BIOINFORMATICS INSTITUTE

Differential RNA expression analysis

May 28, 2024

Homework number 6

by Sapozhnikov N.A.

1 Introduction

Yeasts, especially *Saccharomyces cerevisiae*, are one of the most important eukaryotic model organisms used in biotechnology. They have been domesticated for over 6,000 years and play a crucial role in processes such as baking and brewing.

Yeasts are facultative anaerobes, meaning they can switch between aerobic respiration and anaerobic fermentation depending on the availability of oxygen. During aerobic respiration, yeasts generate ATP efficiently using glucose and oxygen. However, in the absence of oxygen, yeasts switch to fermentation, converting sugars into ethanol and carbon dioxide, a process critical in baking and alcoholic fermentation.

This study aims to investigate the changes in RNA expression in yeast cells before and during fermentation, providing insights into the metabolic shifts that occur during this process. Differential expression analysis helps identify genes that are significantly regulated under different conditions, revealing the underlying biological mechanisms and pathways involved.

2 Methods

2.1 Raw Data

The study uses RNA-seq data from yeast cells collected at two time points: before fermentation and 30 minutes into fermentation. There are two biological replicates for each condition:

- **Before fermentation:** SRR941816, SRR941817
- **During fermentation:** SRR941818, SRR941819

2.2 Reference Data

The reference genome and annotation files for *Saccharomyces cerevisiae* (strain S288c, assembly R64) were downloaded from NCBI:

- Reference genome: `GCF_000146045.2_R64_genomic.fna.gz`
- Annotation file: `GCF_000146045.2_R64_genomic.gff.gz`

2.3 Tools and Parameters

2.3.1 Alignment with HISAT2

```
build genome index:
hisat2-build data/GCF_000146045.2_R64_genomic.fna data/GCF_000146045.2_R64_genomic.index

align reads to the reference genome for each sample:
hisat2 -p 5 -x data/GCF_000146045.2_R64_genomic.index -U data/SRR941816.fastq.gz | samtools sort > da
hisat2 -p 5 -x data/GCF_000146045.2_R64_genomic.index -U data/SRR941817.fastq.gz | samtools sort > da
hisat2 -p 5 -x data/GCF_000146045.2_R64_genomic.index -U data/SRR941818.fastq.gz | samtools sort > da
hisat2 -p 5 -x data/GCF_000146045.2_R64_genomic.index -U data/SRR941819.fastq.gz | samtools sort > da
```

Run featureCounts to count reads:

```
featureCounts -t gene -g ID -a data/GCF_000146045.2_R64_genomic.gff -o data/gene_counts.txt data/SRR
```

Simplify the counts file:

```
cat data/gene_counts.txt | cut -f 1,7-10 > data/simple_counts.txt
```

2.3.2 Differential Expression Analysis with DESeq2

Run DESeq2 to find differentially expressed genes:

```
cat simple_counts.txt | R -f deseq2.r
```

This generates:

- `result.txt`: Differential expression metrics
- `norm-matrix-deseq2.txt`: Normalized counts for visualization

2.3.3 Visualization

Draw a heatmap:

```
cat data/simple_counts.txt | R -f src/deseq2.r
```

3 Results

3.1 Alignment with HISAT2

Successfully aligned reads to the reference genome, producing sorted BAM files for each sample.

3.2 Quantification with featureCounts

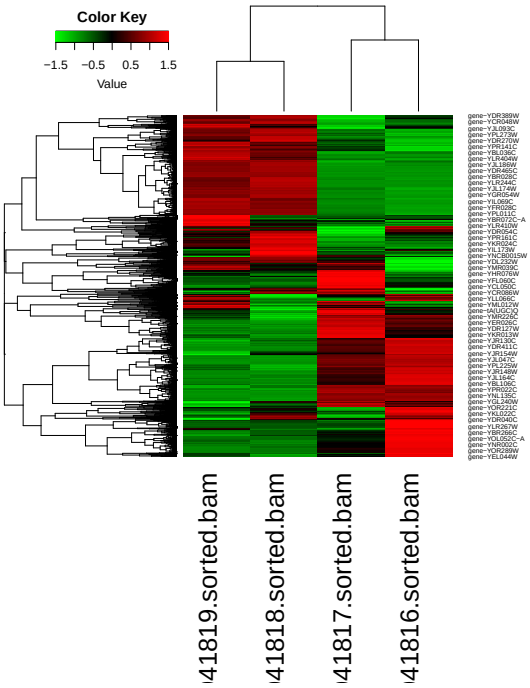
Generated a counts matrix (`counts.txt`) with read counts for each gene across all samples. Simplified counts matrix (`simple_counts.txt`) for input into DESeq2.

3.3 Differential Expression Analysis with DESeq2

Identified differentially expressed genes, with results stored in `result.txt`. Produced normalized count matrix (`norm-matrix-deseq2.txt`) for visualization.

3.4 Visualization

Generated a heatmap from normalized counts to visually assess changes in gene expression.



4 Discussion

4.1 Number of Genes and GO Terms Changed

From `result.txt`, the first 50 genes with the lowest adjusted p-values were extracted and analyzed using GO Slim Mapper. Significant changes were observed in the expression of multiple genes and their associated GO terms before and during fermentation.

4.2 Upregulated Process

Example Process: One of the upregulated processes was “carbohydrate metabolic process.”
Hypothesis: During fermentation, yeast cells likely increase the expression of genes involved in carbohydrate metabolism to efficiently convert glucose to ethanol and CO2, facilitating energy production under anaerobic conditions.

4.3 Downregulated Process

Example Process: Downregulated processes include “transcription by RNA polymerase II“ and “RNA splicing“
Hypothesis: In eukaryotes translation is spatially and temporarily separated. Thus beginning of active fermentation suppresses all transcription associated processes.

All GO enrichment results are available in the notebook supplementary.

5 Conclusion

This study successfully identified changes in RNA expression in yeast cells before and during fermentation. The differential expression analysis revealed significant upregulation of genes involved in carbohydrate metabolism and downregulation of genes associated with active transcription phase, highlighting the metabolic adaptations of yeast cells during fermentation. These findings provide valuable insights into the molecular mechanisms underlying yeast fermentation, with potential applications in biotechnology and fermentation industries.