

Pemanfaatan edgeR dan Pemodelan Jaringan STRING v12.0 pada Analisis Data RNA Kanker Nasofaring

Dicka Wahyu Setiasari^{1*}, Fandisya Rahman²

¹ Bioteknologi, Fakultas Sains dan Teknologi, UIN Raden Mas Said Surakarta, Surakarta, Indonesia

² Informatika (Kampus Kota Surabaya), Universitas Telkom, Indonesia

E-mail: dicka.wahyu@staff.uinsaid.ac.id^{1*}, fandisya@telkomuniversity.ac.id²

INFO ARTIKEL

Sejarah Artikel

Diterima : 28/08/2026

Direvisi : 14/09/2026

Diterbitkan : 01/12/2026

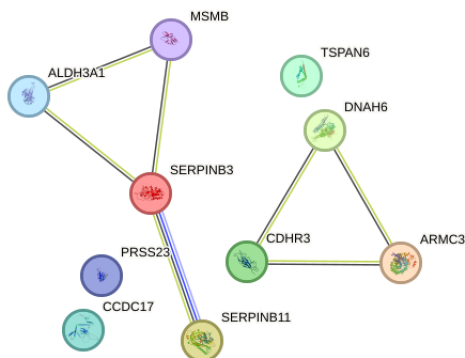
*Corresponding author

dicka.wahyu@staff.uinsaid.ac.id

DOI: 10.70247/jumistik.v5i2.371

<https://ojs.amiklps.ac.id>

GRAPHICAL ABSTRACT



ABSTRACT

Nasopharyngeal carcinoma (NPC) is an epithelial malignancy with high prevalence in Indonesia. This study aims to perform a comprehensive bioinformatic transcriptomic analysis and reconstruct the protein-protein interaction (PPI) network in Indonesian NPC patients. Whole-transcriptome RNA sequencing (RNA-seq) data from 7 NPC biopsy samples and 2 healthy controls were processed via a computational pipeline using HISAT2 v2.1.0 for genome alignment and HTSeq v0.11.1 for transcript quantification. Differential gene expression (DEG) analysis executed the edgeR v3.28.0 statistical algorithm (TMM normalization, FDR < 0.05, log2 FC > 1), identifying 2,043 significantly upregulated genes and 156 downregulated genes. Computational filtering revealed that 5 of the top 10 upregulated genes were long non-coding RNAs (lncRNAs). Network graph reconstruction via STRING v12.0 identified two major upregulated sub-clusters (SOX11-SALL3 and GABRG2-KCNQ2) and a downregulated central hub node SERPINB3 connected to ALDH3A1, MSMB, and SERPINB11. These findings demonstrate the efficacy of the bioinformatic pipeline in extracting candidate molecular biomarkers for NPC.

Keywords: Bioinformatics; Biomedical Informatics; EdgeR Algorithm; Nasopharyngeal Carcinoma; RNA-seq Pipeline; STRING v12.0

ABSTRAK

Karsinoma nasofaring (NPC) merupakan keganasan epitelial yang memiliki prevalensi tinggi di Indonesia. Penelitian ini bertujuan untuk melakukan analisis bioinformatika transkriptomik secara menyeluruh dan merekonstruksi jaringan interaksi protein-protein (PPI) pada pasien NPC Indonesia. Data sekuens RNA utuh (RNA-seq) dari 7 sampel biopsi pasien NPC dan 2 kontrol sehat diproses menggunakan pipeline komputasional HISAT2 v2.1.0 untuk alignment genome dan HTSeq v0.11.1 untuk kuantifikasi transkrip. Analisis ekspresi gen diferensial (DEG) mengeksekusi algoritma statistik edgeR v3.28.0 (TMM normalization, FDR < 0,05, log2 FC > 1), menghasilkan 2.043 gen ter-upregulasi dan 156 gen ter-downregulasi signifikan. Eliminasi komputasional mendeteksi bahwa 5 dari top 10 gen upregulated merupakan long non-coding RNA (lncRNA). Rekonstruksi grafik jaringan PPI menggunakan algoritma STRING v12.0 mengidentifikasi dua sub-kluster upregulated utama (SOX11-SALL3 dan GABRG2-KCNQ2) serta central hub node downregulated SERPINB3 yang terhubung dengan ALDH3A1, MSMB, dan SERPINB11. Hasil ini menegaskan efektivitas pipeline bioinformatika dalam mengekstraksi biomarker molekuler NPC.

Kata kunci: Bioinformatika; Pipeline RNA-seq; Algoritma edgeR; STRING v12.0; Karsinoma Nasofaring; Informatika Biomedis;

© 2020@ Penerbit STMIK Amika Soppeng. All rights reserved

PENDAHULUAN

Karsinoma nasofaring (NPC) merupakan keganasan epitelial yang memiliki distribusi geografis dan etnis yang sangat khas. Berdasarkan estimasi statistik kanker global, NPC mencatat lebih dari 133.000 kasus baru dan 80.000 kematian per tahun secara global [1],[2]. Meskipun relatif jarang di negara-negara Barat, NPC bersifat endemik di Asia Tenggara, Tiongkok Selatan, dan Afrika Utara, dengan insidensi regional mencapai 10 hingga 20 kasus per 100.000 penduduk per tahun [3]. Tingginya beban klinis ini dipicu oleh interaksi kompleks antara kerentanan genetik populasi, paparan karsinogen lingkungan, dan infeksi Epstein-Barr Virus (EBV) [4].

Indonesia diakui sebagai salah satu negara dengan beban kasus NPC tertinggi di Asia Tenggara, dengan insidensi berkisar antara 4 hingga 6 kasus per 100.000 penduduk per tahun atau melebihi 10.000 kasus baru setiap tahunnya. Namun demikian, studi molekuler berbasis data lokal masih sangat terbatas. NPC menunjukkan heterogenitas genetik yang signifikan antar-populasi; misalnya, varian alel HLA kelas I dan mutasi somatic CYLD sangat dominan pada kohort Tiongkok Selatan, sementara arsitektur genomik dan interaksi virus-inang pada populasi ras Austronesia/Melayu seperti di Indonesia belum terpetakan secara utuh.

Sebagian besar pemahaman molekuler NPC saat ini berasal dari penelitian di Tiongkok Selatan dan Hong Kong [5]. Namun, perbedaan latar belakang genetik ancestral (Melayu vs Han) serta variasi strain lokal EBV berpotensi menghasilkan fenotipe tumor dan profil ekspresi gen yang berbeda. Kegagalan memahami lanskap spesifik populasi ini dapat menyebabkan tidak efektifnya biomarker diagnosis maupun target terapi yang ditransfer dari populasi luar. Oleh karena itu, pengolahan data transkriptomik primer populasi lokal menjadi krusial [6].

Kemajuan teknologi Next-Generation Sequencing berbasis profil transkriptom utuh (RNA-seq) memberikan pendekatan yang sangat sensitif dan berkapasitas dinamis tinggi untuk mengukur perubahan ekspresi puluhan ribu gen secara bersamaan tanpa tergantung pada probe awal [7]. Dibandingkan dengan teknologi microarray konvensional, RNA-seq mampu mendeteksi

sinyal jaringan yang subtle namun terkoordinasi secara obyektif dari jaringan biopsi pasien. Penelitian ini bertujuan untuk menganalisis data transkriptomik pasien NPC Indonesia menggunakan pipeline komputasional bertahap yang mengintegrasikan algoritma statistik edgeR dan pemodelan grafik jaringan STRING v12.0.

TINJAUAN PUSTAKA

Karsinoma nasofaring (nasopharyngeal carcinoma/NPC) merupakan keganasan epitel yang memiliki karakteristik epidemiologi dan molekuler yang berbeda antar-populasi. Di Asia Tenggara, termasuk Indonesia, NPC menjadi salah satu keganasan yang memiliki beban klinis penting. Perkembangan NPC dipengaruhi oleh interaksi antara faktor genetik, lingkungan, dan infeksi Epstein-Barr Virus (EBV). Meskipun penelitian mengenai NPC telah berkembang pesat, sebagian besar informasi molekuler masih berasal dari populasi Asia Timur, sehingga karakteristik transkriptomik NPC pada populasi Indonesia masih perlu dikaji lebih lanjut.

Teknologi RNA sequencing (RNA-seq) memberikan pendekatan komprehensif untuk mempelajari perubahan ekspresi gen pada jaringan kanker. Berbeda dengan microarray yang bergantung pada probe yang telah ditentukan sebelumnya, RNA-seq dapat mendeteksi transkrip dalam cakupan yang lebih luas serta memberikan resolusi dan rentang dinamis yang tinggi. Oleh karena itu, RNA-seq banyak digunakan untuk mengidentifikasi differentially expressed genes (DEGs) dan perubahan jalur biologis yang berkaitan dengan perkembangan kanker [8]. Dalam analisis RNA-seq, pemilihan metode alignment, kuantifikasi, dan analisis statistik dapat memengaruhi hasil akhir. Penggunaan HISAT2 untuk pemetaan sekuens dan edgeR untuk analisis ekspresi diferensial merupakan salah satu pendekatan yang telah digunakan dalam penelitian transkriptomik kanker.

Penelitian sebelumnya pada pasien NPC Indonesia menggunakan sembilan sampel yang terdiri atas tujuh jaringan NPC dan dua sampel kontrol normal. Dengan pendekatan RNA-seq menggunakan Illumina NextSeq 550 dan analisis HISAT2, HTSeq, edgeR, serta PANTHER, penelitian tersebut mengidentifikasi

1.956 gen yang mengalami upregulasi dan 90 gen yang mengalami downregulasi pada jaringan NPC. Gen-gen yang mengalami perubahan ekspresi terutama berkaitan dengan regulasi proliferasi sel melalui jalur apoptosis, IGF, Notch, dan PI3K. Temuan tersebut menunjukkan bahwa analisis transkriptomik dapat mengungkap perubahan molekuler yang berhubungan dengan karakteristik proliferasi NPC pada populasi Indonesia [9].

Analisis DEG kemudian dapat dikembangkan lebih lanjut dengan menghubungkan gen-gen yang mengalami perubahan ekspresi ke dalam jaringan interaksi protein-protein (protein-protein interaction/PPI). Pendekatan ini penting karena perubahan ekspresi satu gen tidak selalu dapat menjelaskan proses biologis secara individual. Hubungan antargen dapat memperlihatkan kelompok protein yang bekerja secara fungsional bersama serta membantu mengidentifikasi hub gene yang mempunyai posisi penting dalam jaringan. Dalam penelitian NPC, Wang et al. menggunakan RNA-seq pada lima sampel NPC dan enam sampel normal, kemudian menggabungkan analisis DEG, gene set enrichment analysis (GSEA), serta jaringan PPI. Penelitian tersebut mengidentifikasi 2.024 DEG dan beberapa hub gene, termasuk CXCL10, IGF1, MMP1, MMP3, ICAM1, dan IL13. Beberapa gen tersebut juga menunjukkan hubungan dengan prognosis dan karakteristik biologis NPC [10].

Pendekatan integratif juga telah digunakan untuk menggabungkan mRNA, microRNA, dan long non-coding RNA (lncRNA) dalam penelitian NPC. H et al. mengintegrasikan data transkriptomik dan beberapa dataset microarray untuk mengidentifikasi perubahan mRNA, miRNA, dan lncRNA pada NPC. Analisis tersebut menunjukkan keterlibatan jalur DNA replication, cell cycle, MAPK, PI3K-AKT, dan apoptosis. Selain itu, analisis PPI menghasilkan sejumlah hub gene yang berhubungan dengan proliferasi dan perkembangan NPC. Hasil tersebut menunjukkan bahwa integrasi beberapa tingkat informasi molekuler dapat memberikan gambaran yang lebih luas mengenai mekanisme biologis NPC [11].

Selain analisis PPI, pendekatan berbasis ko-ekspresi juga dapat digunakan untuk menemukan modul gen yang berhubungan dengan karakteristik klinis kanker. Pan dan Liu menggunakan weighted gene co-expression network analysis (WGCNA) pada data NPC dan mengidentifikasi modul yang berhubungan dengan grade histologis serta stadium tumor. Dari analisis tersebut diperoleh CRIP1, KITLG, MARK1, dan PGAP1 sebagai kandidat biomarker. Temuan ini memperlihatkan bahwa analisis berbasis jaringan dapat membantu mengurangi kompleksitas data transkriptomik dan memprioritaskan gen yang berpotensi memiliki nilai diagnostik atau prognostic [12].

Komponen penting lain dalam data transkriptomik adalah lncRNA. Molekul ini tidak menghasilkan protein, tetapi dapat berperan dalam pengaturan ekspresi gen melalui berbagai mekanisme, termasuk interaksi dengan miRNA, regulasi transkripsi, dan pengaruh

terhadap proses epigenetik. Dalam kanker, perubahan ekspresi lncRNA dapat membentuk jaringan regulasi yang berhubungan dengan proliferasi, invasi, metastasis, dan respons terhadap terapi. Studi mengenai lncRNA pada NPC menunjukkan bahwa kelompok molekul tersebut memiliki potensi sebagai biomarker dan sebagai regulator proses biologis kanker. Oleh sebab itu, keberadaan lncRNA dalam hasil DEG perlu dipertimbangkan secara terpisah dari gen pengode protein dan tidak seharusnya langsung dimasukkan ke dalam analisis PPI yang berbasis interaksi produk protein [13].

Database STRING dapat digunakan untuk memetakan hubungan fungsional antargenode protein berdasarkan berbagai sumber bukti, termasuk data eksperimen, ko-ekspresi, basis data, dan text mining. Pemodelan jaringan tersebut memungkinkan identifikasi struktur seperti sub-cluster, hub, dan isolated node [14]. Dengan demikian, integrasi hasil DEG dengan STRING dapat memberikan informasi tambahan mengenai organisasi fungsional gen yang mengalami perubahan ekspresi, dibandingkan hanya melihat daftar gen secara individual.

Berdasarkan penelitian terdahulu, sebagian besar studi NPC telah berhasil mengidentifikasi DEG, jalur biologis, hub gene, atau jaringan ceRNA melalui data dari populasi Asia Timur maupun database publik. Penelitian pada populasi Indonesia sebelumnya telah menggambarkan perubahan ekspresi gen dan kaitannya dengan proliferasi sel, tetapi analisis jaringan PPI yang secara khusus memetakan topologi gen pengode protein dari hasil DEG pada sampel NPC Indonesia masih terbatas. Selain itu, penelitian sebelumnya belum secara khusus menekankan pemisahan lncRNA dari gen pengode protein sebelum rekonstruksi jaringan PPI. Oleh karena itu, penelitian ini mengembangkan analisis transkriptomik NPC Indonesia melalui integrasi HISAT2, HTSeq, edgeR, dan STRING v12.0 untuk mengidentifikasi pola DEG serta struktur jaringan PPI pada gen-gen yang mengalami perubahan ekspresi. Pendekatan tersebut menghasilkan identifikasi sub-kluster SOX11-SALL3 dan GABRG2-KCNQ2 pada kelompok upregulated serta SERPINB3 sebagai central hub pada kelompok downregulated, sehingga memberikan perspektif berbasis jaringan terhadap kandidat biomarker molekuler NPC Indonesia.

METODOLOGI PENELITIAN

Penelitian ini menerapkan arsitektur pipeline pemrosesan data bioinformatika 4 tahap yang disetujui Komite Etika Penelitian Kesehatan FK-KMK UGM / RSUP Dr. Sardjito (No. KE/FK/0250/EC/2019):

Ingesti Data dan Kontrol Kualitas (QC)

Ekstraksi RNA total menggunakan RNeasy Mini Kit (Qiagen), dilanjutkan pembuatan pustaka cDNA dengan TruSeq RNA Library Prep Kit v2 (Illumina). Pustaka cDNA dievaluasi kualitatif dan kuantitatif menggunakan qPCR dan elektroforesis gel agarosa 2% untuk memastikan insert fragment berkisar 400 bp. Data

sekuens RNA utuh diproses pada platform Illumina NextSeq 550 (paired-end read, panjang 35–76 bp) dari BioProject NCBI accession PRJNA579573 yang mencakup 7 sampel biopsi NPC (RSK Dharmais, Jakarta) dan 2 sampel brushing nasofaring kontrol sehat (RSUP Dr. Sardjito, Yogyakarta). Evaluasi kualitas membaca sekuens dijalankan menggunakan algoritma FastQC.

Alignment Genome dan Kuantifikasi Transkrip

Pembacaan sekuens dipetakan ke genom rujukan manusia (*Homo sapiens* UCSC hg38) menggunakan algoritma alignment HISAT2 v2.1.0. Kuantifikasi jumlah read per gen dieksekusi oleh HTSeq v0.11.1 berdasarkan anotasi Ensembl *Homo sapiens*.

Analisis Ekspresi Gen Diferensial (DEG) berbasis Algoritma edgeR

Normalisasi ukuran pustaka silang antar-sampel dilakukan menggunakan algoritma Trimmed Mean of M-values (TMM). Pengujian hipotesis statistik mengimplementasikan model statistik Negative Binomial pada paket edgeR v3.28.0 di lingkungan RStudio, dengan kriteria statistik False Discovery Rate (FDR) < 0,05 dan absolute log2 Fold Change ($|\log_2 FC| > 1$).

Pemodelan Grafik Jaringan Protein (STRING v12.0)

Rekonstruksi jaringan interaksi protein-protein (PPI) dijalankan menggunakan algoritma confidence scoring STRING v12.0 pada threshold low confidence (> 0,150) untuk mengevaluasi topologi grafik (sub-cluster, central hub, dan isolated nodes) dari kelompok top 10 gen upregulated dan downregulated.

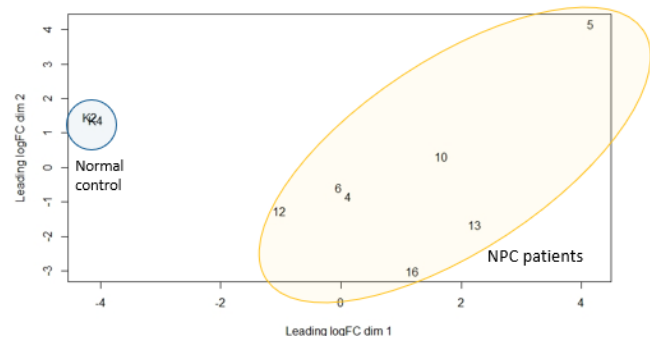
HASIL DAN PEMBAHASAN

Sekuensing RNA berpasangan (paired-end) dilakukan menggunakan platform Illumina NextSeq 550 pada sembilan sampel, yang menghasilkan panjang read antara 35 hingga 76 bp. Hasil evaluasi kualitas awal menggunakan FastQC menunjukkan skor Phred yang memadai untuk seluruh pustaka sekuensing. Oleh karena itu, data mentah langsung diproses ke tahap berikutnya tanpa perlu filtrasi atau pemotongan (trimming) basa berkualitas rendah. Rincian performa dan total luaran sekuensing untuk setiap sampel tercantum pada Tabel 1.

Tabel 1. Statistik Pemrosesan Data Sekuensing RNA (Illumina NextSeq 550)

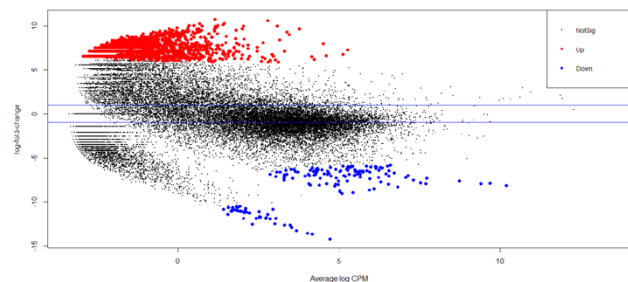
no	Nama Sampel	Kelompok	Total Sequence (bp)	Panjang Read (bp)	%GC
1	K2	Kontrol	33.097.617	35-76	53%
2	K4	Kontrol	169.898.666	35-76	50%
3	4	Kasus NPC	186.979.159	35-76	57%

4	5	Kasus NPC	20.413.210	35-76	56%
5	10	Kasus NPC	18.128.296	35-76	49%
6	12	Kasus NPC	87.652.547	35-76	56%
8	13	Kasus NPC	67.777.774	35-76	53%
9	6	Kasus NPC	36.803.557	35-76	53%
10	16	Kasus NPC	64.223.909	35-76	48%



Gambar 1. Plot Multidimensional Scaling (MDS) Leading LogFC Sampel NPC dan Kontrol

Analisis komponen eksploratif pada Gambar 1 menunjukkan kluster spasial yang terpisah tegas antara kelompok kontrol sehat (K2, K4 pada Dimensi 1, menjelaskan 36% varians total) dan kelompok kasus NPC (sampel 4, 5, 10, 12, 13, 6, 16 pada Dimensi 2, menjelaskan 20% varians). Pemisahan koordinat yang tajam ini mengonfirmasi pergeseran transkriptomik yang masif pada jaringan tumor nasofaring.

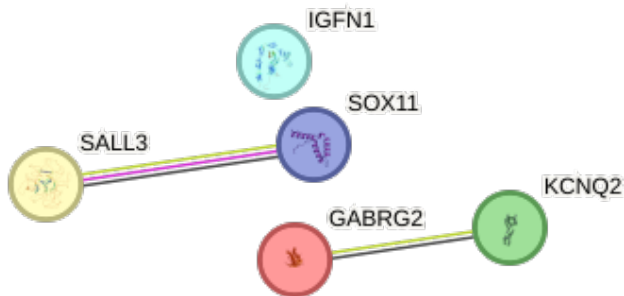


Gambar 2. Sebaran Ekspresi Gen Diferensial (DEG) Antara NPC dan Kontrol

Visualisasi sebaran logCPM vs log2FC pada Gambar 2 memperlihatkan distribusi 36.209 transkrip. Titik merah mewakili 2.043 gen ter-upregulasi signifikan, titik biru mewakili 156 gen ter-downregulasi signifikan (FDR < 0,05; $\log_2 FC > 1$), dan titik hitam merepresentasikan 34.010 gen yang tidak signifikan.

Hasil penapisan (filtering) komputasional terhadap top 10 upregulated genes dari pipeline analisis data NGS NPC menunjukkan karakteristik data heterogen. Lima entitas teratas teridentifikasi sebagai molekul long non-coding RNA (lncRNA: LINC02555, MIR4458HG, LINC01115, AC005082.1, dan AL022316.1). Karena

atribut entitas lncRNA ini tidak memiliki produk translasi berupa sekuens asam amino (protein), entitas tersebut secara otomatis diekskusi dari skema data preprocessing sebelum diinput ke dalam algoritma pemodelan jaringan interaksi protein-protein (PPI) berbasis graph pada STRING v12.0. Langkah eliminasi ini merupakan kriteria inklusi metodologis yang krusial, mengingat pangkalan data STRING beroperasi secara eksklusif pada pemetaan node-edge antar-produk proteomic [15]. Walaupun demikian, secara bioinformatika sistem, variabel lncRNA yang terdeteksi ini berpotensi bertindak sebagai master regulator atau molekul ceRNA (competing endogenous RNA) dalam arsitektur jaringan epigenetik NPC [16].



Gambar 3. Pemodelan Grafik Jaringan Interaksi Protein STRING v12.0 (Top 10 Gen Upregulated)

Rekonstruksi dan visualisasi graph network PPI terhadap lima gen pengode protein (SOX11, SALL3, GABRG2, KCNQ2, dan IGFN1) pada Gambar 3 menghasilkan topologi jaringan dengan pembentukan dua sub-network (sub-kluster) independen yang terpisah secara topologis:

Sub-kluster Faktor Transkripsi (SOX11-SALL3)

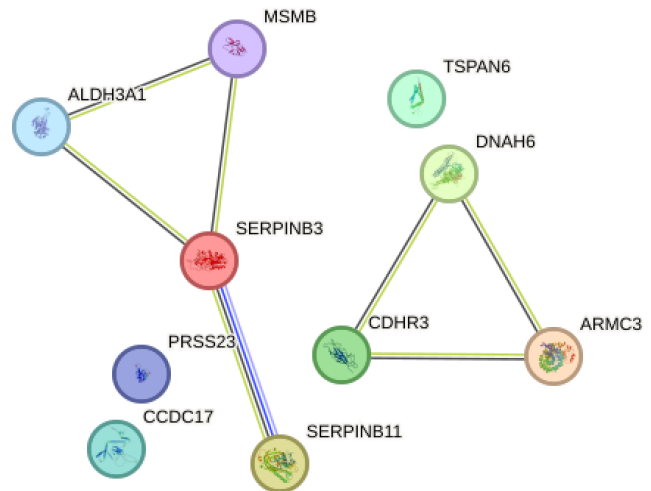
Edge yang menghubungkan simpul (node) SOX11 dan SALL3 dikonfigurasi berdasarkan bobot skor komputasional dari integrasi data eksperimental, data ko-ekspresi, dan ekstraksi informasi teks (text mining). Secara fungsional, kedua simpul ini merepresentasikan hubs regulator transkripsional yang mengontrol pemrograman ulang status seluler. Dalam model biokomputasi kanker, de-regulasi hub SOX11–SALL3 sering berkorelasi dengan modulasi status metilasi promoter yang memicu tumorigenesis [17].

Sub-kluster Reseptor/Kanal Ion (GABRG2-KCNQ2)

Edge pada modul ini dibangun melalui evidence score dari analisis ko-ekspresi genetik dan text mining. Dari kacamata pathway informatics, meskipun GABRG2 dan KCNQ2 secara klasik terpetakan pada jalur sinyal eksitabilitas membran saraf, fenomena aberrant expression dari kanal ion ini sering kali terdeteksi sebagai fitur komputasional (channelopathy) yang meregulasi fluks kation dan potensial membran sel kanker [18].

Di sisi lain, IGFN1 teridentifikasi secara komputasional sebagai simpul terisolasi (isolated node/degree of zero). Hal ini menunjukkan bahwa

dalam adjacency matrix atau pangkalan data relasional STRING v12.0, belum terekam adanya keterhubungan (edge/link) berbasis bukti eksperimental maupun komputasional yang signifikan antara IGFN1 dengan kedua sub-kluster lainnya. Dari perspektif eksplorasi data (data mining), ketiadaan edge ini memposisikan IGFN1 sebagai entitas independen yang berpotensi menjadi kandidat novel biomarker dalam pemodelan sistem biologi NPC.



Gambar 4. Pemodelan Grafik Jaringan Interaksi Protein STRING v12.0 (Top 10 Gen Downregulated)

Rekonstruksi topologi jaringan PPI berbasis graph pada top 10 downregulated genes (Gambar 4) memperlihatkan arsitektur interaksi yang terfragmentasi menjadi dua modul utama dan dua simpul terisolasi (isolated nodes). Berbeda dari kelompok upregulated, sepuluh entitas downregulated ini secara utuh direpresentasikan oleh gen pengode protein, sehingga seluruh entitas dipertahankan dalam skema data preprocessing tanpa mengalami eliminasi berbasis tipe biotipe. Analisis topologi jaringan dan pemrosesan data interaksi memodelkan struktur graph sebagai berikut:

Modul Protease & Respons Metabolik (Central Hub Node)

Algoritma STRING v12.0 mengidentifikasi SERPINB3 sebagai central hub node dengan derajat konektivitas (degree centrality) tertinggi dalam modul ini. SERPINB3 membentuk star-like subgraph yang terhubung langsung (direct edges) melalui interaksi fungsional dengan ALDH3A1, MSMB, dan SERPINB11. Secara biologis-sistem, downregulation pada hub SERPINB3 dan paralognya (SERPINB11) menandai deaktivasi sistem inhibisi cysteine/serine protease epitelial, yang berakibat pada deregulasi homeostasis proteostasis dan rusaknya pertahanan mukosa nasofaring (Weidle et al., 2018). Di sisi lain, ko-penurunan ekspresi ALDH3A1 (enzim metabolisme aldehid) mengindikasikan

gangguan pada modul pertahanan terhadap stres oksidatif seluler [19]

Sub-kluster Silia & Adhesi Epitel (Closed Triangular Graph)

Rekonstruksi jaringan mengekstrak sebuah clique berukuran tiga (triad) yang membentuk struktur grafik segitiga tertutup (closed-loop motif) antara DNAH6, CDHR3, dan ARMC3. Dalam teori grafik bioinformatika, motif triad tertutup menunjukkan tingkat ko-ekspresi dan redundansi fungsional yang sangat tinggi (high clustering coefficient). Modul ini bertanggung jawab penuh atas integritas struktur siliogenesis dan adhesi seluler epitel respiratorik. Downregulasi terkoordinasi pada ketiadaan simpul dalam motif ini (DNAH6 sebagai komponen motor dynein, CDHR3 sebagai molekul cadherin mukosa, dan ARMC3 sebagai protein siliari) mencerminkan sinyal komputasional dari kerusakan masif pada arsitektur silia serta hilangnya polaritas epitel nasofaring [20]

Simpul Terisolasi (Isolated Nodes / Degree of Zero)

Entitas TSPAN6 (Tetraspanin-6) dan CCDC17, serta interaksi minimal pada PRSS23, terpetakan sebagai simpul terpisah tanpa edge yang menghubungkannya ke modul utama pada ambang batas skor keyakinan (stringency threshold) default STRING v12.0. Ketiadaan keterhubungan (null adjacency) ini mengindikasikan bahwa keterlibatan TSPAN6 (penekan sinyal Wnt/beta-catenin) dan PRSS23 dalam pemodelan patogenesis NPC berjalan melalui jalur independen di luar sumbu utama SERPINB3–siliogenesis [21]

Secara keseluruhan, penekan ekspresi yang terkoordinasi pada modul central hub SERPINB3 dan triad motif silia-adhesi mengonfirmasi secara *in silico* hilangnya fungsi proteksi protease epitelial serta rusaknya integritas arsitektur mukosa nasofaring pada pasien NPC.

Penerapan pipeline bioinformatika komputasional yang mengintegrasikan pemetaan HISAT2, kuantifikasi HTSeq, analisis statistik edgeR, dan pemodelan grafik STRING v12.0 terbukti sangat efektif dalam mengekstraksi arsitektur transkriptomik dan jaringan protein pasien NPC Indonesia. Identifikasi central hub SERPINB3 dan sub-kluster SOX11-SALL3 menyediakan kandidat biomarker komputasional potensial untuk pengembangan sistem informasi diagnosis presisi berbasis data omik.

KESIMPULAN DAN SARAN

Penelitian ini berhasil menerapkan pipeline bioinformatika yang mengintegrasikan HISAT2, HTSeq, edgeR, dan STRING v12.0 untuk menganalisis perubahan ekspresi gen dan jaringan interaksi protein pada data RNA-seq pasien karsinoma nasofaring Indonesia. Analisis ekspresi diferensial mengidentifikasi 2.043 gen yang mengalami upregulasi dan 156 gen yang mengalami downregulasi berdasarkan kriteria $FDR < 0,05$ dan $|\log_2 FC| > 1$. Lima dari sepuluh gen dengan ekspresi tertinggi merupakan lncRNA sehingga

tidak dimasukkan dalam pemodelan PPI berbasis protein.

Analisis jaringan STRING v12.0 pada gen pengode protein yang mengalami upregulasi menghasilkan dua sub-kluster utama, yaitu SOX11-SALL3 dan GABRG2-KCNQ2, sedangkan IGFN1 teridentifikasi sebagai simpul terisolasi. Pada kelompok gen yang mengalami downregulasi, SERPINB3 teridentifikasi sebagai central hub yang berinteraksi dengan ALDH3A1, MSMB, dan SERPINB11. Selain itu, DNAH6, CDHR3, dan ARMC3 membentuk modul yang berkaitan dengan struktur silia dan adhesi epitel. Hasil tersebut menunjukkan bahwa perubahan transkriptomik NPC tidak hanya berupa perubahan ekspresi gen secara individual, tetapi juga membentuk pola jaringan fungsional yang dapat digunakan untuk memprioritaskan kandidat biomarker molekuler.

Keunggulan penelitian ini adalah integrasi analisis ekspresi diferensial dengan pemodelan jaringan PPI serta penggunaan data transkriptomik pasien NPC Indonesia. Keterbatasan utama penelitian adalah jumlah sampel yang relatif kecil dan analisis yang masih bersifat *in silico*, sehingga kandidat biomarker yang diperoleh belum dapat dianggap sebagai biomarker klinis tervalidasi tanpa validasi eksperimental dan kohort independen.

Penelitian selanjutnya disarankan menggunakan jumlah sampel NPC dan kontrol yang lebih besar serta melibatkan kohort independen untuk meningkatkan kekuatan statistik dan menguji reproduktibilitas hasil. Kandidat hub gene, termasuk SERPINB3, SOX11, SALL3, GABRG2, dan KCNQ2, perlu divalidasi menggunakan metode eksperimental seperti qRT-PCR, imunohistokimia, atau analisis proteomik. lncRNA yang teridentifikasi pada kelompok upregulated juga perlu dikaji lebih lanjut melalui analisis ceRNA, interaksi miRNA, dan eksperimen fungsional untuk mengetahui perannya dalam perkembangan NPC. Penelitian berikutnya juga dapat mengintegrasikan data klinis, genomik, epigenomik, dan proteomik sehingga kandidat biomarker yang diperoleh dapat dievaluasi lebih lanjut untuk kebutuhan diagnosis, prognosis, dan pengembangan terapi presisi pada pasien NPC Indonesia.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kemenristekdikti dan Universitas Gadjah Mada atas dukungan pendanaan Hibah Penelitian Fundamental (No. 2816/UNI.DITLIT/DIT-LIT/LT/2019), serta seluruh staf klinis RSK Dharmas dan RSUP Dr. Sardjito.

DAFTAR PUSTAKA

- [1] R. Andrijes, R. K. Hejmadi, M. Pugh, S. Rajesh, V. Novitskaya, M. Ibrahim, F. Berditchovski, "Tetraspanin 6 is a regulator of carcinogenesis in colorectal cancer," *Proc. Natl. Acad. Sci.*, vol. 118, no. 39, p. e2011411118, 2021. doi: 10.1073/pnas.2011411118.
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185

- countries," *CA Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018. doi: 10.3322/caac.21492.
- [3] S. M. Chan, I. C. Paterson, L. F. Yap, "Nasopharyngeal carcinoma in Southeast Asia: Current landscape and future priorities," *Br. J. Biomed. Sci.*, vol. 82, p. 15902, 2025. doi: 10.3389/bjbs.2025.15902.
 - [4] E. T. Chang, W. Ye, Y. X. Zeng, dkk., "The evolving epidemiology of nasopharyngeal carcinoma," *Cancer Epidemiol. Biomarkers Prev.*, vol. 30, no. 6, pp. 1035–1047, 2021. doi: 10.1158/1055-9965.EPI-20-1702.
 - [5] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, A. Mortazavi, "A survey of best practices for RNA-seq data analysis," *Genome Biol.*, vol. 17, no. 1, p. 13, 2016. doi: 10.1186/s13059-016-0881-8.
 - [6] P. F. Langton, S. Kakugawa, J. P. Vincent, "Making, exporting, and modulating Wnts," *Trends Cell Biol.*, vol. 26, no. 10, pp. 756–765, 2016. doi: 10.1016/j.tcb.2016.05.011.
 - [7] D. C. Lin, X. Meng, M. Hazawa, H. P. Koeffler, "The genomic landscape of nasopharyngeal carcinoma," *Nat. Genet.*, vol. 46, no. 8, pp. 866–871, 2014. doi: 10.1038/ng.3006.
 - [8] R. Lutter, L. Ravanetti, "Cadherin-related family member 3 (CDHR3) drives differentiation of ciliated bronchial epithelial cells and facilitates rhinovirus C infection, although with a little help," *J. Allergy Clin. Immunol.*, vol. 144, no. 4, pp. 926–927, 2019. doi: 10.1016/j.jaci.2019.07.021.
 - [9] G. Muzio, M. Maggiora, E. Paiuzzi, M. Oraldi, R. A. Canuto, "Aldehyde dehydrogenases and cell proliferation," *Free Radic. Biol. Med.*, vol. 52, no. 4, pp. 735–746, 2012. doi: 10.1016/j.freeradbiomed.2011.11.033.
 - [10] I. M. Nafisa, M. S. Utama, M. A. Sunardi, dkk., "Profile of nasopharyngeal cancer patients who underwent radiotherapy in Dr. Hasan Sadikin General Hospital Bandung," *Indones. J. Cancer*, vol. 16, no. 2, p. 88, 2022. doi: 10.33371/ijoc.v16i2.869.
 - [11] N. Prevarskaya, R. Skryma, Y. Shuba, "Ion channels in cancer: Are cancer hallmarks oncochannelopathies?" *Physiol. Rev.*, vol. 98, no. 2, pp. 559–621, 2018. doi: 10.1152/physrev.00044.2016.
 - [12] M. F. Rai, E. D. Tycksen, L. J. Sandell, R. H. Brophy, "Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling," *J. Orthop. Res.*, vol. 36, no. 2, pp. 484–497, 2018. doi: 10.1002/jor.23661.
 - [13] A. Raymos, S. F. Boesoirie, Y. A. Dewi, "Epidemiology of nasopharyngeal carcinoma: Understanding the interplay of genetic and environmental factors," dalam *Proc. PERHATIKL 2022*, him. 212–217, 2023. doi: 10.2991/978-94-6463-280-4_39.
 - [14] S. Romero-Garcia, H. Prado-Garcia, A. Carlos-Reyes, "Role of DNA methylation in the resistance to therapy in solid tumors," *Front. Oncol.*, vol. 10, p. 1152, 2020. doi: 10.3389/fonc.2020.01152.
 - [15] R. Stark, M. Grzelak, J. Hadfield, "RNA sequencing: The teenage years," *Nat. Rev. Genet.*, vol. 20, no. 11, pp. 631–656, 2019. doi: 10.1038/s41576-019-0150-2.
 - [16] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, 2021. doi: 10.3322/caac.21660.
 - [17] D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Patel, A. L. Gable, C. V. Mering, "The STRING database in 2023: Protein-protein networks integrated with functional predictions for all sequenced organisms," *Nucleic Acids Res.*, vol. 51, no. D1, pp. D638–D646, 2023. doi: 10.1093/nar/gkac1000.
 - [18] Y. Tang, X. He, "Long non-coding RNAs in nasopharyngeal carcinoma: Biological functions and clinical applications," *Mol. Cell. Biochem.*, vol. 476, no. 9, pp. 3537–3550, 2021. doi: 10.1007/s11010-021-04176-4.
 - [19] The HUGO Pan-Asian SNP Consortium, "Mapping human genetic diversity in Asia," *Science*, vol. 326, no. 5959, pp. 1541–1545, 2009. doi: 10.1126/science.1177074.
 - [20] U. H. Weidle, F. Birzele, G. Tiefenthaler, "Potential of protein-based anti-metastatic therapy with serpins and inter a-trypsin inhibitors," *Cancer Genomics Proteomics*, vol. 15, no. 4, pp. 225–238, 2018. doi: 10.21873/cgp.20081.
 - [21] R. Zhang, Y. He, B. Wei, dkk., "Nasopharyngeal carcinoma burden and its attributable risk factors in China," *Int. J. Environ. Res. Public Health*, vol. 20, no. 4, p. 2926, 2023. doi: 10.3390/ijerph20042926.