

gene-expression

Diego Cesar Villa Almeyda

2025-06-21

Table of contents

Preface	3
1 Introduction	4
2 Methods	5
2.1 Experimental design and data	5
2.2 Pre-filtering	6
2.3 Data transformation	6
2.4 Exploratory analysis	6
2.5 Differential expression analysis	6
2.6 Motif enrichment analysis	6
3 Summary	7
References	8

Preface

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

1 + 1

[1] 2

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

2 Methods

2.1 Experimental design and data

The study involved culturing cells from both the wild-type (WT) and FLC1 Δ mutant strains under various environmental conditions. Cells were grown in four different media: YPD (Y), YPD with CFW (YC), YPD with EGTA (YE), and YPD with both CFW and EGTA (YCE), all at 37°C to induce stress. Additionally, both strains were grown under baseline conditions in standard YPD at 30°C. In total, 10 distinct experimental conditions were tested (5 environmental conditions \times 2 strains), each with 3 biological replicates, resulting in 30 samples.

Table 2.1 presents the experimental design along with the sample labels for each condition. The 10 experimental conditions are labeled by combining the levels of the three experimental factors: strain, temperature, and growth media, in that order. We use the label FL to denote the FLC1 Δ strain. For example, the condition involving the wild-type strain grown in YCE medium at 37°C is labeled WT-37-YCE, while the corresponding condition for the FLC1 Δ strain is labeled FL-37-YCE.

The primary output of the experiment was a raw count matrix containing RNA abundance measurements for 6795 genes across 30 samples, resulting in a 6795×30 matrix. Each entry in the matrix represents the number of sequencing reads mapped to a specific gene in a given sample.

Table 2.1: Overview of the experimental design showing the combinations of temperature and growth media (environmental conditions) used to grow cells of the WT and FLC1 Δ (FL) strains. Each cell lists the replicate sample labels (S1–S30) corresponding to each unique experimental condition.

Strain	Enviromental conditions				
	30°C	37°C			
	Y	Y	YC	YE	YCE
WT	S1, S2, S3	S4, S5, S6	S7, S8, S9	S10, S11, S12	S13, S14, S15
FL	S16, S17, S18	S19, S20, S21	S22, S23, S24	S25, S26, S27	S28, S29, S30

2.2 Pre-filtering

Some genes may exhibit low or zero counts across all samples, suggesting they are not expressed under the experimental conditions. These genes are unlikely to be identified as differentially expressed and are typically filtered out prior to analysis. The biological justification for pre-filtering low count or uninformative genes is that a gene typically needs to be expressed above a minimal threshold to be translated into a functional protein or to exert a meaningful biological effect (Chen, Lun, and Smyth 2016). Moreover, low-expression genes often reflect sampling noise rather than true biological signal (Sha, Phan, and Wang 2015).

DE analysis involves performing a hypothesis test for each gene to evaluate whether expression levels differ between experimental conditions (see Section 2.5). This results in a multiple testing problem, which is typically addressed by adjusting p-values to control the false discovery rate (FDR). However, such corrections reduce statistical power by raising the threshold for significance. As the number of tests increases, this loss of power aggravates, further limiting the ability to detect truly differentially expressed (DE) genes, especially when they represent a small proportion of the total set (Bourgon, Gentleman, and Huber 2010). To mitigate this, it is recommended to filter out low-expression genes prior to DE analysis, thereby reducing the total number of hypotheses. This makes the multiple testing correction less stringent and increases the chance of correctly identifying DE genes (Bourgon, Gentleman, and Huber 2010; Sha, Phan, and Wang 2015).

2.3 Data transformation

2.4 Exploratory analysis

2.5 Differential expression analysis

2.6 Motif enrichment analysis

3 Summary

In summary, this book has no content whatsoever.

1 + 1

[1] 2

References

```
r dim(counts_raw)[1]
```

- Bourgon, Richard, Robert Gentleman, and Wolfgang Huber. 2010. “Independent Filtering Increases Detection Power for High-Throughput Experiments.” *Proceedings of the National Academy of Sciences* 107 (21): 9546–51.
- Chen, Y., ATL. Lun, and GK. Smyth. 2016. “From Reads to Genes to Pathways: Differential Expression Analysis of RNA-Seq Experiments Using Rsubread and the edgeR Quasi-Likelihood Pipeline [Version 2; Peer Review: 5 Approved].” *F1000Research* 5 (1438). <https://doi.org/10.12688/f1000research.8987.2>.
- Knuth, Donald E. 1984. “Literate Programming.” *Comput. J.* 27 (2): 97–111. <https://doi.org/10.1093/comjnl/27.2.97>.
- Sha, Ying, John H Phan, and May D Wang. 2015. “Effect of Low-Expression Gene Filtering on Detection of Differentially Expressed Genes in RNA-Seq Data.” In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6461–64. IEEE.