

Analisis Differentially Expressed Genes pada *Saccharomyces cerevisiae* yang Dikultivasi dalam Lignocellulosic Hydrolysate Menggunakan GEO2R

Nur Aini

1. Pendahuluan

Produksi bioetanol berbasis lignoselulosa memerlukan proses hidrolisis biomassa yang efisien serta fermentasi gula oleh mikroorganisme. Namun, tahap pre-treatment lignoselulosa menghasilkan berbagai senyawa inhibitor seperti furfural, asam lemah (asam asetat dan format), dan senyawa fenolik yang dapat menghambat pertumbuhan dan aktivitas fermentasi *Saccharomyces cerevisiae*. Paparan senyawa-senyawa tersebut dapat menurunkan viabilitas sel, mengganggu metabolisme energi, memperpanjang fase lag, serta meningkatkan pembentukan *reactive oxygen species* (ROS) yang menyebabkan stres oksidatif. Oleh karena itu, pemahaman mengenai respons molekuler *S. cerevisiae* terhadap hidrolisat penting dipahami untuk meningkatkan efisiensi fermentasi bioetanol.

Berbagai pendekatan telah dilakukan untuk meningkatkan toleransi *S. cerevisiae* terhadap inhibitor lignoselulosa, termasuk rekayasa evolusioner, rekayasa genetik, serta adaptasi jangka pendek melalui paparan awal terhadap hidrolisat. Studi transkriptomik sebelumnya menunjukkan bahwa mekanisme stres oksidatif, enzim detoksifikasi (seperti reduktase furfural dan HMF), transporter membran, serta jalur biosintesis vitamin berperan penting dalam proses adaptasi tersebut (Almeida dkk., 2023). Hal ini menunjukkan bahwa perubahan ekspresi gen secara global merupakan kunci dalam memahami mekanisme toleransi dan adaptasi *S. cerevisiae* terhadap kondisi fermentasi berbasis lignoselulosa.

Berdasarkan latar belakang tersebut, pada tugas ini dilakukan eksplorasi dataset transkriptomik publik dari *Gene Expression Omnibus* (GEO) untuk menganalisis perbedaan ekspresi gen (DEG) antara kondisi kultur *S. cerevisiae* pada medium hidrolisat lignoselulosa dan medium berbasis gula sebagai kontrol. Analisis dilakukan menggunakan tool berbasis web GEO2R untuk mengidentifikasi gen-gen yang mengalami peningkatan ekspresi (*upregulated*) maupun penurunan ekspresi (*downregulated*). Hasil analisis ini diharapkan dapat memberikan gambaran awal mengenai respons molekuler dan jalur biologis yang terlibat dalam adaptasi *S. cerevisiae* terhadap stres hidrolisat lignoselulosa.

2. Metode

2.1 Dataset

Dataset yang digunakan dalam analisis ini adalah GSE218764 yang diperoleh dari database publik *Gene Expression Omnibus* (GEO), NCBI. Dataset ini berjudul “*Physiological and molecular characterization of yeast cultures pre-adapted for fermentation of lignocellulosic hydrolysate*” dan menggunakan platform GPL (microarray Affymetrix) untuk menganalisis ekspresi gen pada *Saccharomyces cerevisiae* strain TMB3500. Sampel yang dipilih untuk analisis adalah kultur *S. cerevisiae* yang ditumbuhkan dalam a) Medium hidrolisat lignoselulosa (9 jam) dan, b) Medium berbasis gula (9 jam) sebagai kontrol pembanding. Masing-masing kondisi terdiri atas dua replikasi biologis, sehingga total empat sampel digunakan dalam perbandingan diferensial.

2.2 Analisis Differentially Expressed Genes (DEG)

Analisis ekspresi gen diferensial dilakukan menggunakan GEO2R, yaitu tool berbasis web yang tersedia pada halaman dataset GEO. Tahapan analisis meliputi:

2.2.1 Pembagian Kelompok (Group Assignment):

Pembagian kelompok disesuaikan dengan

- a. *Group 1*: Hidrolisat (9h)
- b. *Group 2*: Media gula (9h)

Series GSE218764		Query DataSets for GSE218764
Status	Public on Feb 21, 2023	
Title	Physiological and molecular characterization of yeast cultures pre-adapted for fermentation of lignocellulosic hydrolysate	
Platform organisms	<i>Schizosaccharomyces pombe</i> ; <i>Saccharomyces cerevisiae</i>	
Sample organism	<i>Saccharomyces cerevisiae</i>	
Experiment type	Expression profiling by array	
Summary	To reach an economically feasible bioethanol process from lignocellulose, efficient fermentation by yeast of all sugars present in the hydrolysate has to be achieved. However, when exposed to lignocellulosic hydrolysate, <i>Saccharomyces cerevisiae</i> is challenged with a variety of inhibitors that reduce yeast viability, growth and fermentation rate, and in addition damage cellular structures. In order to evaluate the yeast capability to adapt to lignocellulosic hydrolysates and to investigate the yeast molecular response to inhibitors, fed-batch cultivation of an industrial <i>S. cerevisiae</i> strain was performed using either spruce hydrolysate or a sugar medium as feed. The physiological effects of cultivating yeast in spruce hydrolysate was comprehensively studied by assessment of yeast performance in simultaneous saccharification and fermentation (SSF), measurement of furaldehyde reduction activity, assessment of conversion of phenolic compounds and genome wide transcription analysis. The yeast cultivated in spruce hydrolysate developed a rapid adaptive response to lignocellulosic hydrolysate, which significantly improved its fermentation performance in subsequent SSF experiments. Yeast adaptation to hydrolysate was shown to involve induction of NADPH-dependent aldehyde reduction activity and conversion of phenolic compounds during the fed-batch cultivation and these properties were correlated to the expression of several genes encoding oxido-reductase activities, notably AAD4, ADH6, OYE2/3 and YML131w. The other most significant transcriptional changes involved genes involved in transport mechanisms, such as YHK8, FLR1 or ATR1. A large set of genes were found to be associated to transcription factors involved in stress response (Msn2p, Msn4p, Yap1p) but also cell growth and division (Gcr4p, Ste12p, Sok2p) that were most likely activated at the post-transcriptional level.	
Overall design	Two biological replicates for each of the following conditions: <i>S. cerevisiae</i> start of fed-batch (0h) <i>S. cerevisiae</i> hydrolysate fed-batch (9h) <i>S. cerevisiae</i> hydrolysate fed-batch (18h) <i>S. cerevisiae</i> sugars fed-batch (9h) <i>S. cerevisiae</i> sugars fed-batch (18h)	
Contributor(s)	Almeida JR, Wiman M, Heer D, Brink DP, Sauer U, Hahn-Hägerdal B, Lidén G, Gorwa-Grauslund MF	
Citation	Almeida JRM, Wiman M, Heer D, Brink DP, Sauer U, Hahn-Hägerdal B, Lidén G, and Gorwa-Grauslund MF. Physiological and Molecular Characterization of Yeast Cultures Pre-Adapted for Fermentation of Lignocellulosic Hydrolysate. <i>Fermentation</i> . 2023; 9(1):72. doi:10.3390/fermentation9010072	

Analyze with GEO2R

Submission date	Nov 25, 2022
Last update date	Feb 22, 2023
Contact name	Daniel P Brink
E-mail(s)	daniel.brink@lmb.lth.se
Organization name	Lund University
Department	Applied Microbiology
Street address	Naturvetarvagen 14
City	LUND
ZIP/Postal code	SE-223 62
Country	Sweden

Gambar 1. DataSet GSE218764 dari database *Gene Expression Omnibus* (GEO) NCBI

2.2.2 Metode Statistik

GEO2R menggunakan paket statistik berbasis limma (*Linear Models for Microarray Data*) untuk menghitung perbedaan ekspresi gen antar kelompok. Limma adalah paket perangkat lunak berbasis R/Bioconductor yang digunakan untuk menganalisis data ekspresi gen dalam studi transkriptomik. Limma memungkinkan analisis diferensial ekspresi dan diferensial splicing pada data mikroarray maupun RNA-seq dengan pendekatan statistik yang kuat. Selain itu, limma mendukung normalisasi dan analisis lanjutan untuk membantu interpretasi biologis hasil ekspresi gen (Ritchie dkk., 2015).

2.2.3 Koreksi Multiple Testing

Metode koreksi Benjamini–Hochberg adjusted p-value digunakan untuk mengontrol *false discovery rate* (FDR).

2.2.4 Kriteria Signifikansi

Gen dinyatakan sebagai DEG apabila memenuhi kriteria:

- Adjusted p-value < 0,05
- $|\log_2 \text{Fold Change} (\log FC)| \geq 1$

Gen dengan logFC positif dikategorikan sebagai *upregulated*, sedangkan gen dengan logFC negatif dikategorikan sebagai *downregulated*. Data jumlah gen *upregulated* dan *downregulated* ditentukan dengan analisis menggunakan Microsoft Excel.

2.3 Replikasi Analisis (Reproducibility)

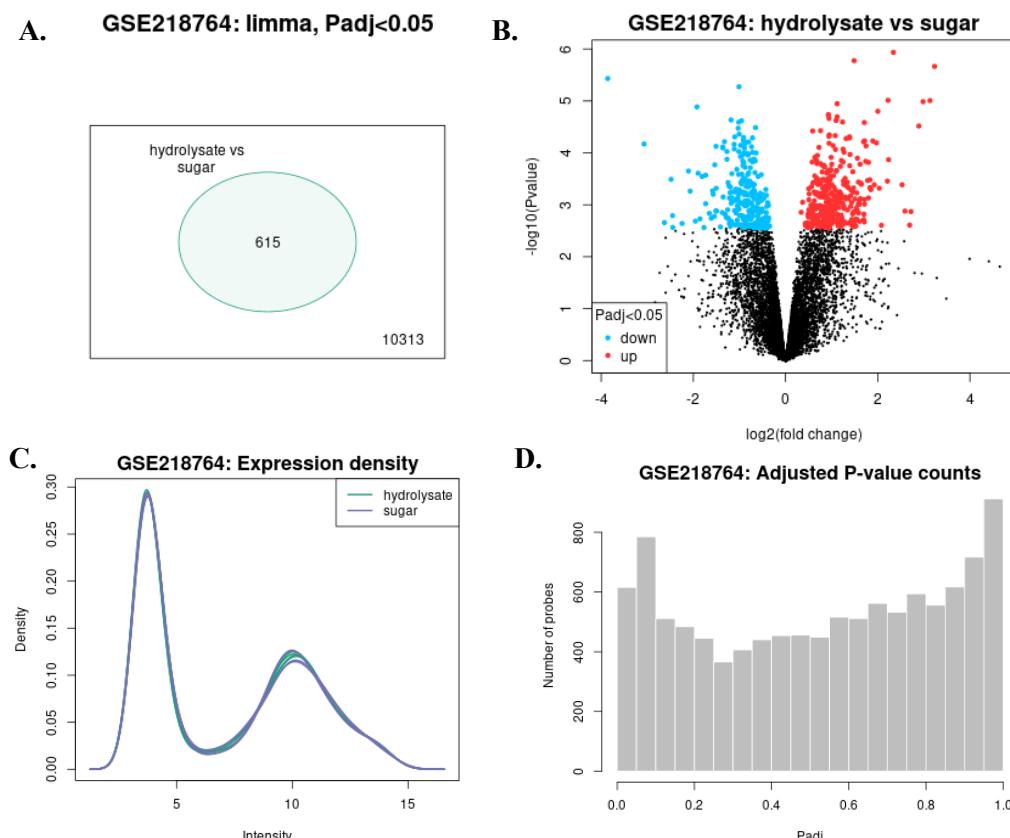
Untuk memastikan konsistensi hasil, analisis dilakukan sebanyak tiga kali replikasi dengan parameter yang sama. Replikasi dilakukan dengan:

- Pengulangan pembagian kelompok yang identik
- Penggunaan parameter statistik dan kriteria signifikansi yang konsisten
- Verifikasi kesamaan pola gen signifikan

Hasil antar replikasi dibandingkan untuk memastikan konsistensi jumlah gen signifikan serta arah perubahan ekspresi (*upregulation* atau *downregulation*).

3. Hasil dan Interpretasi

Analisis diferensial ekspresi gen pada dataset GSE218764 dilakukan menggunakan metode limma melalui platform GEO2R dengan kriteria signifikansi adjusted p-value ($\text{Padj} < 0,05$). Dari total 10.928 gen yang dianalisis, sebanyak 615 gen teridentifikasi sebagai *differentially expressed genes* (DEGs) antara kondisi kultivasi menggunakan hidrolisat lignoselulosa dan medium gula (Gambar 2A). Dari jumlah tersebut, 344 gen mengalami peningkatan ekspresi (*upregulated*), sedangkan 271 gen mengalami penurunan ekspresi (*downregulated*) pada kondisi hidrolisat dibandingkan dengan kontrol gula.



Gambar 2. Analisis ekspresi gen diferensial pada *S. cerevisiae* yang dikultur dalam medium lignoselulosa hidrolisat dibandingkan dengan medium gula. (A) Ringkasan jumlah gen yang teridentifikasi sebagai berbeda signifikan berdasarkan analisis limma (adjusted p-value $< 0,05$). (B) Volcano plot yang menunjukkan distribusi gen *upregulated* dan *downregulated*. (C) Plot densitas distribusi intensitas ekspresi ternormalisasi pada masing-

masing kondisi, yang menunjukkan pola distribusi yang relatif serupa. (D) Histogram distribusi adjusted p-value dari seluruh gen yang dianalisis.

Visualisasi volcano plot menunjukkan pemisahan yang jelas antara gen yang meningkat dan menurun ekspresinya berdasarkan nilai log₂ (fold change) dan -log₁₀ (p-value) (Gambar 2B). Sebagian besar gen signifikan terkonsentrasi pada rentang log₂FC sekitar ±1 hingga ±3, dengan beberapa gen menunjukkan perubahan ekspresi yang lebih ekstrem (log₂FC > 3). Jumlah gen yang mengalami peningkatan ekspresi tampak lebih dominan dibandingkan dengan gen yang mengalami penurunan ekspresi, yang mengindikasikan adanya respons transkriptomik aktif terhadap paparan hidrolisat. Kurva *expression density* memperlihatkan pola distribusi intensitas ekspresi yang relatif serupa antar kelompok sampel (Gambar 2C-D). Hal ini menunjukkan bahwa proses normalisasi data telah dilakukan dengan baik sehingga variasi yang teramat lebih merefleksikan perbedaan biologis dibandingkan variasi teknis.

Dominasi gen yang mengalami peningkatan ekspresi (*upregulated*) menunjukkan bahwa paparan hidrolisat lignoselulosa memicu respons adaptif yang luas pada *S. cerevisiae*. Kondisi hidrolisat yang mengandung senyawa inhibitor seperti furfural, asam lemah, dan senyawa fenolik kemungkinan menginduksi aktivasi jalur respons stres, mekanisme detoksifikasi, sistem pertahanan terhadap stres oksidatif, serta regulasi transporter dan metabolisme energi untuk mempertahankan homeostasis seluler. Sebaliknya, gen yang mengalami penurunan ekspresi (*downregulated*) merepresentasikan jalur metabolismik yang ditekan sementara sebagai strategi efisiensi energi dalam kondisi stres. Secara keseluruhan, hasil ini menunjukkan bahwa adaptasi terhadap spruce hydrolysate melibatkan reprogramming transkriptomik yang kompleks dengan kecenderungan aktivasi gen lebih dominan dibandingkan represi gen (Almeida dkk., 2023).

4. Kesimpulan

Analisis dataset GSE218764 menggunakan GEO2R mengidentifikasi 615 gen yang berbeda signifikan (adjusted p-value < 0,05 dan |log₂FC| ≥ 1) dari total 10.928 gen yang dianalisis, terdiri atas 344 gen *upregulated* dan 271 gen *downregulated* pada kondisi hidrolisat dibandingkan medium gula. Visualisasi hasil menunjukkan distribusi gen signifikan yang jelas serta kualitas normalisasi data yang baik. Dominasi gen yang mengalami peningkatan ekspresi mengindikasikan bahwa *S. cerevisiae* merespons paparan hidrolisat lignoselulosa melalui aktivasi berbagai mekanisme adaptif. Hasil ini menunjukkan bahwa adaptasi terhadap hidrolisat melibatkan perubahan ekspresi gen yang luas dan spesifik terhadap kondisi lingkungan serta menunjukkan bahwa analisis berbasis GEO2R efektif untuk eksplorasi awal respons transkriptomik.

Daftar Pustaka

- Almeida, J. R. M., Wiman, M., Heer, D., Brink, D. P., Sauer, U., Hahn-Hägerdal, B., Lidén, G., & Gorwa-Grauslund, M. F. (2023). Physiological and Molecular Characterization of Yeast Cultures Pre-Adapted for Fermentation of Lignocellulosic Hydrolysate. *Fermentation*, 9(1), 72. <https://doi.org/10.3390/fermentation9010072>
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47–e47. <https://doi.org/10.1093/nar/gkv007>