results of your functional annotation and ORA/GSEA.
4. **Discussion (5):** Discuss the biological implications of your results, with appropriate citations. It is sufficient to discuss only a few genes/functional annotations.
5. **Formatting (5):** All code and results present and well-commented, spelling and grammar correct, proper use of Git.

Sections (10 total)	Excellent	Good	Satisfactory	Needs Improvement
Introduction (5) • Biological background (3) • Methods comparison (2)	(5) Clearly outlines the goals and objectives. Shows understanding of biological context. Understands limitations and trade-offs of methods. Cites 5+ relevant sources, showing clear understanding of material.	(4) Clearly outlines the goals and objectives. Compares methods in some detail and supports a choice of method. Cites relevant sources supporting conclusions.	(3) Compares methods but may be missing detail or focused on the wrong metrics. Lacks a clear understanding of biological context. Cites relevant source but connection to conclusions is weak – sources may not compare methods.	(1-2) Minimal details included in background.
Methods (5) • References planned methods and tools with citations (2) • Includes quantification, differential expression, functional annotation (1) • Includes full code used (2) • Everything except your raw data or other files over 100 MB must be uploaded.	(5) Clearly details starting data, QC, quantification, differential expression, functional annotation. Describes steps taken in planned order. Includes version details and appropriate citations. Includes any specified parameters that differ from the default.	(4) Describes tools used. May lack specific details on versions or be missing appropriate citations. Mostly includes version details and appropriate citations. Includes any specified parameters that differ from the default.	(3) Describes data and QC, as well as assembler and aligner. Missing information on specific parameters Some code may be unclear, or missing.	(1-2) Missing details or described software is inappropriate to the task. Missing code.
Organization and Formatting (5) • Uploaded to GitHub or other repository • Grammar and spelling • Citations are formatted correctly	(5) Well-organized. Proper citations. Excellent use of GitHub's markdown. Includes multiple commits, indicating use of Git for versioning. Well-formatted figures and captions.	(4) Minimal formatting issues. Includes multiple commits, indicating use of Git for versioning.	(3) Disorganized, or missing citations. Project uploaded as single commit.	(1-2) Difficult to read. Poor organization.

Sections (10 total)	Excellent	Good	Satisfactory	Needs Improvement
<p>Results (10)</p> <p>Includes a differential expression analysis. (5)</p> <p>Includes functional annotation. (5)</p> <p>Paragraph form results section in standard scientific style, referencing figures.</p> <p>Effective visualizations of overall data structure, differential expression, functional annotation.</p> <p>Figures include plots with captions.</p>	<p>(10)</p> <p>Results are meaningful and well-presented in paragraph form.</p> <p>Includes at least three plots showcasing results.</p> <p>Plots are well organized</p>	<p>(8)</p> <p>Results lack some depth.</p> <p>Included three plots but could be improved.</p>	<p>(6)</p> <p>Results may be missing or slightly inaccurate.</p> <p>Plots are missing or uninformative.</p>	<p>(0-5)</p> <p>Minimal details included</p> <p>Uninformative results.</p> <p>Plots not included or inaccurate.</p>
<p>Discussion (5)</p> <p>Interpretation of results and analysis.</p> <p>Discusses the results from the DE analysis.</p> <p>Discusses specific genes and functional groups, as well as the biological implications of the change in expression.</p>	<p>(5)</p> <p>Effectively analyses the results and the potential biological effects.</p> <p>Clear and concise, in the style of a scientific discussion section.</p> <p>Relates findings to multiple experimental papers (5+), i.e. studies on the genes mentioned, or on the organism and condition.</p>	<p>(4)</p> <p>Mostly organized discussion on major topics.</p> <p>Some inaccuracies in interpreting results.</p> <p>Discusses at least one major topic related to findings and relates to the literature.</p>	<p>(3)</p> <p>Basic discussion and interpretations of results.</p> <p>Could use more details</p>	<p>(0-2)</p> <p>Missing details</p> <p>Lacks meaningful discussion regarding findings</p>

Two options:

- Use a provided yeast dataset (Mardanov et al. 2020)
- Find another dataset you're interested in
 - Must be a bulk transcriptomics experiment
 - Must include at least 3 groups

Yeast data

- <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA592304>

Stage	Sample ID	SRA accession
Early biofilm	IL20	SRR10551665
	IL21	SRR10551664
	IL22	SRR10551663
Thin Biofilm	IL23	SRR10551662
	IL24	SRR10551661
	IL25	SRR10551660
Mature Biofilm	IL29	SRR10551659
	IL30	SRR10551658
	IL31	SRR10551657

Additional metadata

Available from <https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2020.00538/full>

Sample designation	Early biofilm	Thin biofilm	Mature biofilm
Days from inoculation to sampling	38	83	109
Ethanol% (v/v)	12.4	10.8	9.6
Volatile acidity (g/l)*	0.2	0.2	0.1
Total acidity (g/l)**	7.8	7.4	7.0
Aldehydes (mg/l)	382.8	531.3	668.8
Acetals (mg/l)	147.2	253.7	280.3
pH	3.6	3.6	3.6
Glucose (g/l)	0.2	0.1	<0.1
Fructose (g/l)	<0.1	<0.1	<0.1
Oligosaccharides (g/l)	0.3	0.2	0.2
Glycerol (g/l)	8.5	7.9	6.8

*Expressed in grams of tartaric acid per liter. **Expressed in grams of acetic acid per liter.

Consider

- How will you (or will you) quality control?
- Where will you find a reference genome/transcriptome?
- How will you quantify your dataset?
- What statistical comparisons will you make?
 - Pairwise comparisons? Or a time course?
- Which statistical software will you use?
- Where will you get your functional annotations?
- How will you plot your important results?

Tips

- Start early – some steps may be computationally intensive, or difficult to get working
- Don't forget to include figure captions and paragraph-form results
- Pick a few of your most significant results to focus on
- Your report doesn't have to be extremely long, but it does need to include all the features mentioned in the rubric
- If you don't have the code working a week before the due date, please ask me for help!