

Analisis Differentially Expressed Genes pada *Saccharomyces cerevisiae* yang Dikultur dalam Hidrolisat Lignoselulosa Menggunakan R

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 laporan ini dilakukan analisis diferensial ekspresi gen (*Differentially Expressed Genes*, DEG) pada *S. cerevisiae* yang dikultivasi dalam medium hidrolisat lignoselulosa dibandingkan dengan medium gula sebagai kontrol menggunakan perangkat lunak R. Dataset transkriptomik publik yang diperoleh dari Gene Expression Omnibus (GEO) dianalisis untuk mengidentifikasi gen-gen yang mengalami peningkatan ekspresi (*upregulated*) maupun penurunan ekspresi (*downregulated*) secara signifikan. Selanjutnya, gen-gen signifikan tersebut dianalisis lebih lanjut melalui pendekatan *enrichment* fungsional menggunakan Gene Ontology (GO) dan KEGG pathway untuk menginterpretasikan keterlibatan proses biologis, fungsi molekuler, serta jalur metabolismik yang terdampak akibat paparan hidrolisat lignoselulosa. Melalui analisis terintegrasi ini, diharapkan diperoleh gambaran yang lebih komprehensif mengenai mekanisme adaptasi transkriptomik *S. cerevisiae* terhadap tekanan metabolik dan senyawa inhibitor yang terkandung dalam hidrolisat lignoselulosa.

2. Metode

2.1 Dataset dan Desain Eksperimen

Dataset yang digunakan dalam analisis ini adalah GSE218764 yang diperoleh dari database publik GEO, NCBI. Dataset ini berjudul “Physiological and molecular characterization of yeast cultures pre-adapted for fermentation of lignocellulosic hydrolysate” dan menggunakan platform GPL2529 (microarray Affymetrix) untuk menganalisis ekspresi gen pada *Saccharomyces cerevisiae* strain TMB3500. Analisis difokuskan pada titik waktu inkubasi yang sama untuk menghindari bias akibat variasi fase pertumbuhan, sehingga

perbandingan hanya merefleksikan perbedaan media kultur. Sampel dikelompokkan menjadi dua kondisi biologis, yaitu hydrolysate dan sugars.

2.2 Pemrosesan Data dan Analisis Ekspresi Gen (DEG)

Seluruh analisis dilakukan menggunakan perangkat lunak R. Data ekspresi gen dalam format matriks diekstraksi dan difilter sesuai dengan desain eksperimental. Analisis diferensial ekspresi gen dilakukan menggunakan paket limma (*Linear Models for Microarray Data*), yang mengimplementasikan pendekatan model linear dan *empirical Bayes moderation* untuk meningkatkan estimasi variansi pada jumlah replikasi terbatas. Design matrix dibuat berdasarkan faktor kondisi media (hydrolysate vs sugars), kemudian contrast matrix ditetapkan untuk membandingkan ekspresi gen antar kondisi. Gen yang menunjukkan nilai adjusted p-value (Benjamini–Hochberg false discovery rate) $< 0,05$ dan $|\log_2 \text{fold change}| > 1$ dikategorikan sebagai *differentially expressed genes* (DEGs).

2.3 Anotasi Gen dan Konversi ID

Anotasi gen dilakukan menggunakan paket **org.Sc.sgd.db**, yang menyediakan pemetaan antara berbagai tipe identifikasi gen *S. cerevisiae*. Karena analisis jalur biologis memerlukan format identifikasi tertentu, ID gen signifikan dikonversi dari format Saccharomyces Genome Database (SGD) ke format Open Reading Frame (ORF) untuk analisis KEGG, serta ke ENTREZ Gene ID untuk analisis Gene Ontology (GO), menggunakan fungsi `bitr()` dari paket **clusterProfiler**.

2.4 Analisis Enrichment

Analisis *enrichment* fungsional dilakukan menggunakan paket **clusterProfiler**. Gene Ontology *enrichment* dianalisis berdasarkan kategori Biological Process (BP) dengan metode koreksi multiple testing Benjamini–Hochberg dan batas signifikansi adjusted p-value $<0,05$. Analisis KEGG pathway dilakukan dengan parameter organisme “sce” untuk *S. cerevisiae*, menggunakan daftar gen dalam format ORF sebagai input. Hasil enrichment divisualisasikan menggunakan dot plot untuk menunjukkan proses biologis dan jalur metabolismik yang secara signifikan lebih aktif pada kondisi hidrolisat lignoselulosa dibandingkan kondisi kontrol.

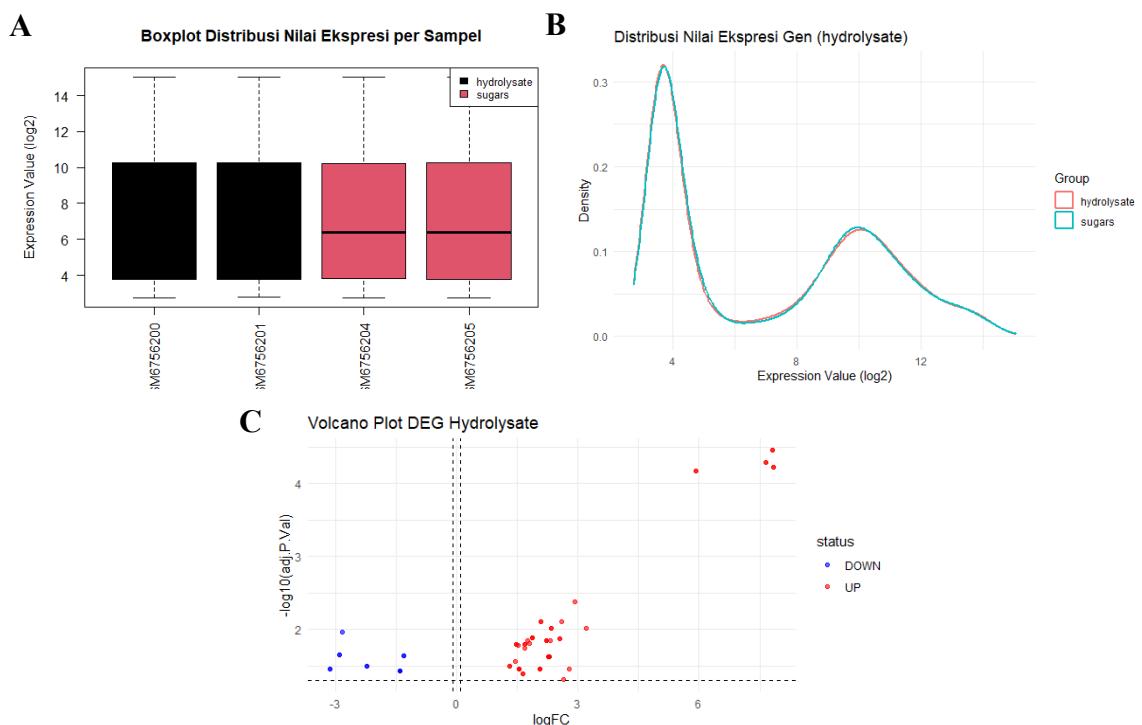
2.5 Visualisasi Data

Visualisasi data dilakukan untuk mengevaluasi kualitas data serta menginterpretasikan hasil analisis secara komprehensif. Boxplot digunakan untuk menilai distribusi dan keseragaman intensitas ekspresi antar sampel sebagai indikator keberhasilan normalisasi data. Distribusi nilai ekspresi gen divisualisasikan dalam bentuk kurva density untuk memastikan pola sebaran yang relatif serupa antar kelompok perlakuan. Volcano plot digunakan untuk menggambarkan hubungan antara nilai $\log_2 \text{fold change}$ dan $-\log_{10} \text{adjusted p-value}$ sehingga gen yang mengalami peningkatan (upregulated) maupun penurunan ekspresi (downregulated) dapat teridentifikasi secara jelas. Heatmap digunakan untuk menunjukkan pola klasterisasi ekspresi antar sampel dan memvalidasi pemisahan biologis antara kondisi hydrolysate dan sugars. Selain itu, hasil analisis KEGG pathway divisualisasikan menggunakan KEGG Mapper (<https://www.genome.jp/kegg/mapper/>) untuk memetakan gen-gen signifikan ke dalam jalur metabolismik spesifik, sehingga memudahkan interpretasi keterlibatan gen dalam proses biologis yang terdampak oleh kondisi hidrolisat lignoselulosa.

3. Hasil dan Interpretasi

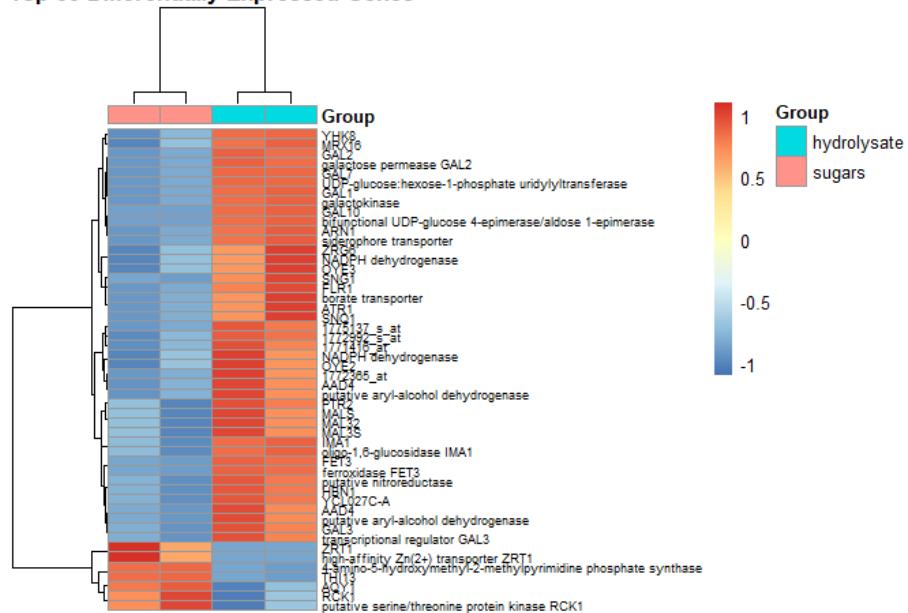
Analisis diferensial ekspresi gen pada dataset GSE218764 dilakukan menggunakan metode *limma* melalui pendekatan model linear dan *empirical Bayes moderation* untuk mengidentifikasi gen yang mengalami perubahan ekspresi signifikan antara kondisi hidrolisat lignoselulosa dan gula sebagai kontrol. Penelitian ini bertujuan untuk menganalisis transkriptom pada sel yang teradaptasi kondisi hidrolisat dan kontrol atau medium gula masing-masing pada waktu 9 jam dan 17 jam. Namun, analisis DEG pada laporan ini difokuskan pada perlakuan waktu 9 jam, didasarkan pada waktu yang dibutuhkan sel untuk reduksi furaldehida sebagai indikator respons seluler terhadap inhibitor dalam hidrolisat lignoselulosa.

Hasil visualisasi awal menunjukkan bahwa distribusi nilai ekspresi gen antar sampel relatif seragam. Boxplot memperlihatkan median dan rentang distribusi ekspresi (\log_2) yang sebanding pada seluruh sampel, mengindikasikan bahwa proses normalisasi data telah dilakukan dengan baik tanpa adanya penyimpangan ekstrem (Gambar 1A). Distribusi kepadatan (*density plot*) juga menunjukkan pola yang hampir tumpang tindih antara kedua kelompok, menandakan bahwa perubahan ekspresi tidak terjadi secara menyeluruh pada seluruh gen, melainkan terbatas pada gen tertentu (Gambar 1B). Hasil ini diperkuat oleh volcano plot yang menunjukkan adanya gen-gen dengan nilai \log_2 fold change yang tinggi dan adjusted p-value signifikan, baik gen-gen yang upregulated maupun downregulated (Gambar 1C). Secara umum, jumlah gen yang mengalami peningkatan ekspresi pada kondisi hidrolisat tampak lebih dominan dibandingkan gen yang mengalami penurunan ekspresi, mengindikasikan adanya aktivasi respons transkriptomik spesifik terhadap kondisi hidrolisat lignoselulosa.



Gambar 1. Analisis ekspresi gen diferensial pada *S. cerevisiae* yang dikultur dalam medium lignoselulosa hidrolisat dibandingkan dengan medium gula. (A) Boxplot distribusi nilai ekspresi gen ter-normalisasi (\log_2). (B) Plot densitas distribusi ekspresi gen. (C) Volcano plot yang menunjukkan distribusi gen *upregulated* dan *downregulated*.

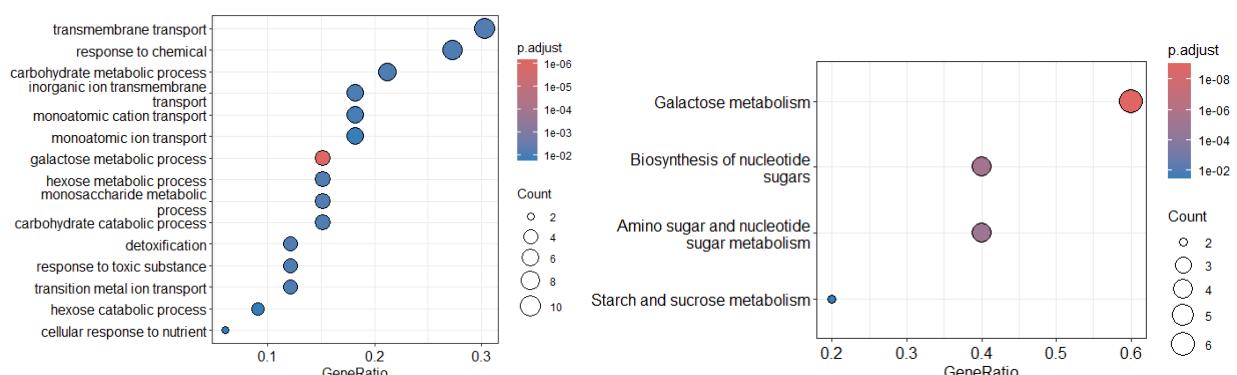
Top 50 Differentially Expressed Genes



Gambar 2. Heatmap 50 gen dengan ekspresi diferensial tertinggi antara kondisi hidrolisat dan kontrol gula.

Visualisasi heatmap terhadap 50 gen dengan nilai diferensial ekspresi tertinggi menunjukkan pengelompokan yang jelas antara kelompok hidrolisat dan gula (Gambar 2). Sampel dari masing-masing kelompok mengelompok secara konsisten dalam cabang dendrogram yang berbeda, menandakan adanya perbedaan profil transkriptomik antar kondisi perlakuan. Sebagian besar gen memperlihatkan pola ekspresi yang meningkat (merah) pada kondisi hidrolisat dan menurun (biru) pada kondisi kontrol, atau sebaliknya, yang mengindikasikan regulasi diferensial yang kuat akibat paparan hidrolisat lignoselulosa. Beberapa gen yang terlibat dalam metabolisme galaktosa seperti *GAL2*, *GAL7*, dan *GAL10*, menunjukkan pola ekspresi tinggi pada salah satu kondisi, menunjukkan adanya penyesuaian jalur metabolisme karbon. Selain itu, gen-gen yang berperan dalam transport, respons stres oksidatif, dan metabolisme energi juga tampak mengalami perubahan signifikan.

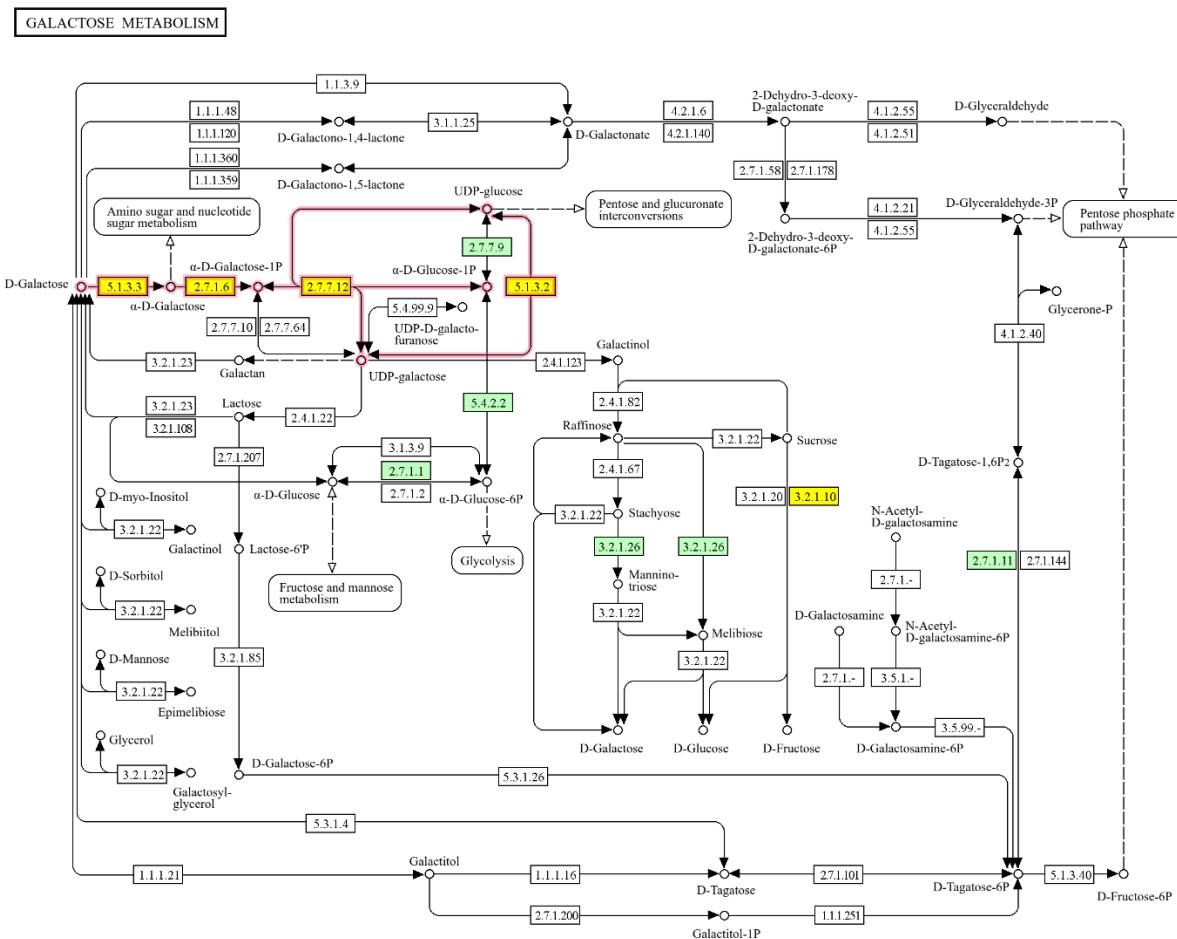
Hasil analisis enrichment Gene Ontology (GO) menunjukkan bahwa gen-gen yang terdiferensiasi secara signifikan diperkaya pada proses yang berkaitan dengan metabolisme karbohidrat, transport membran, serta respons terhadap senyawa kimia dan stres (Gambar 3).

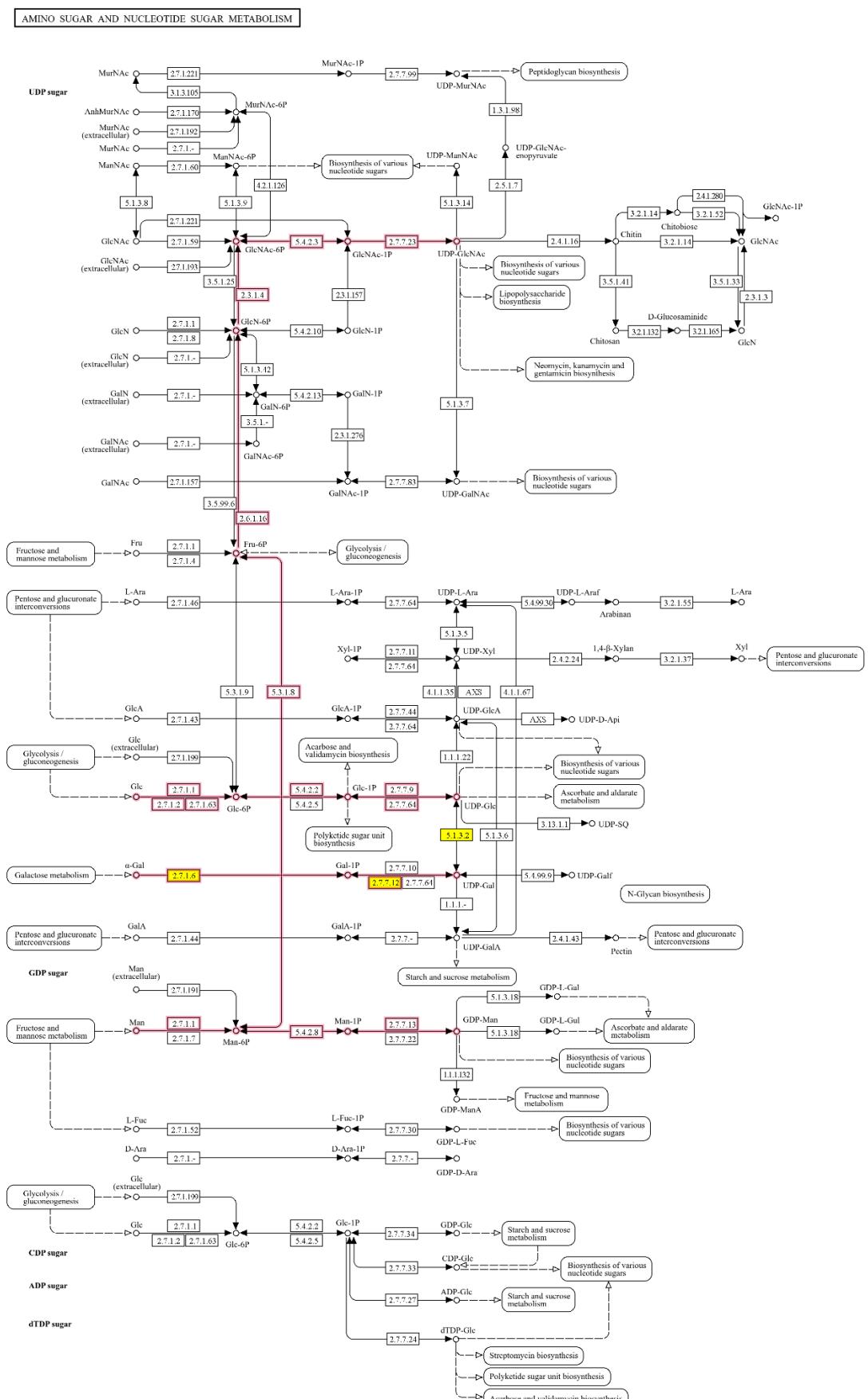


Gambar 3. Analisis enrichment Gene Ontology (kiri) dan KEGG pathway (kanan) pada gen diferensial antara kondisi hydrolysate dan kontrol gula.

Istilah dengan GeneRatio tinggi seperti *transmembrane transport*, *response to chemical*, dan *carbohydrate metabolic process* mengindikasikan bahwa kondisi lignocellulosic hydrolysate memicu aktivasi mekanisme transport molekul melintasi membran serta penyesuaian metabolisme karbon. Selain itu, munculnya istilah seperti *detoxification*, *response to toxic substance*, dan *transition metal ion transport* menunjukkan adanya respons sel terhadap senyawa inhibitor atau tekanan kimia yang umum terdapat dalam hidrolisat lignoselulosa. Secara konsisten, analisis KEGG pathway memperlihatkan pengayaan signifikan pada jalur *Galactose metabolism* dengan nilai GeneRatio dan signifikansi tertinggi, diikuti oleh *Biosynthesis of nucleotide sugars*, *Amino sugar and nucleotide sugar metabolism*, serta *Starch and sucrose metabolism*.

Hasil pemetaan menggunakan KEGG Mapper menunjukkan bahwa gen-gen diferensial secara jelas terdistribusi pada jalur metabolisme galaktosa dan metabolisme gula amino dan gula nukleotida (Gambar 4). Pada jalur metabolisme galaktosa, beberapa enzim kunci seperti galaktokinase (GAL1), galaktosa-1-fosfat uridiltransferase (GAL7), dan UDP-glukosa 4-epimerase (GAL10) terpetakan dan teraktivasi pada tahapan konversi D-galaktosa menjadi UDP-glukosa. Aktivasi rangkaian enzim ini mengindikasikan adanya peningkatan fluks metabolik dari galaktosa menuju intermediat glikolisis yang menunjukkan penyesuaian





Gambar 4. Pemetaan DEG pada jalur metabolisme galaktosa dan metabolisme gula amino dan gula nukleotida menggunakan KEGG Mapper.

metabolisme karbon dalam kondisi hidrolisat lignoselulosa. Selain itu, keterlibatan enzim pada cabang pembentukan gula nukleotida menunjukkan bahwa sel tidak hanya memanfaatkan galaktosa sebagai sumber energi, tetapi juga untuk sintesis komponen struktural dan prekursor biosintetik.

Pada jalur metabolisme gula amino dan gula nukleotida, gen-gen yang terpetakan berperan dalam konversi glukosa-6-fosfat menjadi berbagai gula nukleotida seperti UDP-glukosa dan UDP-galaktosa, yang merupakan prekursor penting dalam biosintesis dinding sel dan glikoprotein. Distribusi DEG pada jalur metabolismik utama, termasuk berhubungan dengan jalur glikolisis dan metabolisme fruktosa/mannosa, menunjukkan adanya perubahan metabolisme pusat (*central carbon metabolism*). Secara keseluruhan, hasil KEGG Mapper mengonfirmasi bahwa respons transkriptomik *S. cerevisiae* terhadap kondisi hidrolisat difokuskan pada optimalisasi pemanfaatan sumber karbon alternatif serta penyesuaian biosintesis komponen seluler, yang berfungsi untuk meningkatkan toleransi terhadap tekanan lingkungan hidrolisat lignoselulosa.

4. Kesimpulan

Analisis diferensial ekspresi gen pada waktu 9 jam menunjukkan bahwa *S. cerevisiae* mengalami perubahan transkriptomik yang spesifik ketika dikultivasi dalam hidrolisat lignoselulosa dibandingkan medium gula kontrol. Perubahan tersebut melibatkan gen-gen yang berperan dalam metabolisme karbohidrat, transport membran, serta respons terhadap stres kimia dan senyawa inhibitor. Hasil *enrichment* GO, KEGG pathway, dan pemetaan KEGG Mapper menunjukkan bahwa adaptasi sel difokuskan pada reorganisasi metabolisme karbon, khususnya jalur metabolisme galaktosa dan biosintesis gula nukleotida, serta aktivasi mekanisme protektif terhadap stres lingkungan. Hasil ini menunjukkan bahwa toleransi *S. cerevisiae* terhadap hidrolisat lignoselulosa terjadi melalui optimalisasi pemanfaatan sumber karbon alternatif dan penyesuaian metabolisme pusat untuk mempertahankan kelangsungan pertumbuhan sel.