

IDENTIFIKASI VARIASI DAN EKSPRESI GEN PADA STROKE ISKEMIK DENGAN PENDEKATAN BIOINFORMATIKA

IDENTIFICATION OF GENE VARIATION AND EXPRESSION IN ISCHEMIC STROKE WITH BIOINFORMATICS APPROACH

Febriana Astuti¹, Rafiastiana Capritasari², Delfi Iskardyani³, Ilham Alifiar⁴, Lalu Muhammad Irham⁵

¹ D3 Farmasi, Politeknik Kesehatan TNI AU Adisutjipto

Jalan Majapahit Blok-R Lanud Adisutjipto Yogyakarta

² S1 Farmasi, Universitas Muhammadiyah Magelang

Jl. Tidar No.21, Magersari, Kec. Magelang Sel., Kota Magelang, Jawa Tengah

³ D3 Radiologi, Politeknik Kesehatan TNI AU Adisutjipto

Jalan Majapahit Blok-R Lanud Adisutjipto Yogyakarta

⁴ S1 Farmasi, Universitas Bakti Tunas Husada

Jalan Cilolohan No.36 Tasikmalaya

⁵ S2 Farmasi, Universitas Ahmad Dahlan

Jalan Prof. Dr. Soepomo Sh, Warungboto, Yogyakarta

E-mail korespondensi: febrianafarmasis@gmail.com

ABSTRACT

Stroke is the leading cause of cerebrovascular disease (CVD) and is a considerable cause of death and disability in society. Approximately 80-85% of all stroke cases are caused by ischemic stroke. Genetic factors play an important role in early-onset stroke. A family history of stroke is associated with a higher prevalence and incidence of stroke. Until now, research on ischemic stroke has been limited to identifying gene variations that affect it, but has not produced reports that specifically show which gene variations are most significant in causing ischemic stroke, especially those that are missense/nonsense. Therefore, this study used genomic databases in bioinformatic analysis to identify gene variations that have the strongest correlation with ischemic stroke. The aim of this research is to identify genetic variations associated with ischemic stroke, prioritizing these genetic variations based on their susceptibility levels through the utilization of GWAS catalog and integration of multiple bioinformatics databases. The method employed in this study utilizes several databases, including the GWAS catalog, PheWAS catalog, HaploReg (v41.), and the GTEx portal. The results showed that GCKR, ALDH2, and F5 genes were significantly associated with ischemic stroke, and showed high expression levels in several tissue types such as liver, subcutaneous adipose, ovary, and whole blood. Gene variation based on the frequency of allelic variants (rs1260326, rs671, rs6025) associated with ischemic stroke showed the highest tissue expression in liver.

Keywords: Gene Expression, Ischaemic Stroke, Bioinformatics

Diterima: 09-02-2023

Direview: 12-02-2023

Diterbitkan: 21-02-2023

ABSTRAK

Stroke adalah penyebab utama penyakit serebrovaskular (CVD) dan merupakan penyebab kematian dan kecacatan yang cukup besar dalam masyarakat. Sekitar 80-85% dari seluruh kasus stroke disebabkan oleh iskemik. Faktor genetika memainkan peran penting dalam stroke yang terjadi lebih awal. Riwayat stroke dalam keluarga dikaitkan dengan prevalensi dan kejadian stroke yang lebih tinggi. Hingga saat ini, penelitian mengenai stroke iskemik masih terbatas pada pengidentifikasian variasi gen yang mempengaruhi, namun belum menghasilkan laporan yang secara spesifik menunjukkan variasi gen mana yang paling signifikan dalam menyebabkan stroke iskemik, terutama yang bersifat *missens/nonsense*. Oleh karena itu, pada penelitian ini digunakan database genomik dalam menganalisis bioinformatik untuk mengidentifikasi variasi gen yang memiliki korelasi paling kuat dengan stroke iskemik. Metode yang digunakan dalam penelitian ini memanfaatkan sejumlah database, termasuk katalog GWAS, katalog PheWAS, HaploReg (v41.), dan portal GTEx. Hasil penelitian menunjukkan bahwa gen GCKR, ALDH2, dan F5 terkait secara signifikan dengan stroke iskemik, serta menunjukkan tingkat ekspresi yang tinggi pada beberapa jenis jaringan seperti hati, adiposa subkutan, ovarium, dan seluruh darah. Variasi gen berdasarkan frekuensi varian alel (rs1260326, rs671, rs6025) terkait stroke iskemik menunjukkan ekspresi jaringan tertinggi di liver.

Kata Kunci: Ekspresi Gen, Stroke Iskemik, Bioinformatik

PENDAHULUAN / INTRODUCING

Stroke iskemik merupakan penyebab kematian kedua terbanyak setelah penyakit jantung di dunia. Stroke iskemik terjadi karena penyumbatan pada arteri yang menyebabkan berkurangnya aliran oksigen ke otak (Stanzione *et al.*, 2020). Etiologi stroke iskemik bersifat multifaktorial. Faktor risiko stroke iskemik ada yang dapat dimodifikasi, seperti hipertensi, merokok, diabetes, dan hiperlipidemia, sedangkan faktor yang tidak dapat dimodifikasi seperti genetika, usia, jenis kelamin (Ekkert *et al.*, 2022). Faktor genetika memainkan peran penting dalam stroke yang terjadi lebih awal. Riwayat stroke dalam

keluarga dikaitkan dengan prevalensi dan kejadian stroke yang lebih tinggi. Faktor genetik diketahui mempengaruhi berbagai aspek stroke iskemik, termasuk perubahan risiko terjadinya stroke individu, modulasi respon pengobatan, dan efektivitas pemulihan fungsional pasca stroke (Zhang *et al.*, 2023). Patofisiologi stroke iskemik melibatkan dua tahap penting, yaitu (i) gangguan pasokan darah ke otak, dan (ii) perkembangan hipoksia dan nekrosis jaringan otak. Pada stroke iskemik, penurunan awal pasokan darah ke otak dan/atau gangguan metabolisme oksigen, glukosa, dan lipid mengakibatkan terjadinya iskemia serebral lokal. Hal ini diikuti oleh

serangkaian proses patofisiologi lainnya, seperti eksitotoksisitas, asidotoksisitas, ketidakseimbangan ion, stres oksidatif dan nitratif, neuroinflamasi, dan apoptosis, yang pada akhirnya mengakibatkan kematian neuron (Zhang *et al.*, 2023).

Pada stroke iskemik berbeda dengan diabetes melitus dan penyakit jantung yang patogenesisnya disebabkan banyak hal. pada stroke iskemik tidak ada biomarker praklinis yang menjadi dasar untuk mempelajari proses molekuler dan fisiologis yang mengarah ke kejadian stroke. Karena predisposisi genetik terhadap stroke telah diketahui secara luas, 1 strategi saat ini untuk mempelajari patogenesis stroke adalah memulai dengan mengidentifikasi gen yang terkait dengan stroke dan kemudian menyelidiki fungsi gen-gen ini. Untuk mencapai tujuan ini, banyak penelitian gen kandidat dan studi asosiasi genom (GWAS) stroke telah dilakukan (Cheng *et al.*, 2014). Tujuan penelitian ini untuk mengidentifikasi variasi gen yang berhubungan dengan stroke iskemik memprioritaskan variasi gen tersebut berdasarkan tingkat kerentanannya melalui pemanfaatan GWAS katalog dan integrasi beberapa basis data bioinformatika

Kemajuan teknologi dalam mengidentifikasi kerentanan genomik saat ini memberikan berbagai kemudahan, terutama dalam deteksi variasi gen yang merupakan faktor utama dalam kesalahan urutan pada susunan basa DNA yang terdiri dari huruf A-T-G dan C. Salah satu sumber informasi yang membahas variasi gen adalah *Genome-Wide Association Studies* (GWAS) (Furqan *et al.*, 2020).

Katalog GWAS adalah database yang menghimpun hasil dari studi GWAS di seluruh dunia. Pendekatan GWAS ini sangat praktis karena mampu mengidentifikasi SNP (*Single Nucleotide Polymorphism*) dalam jumlah yang sangat besar (Lifia *et al.*, 2023)

METODE PENELITIAN

Identifikasi variasi genetik sangat penting tidak hanya untuk memahami struktur gen tetapi untuk mendapatkan informasi biologi penyakit. Dalam penelitian ini, kami menggunakan beberapa pendekatan berbasis bioinformatika untuk mengintegrasikan varian yang terkait dengan stroke iskemik. *Single-Nucleotide Polymorphisms* (SNP) yang terkait dengan penyakit stroke iskemik diperoleh dari *National Human Genome Research Institute* (NHGRI) GWAS Katalog yang diunduh pada tanggal 15 Oktober 2023 dengan istilah " ischemic stroke " (**EFO ID: EFO_0002690**) (GWAS, n.d.). Pada studi ini diperoleh hasil 577 SNP yang terkait dengan stroke iskemik. Kemudian hasil difokuskan pada varian *missens* karena yang berpotensi untuk terjadi mutasi atau perubahan asam amino atau protein. Hasil yang memenuhi varian *missens* diperoleh 41 SNP. Identifikasi jenis SNP yang bersifat *missens* dengan menggunakan database HaploReg v4.2 (HaploReg, n.d.). Dari 41 SNP tersebut difokuskan kembali berdasarkan $p\text{-value} < 10^{-8}$ untuk membedakan positif sejati dari positif palsu. Hasilnya diperoleh 13 SNP yang memenuhi kriteria tersebut.

Database bioinformatika HaploReg v4.2 dirancang untuk kemudahan penggunaan dan integrasi, terutama dalam tujuan identifikasi SNP. Pengidentifikasi SNP yang terkait dengan stroke iskemik dilakukan dengan menggunakan kriteria $p\text{-value} < 10^{-8}$ dari katalog GWAS.

Langkah berikutnya adalah mengevaluasi SNP dengan memanfaatkan GTEX Portal untuk mengidentifikasi profil ekspresi genetik dalam jaringan yang berkaitan dengan stroke iskemik (Portal, n.d.). Dengan menggunakan jenis alel dari setiap SNP, distribusi kerentanan gen pada populasi diukur. Frekuensi alel untuk populasi Eropa, Afrika, Amerika, dan Asia dihitung dengan database genetik HaploReg v 4.2. (HaploReg, n.d.) dan browser Ensembl Genome (Institute, n.d.) .

HASIL DAN PEMBAHASAN

Dari hasil penelusuran menggunakan GWAS untuk menganalisis SNP pada stroke iskemik. Pendekatan GWAS berdasarkan kemampuannya dapat menyaring ratusan ribu. Salah satu penanda genetik yang digunakan untuk menentukan kelainan penyakit adalah *Single Nucleotide Polymorphism* (SNP) yang ditemukan di seluruh genom individu (Bush & Moore, 2012). Data SNP yang memenuhi kriteria missens sebanyak 41 SNP. Penerapan metode komputasi pada SNP *missens* manusia menunjukkan bahwa SNP *missens* memiliki dampak merusak protein. Analisis yang dilakukan menunjukkan bahwa SNP *missens* berkontribusi berbagai fungsi protein dan kemungkinan terlibat dalam mekanisme penyakit. Namun, SNP *missense* yang

memiliki dampak lebih rendah juga dapat berkontribusi pada penyakit sifat kompleks meskipun secara efektif netral (Pal & Moul, 2015). Dari 41 SNP missens, 13 SNP memiliki $p\text{-value} < 10^{-8}$ (Tabel 1). Nilai $p\text{-value} < 10^{-8}$ dipilih dalam penelitian ini untuk memastikan bahwa hasilnya memiliki tingkat kepercayaan yang tinggi, dan bahwa SNP yang diidentifikasi benar-benar memiliki korelasi yang signifikan dengan stroke iskemik. Selain itu, mengurangi kemungkinan kesalahan interpretasi dan menghindari nilai positif yang salah terkait dengan hubungan antara SNP dengan fenotip (Kim *et al.*, 2021).

Gambar 1. Hasil katalog GWAS 13 SNP yang berkaitan dengan stroke iskemik.

No	Kode Gen	SNP	P-Value
1	PLCB2	rs200787930	4,00E ⁻⁰⁹
2	VPS33B	rs199921354	6,00E ⁻⁰⁹
3	ADGRL3	rs192210727	6,00E ⁻⁰⁹
4	TMOD4	rs115287176	9,00E ⁻⁰⁹
5	PMF1- BGLAP, PMF1	rs1052053	8,00E ⁻¹¹
6	GCKR	rs1260326	1,00E ⁻³⁰
7	SH2B3, ATXN2	rs3184504	6,00E ⁻¹⁵
8	VWF	rs1063856	2,00E ⁻²⁵
9	CLEC4M	rs2277998	9,00E ⁻¹⁵
10	KNG1, HRG-AS1	rs710446	5,00E ⁻³⁰⁷
11	HNF4A	rs1800961	1,00E ⁻¹⁰
12	ALDH2	rs671	2,00E ⁻¹⁸
13	F5	rs6025	4,00E ⁻¹³⁷

SNP yang diperoleh berdasarkan identifikasi yang dilakukan kemudian dibandingkan menggunakan database Ensembl (Institute, n.d.)

Ensembl merupakan database yang lebih khusus memprediksi mutasi protein yaitu Polyphen2 (Puspitaningrum *et al.*, 2022). Database ini mengelompokkan protein

menjadi 3 kategori berdasarkan skor dan tingkat kerusakan SNP pada protein. Skor 0,00- 0,15 (jinak), 0,15-0,85 kemungkinan merusak, dan 0,85-1,00 mungkin merusak. Hasil dentifikasi SNPs pada stroke iskemik dari 13 SNP yang memiliki nilai $p\text{-value} < 10^{-8}$ ada 3 SNP yang memiliki resiko tinggi pada penyakit stroke iskemik. tiga SNP tersebut yaitu rs1260326, rs671 dan rs6025 dengan kode GCKR, ALDH2 dan F5. Skore tertinggi dari ketiga SNP tersebut adalah rs6025 dengan nilai 1,00 dan prediksi adalah *probably damaging*.

Tabel 2. Identifikasi SNPs stroke iskemik dan efeknya pada tingkat protein

SNP	Chromo	Gene	Score	Prediction
rs1260326	Chr2	GCKR	0.735	<i>Possibly Damaging</i>
rs671	chr12	ALDH2	0.761	<i>Possibly Damaging</i>
rs6025	Chr1	F5	1.00	<i>Probably Damaging</i>

Selanjutnya, peneliti menggunakan database GTEx untuk mengidentifikasi eQTL yang terkait dengan ekspresi gen stroke iskemik. Tabel 2 menunjukkan beberapa allele kecil yang terkait dengan penyakit stroke iskemik. Rs1260326, Rs671, dan Rs6025 adalah

beberapa SNP yang ditemukan memiliki ekspresi tinggi. Berdasarkan data website HaploReg v4.2., peneliti menemukan frekuensi alel yang terkait dengan penyakit stroke iskemik di berbagai populasi di seluruh dunia, seperti Afrika, Amerika, Asia, dan Eropa (Tabel 3). Tujuan pemetaan ini adalah untuk mengetahui bagaimana kerentanan itu tersebar di berbagai populasi di seluruh dunia. Frekuensi Alel untuk setiap SNP berbeda dari populasi ke populasi. Frekuensi alel yang terkait dengan gen GCKR pada stroke iskemik ditentukan oleh SNP rs1260326, yang menunjukkan bahwa di Afrika 0,88% jauh lebih tinggi daripada 0,59% di Amerika Serikat, 0,59% di Eropa, dan 0,44% di Asia. Selain itu jenis SNP rs671 menunjukkan bahwa Asia 0,22% yang paling rentan untuk kasus stroke iskemik dan belum ditemukan sama sekali di Amerika, Eropa dan Afrika. Pada SNP rs6025 frekuensi alel terkait gen F5 menunjukkan bahwa di Afrika dan Asia 1% sedangkan di Amerika dan Eropa 0,99%. Metode genomik dapat menggunakan hasil alel untuk menentukan kerentanan penyakit (Shallis *et al.*, 2018)

Tabel 3. Frekuensi alel pada setiap SNP

No	SNP	Allele		African	American	Asean	Europa	chr
		REF allele	ALT allele					
1	rs1260326	T	C	0.88	0.59	0.44	0.59	Chr2
2	rs671	G	A	0	0	0.22	0	Chr12
3	rs6025	T	C	1	0.99	1	0.99	Chr1

Peneliti menggunakan e-QTL menggunakan database GTEx portal (Portal, n.d.) untuk mengevaluasi perbedaan genetik dalam stroke iskemik di berbagai jaringan. Tingkat ekspresi

gen ditunjukkan pada gambar 1-3 untuk masing-masing jaringan. Tujuan penggunaan data ini adalah untuk mengidentifikasi jenis ekspresi gen pada berbagai jaringan. Gen-gen

Gen ALDH2 termasuk jenis protein aldehyd dehidrogenase, yang merupakan enzim kedua dalam jalur metabolisme alkohol. Ada dua isoform utama aldehyd dehidrogenase hati, yaitu sitosolik dan mitokondria, yang dapat dibedakan berdasarkan elektroforesis, kinetika, dan lokalisasi sel. Orang Kaukasia umumnya memiliki kedua isoform tersebut, sementara sekitar 50% orang Asia Timur hanya memiliki isoform sitosolik. Tingginya frekuensi keracunan alkohol akut di kalangan orang Asia Timur dapat dikaitkan dengan ketiadaan isoform mitokondria yang aktif secara katalitik. Kekurangan isoform mitokondria ini juga dapat meningkatkan risiko terkena berbagai jenis kanker karena paparan yang lebih tinggi terhadap asetaldehida. Gen ini mengkode isoform mitokondria yang memiliki afinitas rendah terhadap asetaldehida, dan terletak dalam matriks mitokondria. Terdapat juga variasi transkrip yang menghasilkan berbagai isoform yang berbeda melalui penyambungan alternatif (NCBI, 2018).

Pada mutasi gen ALDH2 diperoleh hasil yang menunjukkan bahwa mutase pada gen tersebut

dapat menyebabkan kerusakan pada beberapa jaringan. Namun yang paling dominan terjadi pada jaringan *liver*, kelenjar susu, jaringan adiposa yang ada dibawah kulit, dan jaringan adiposa yang ada di bawah perut. Hasil tersebut belum ada referensi yang membuktikan hubungan antara gen ALDH2 dengan gangguan pada liver, kelenjar susu, jaringan adiposa yang ada dibawah kulit, jaringan adiposa yang ada dibawah kulit perut. Namun berdasarkan penelitian yang dilakukan (Irham *et al.*, 2023) menyatakan bahwa gen ALDH2 memiliki efek yang meningkatkan resiko terjadinya hipertensi.

Defisiensi enzim ALDH2 telah diidentifikasi pada tikus hipertensi dapat menyebabkan stroke. ALDH2 berperan dalam cedera saraf karena ekspresi berlebih atau aktivasi ALDH2. Oleh karena itu, aktivasi jalur ALDH2 dapat digunakan sebagai indeks dalam mengidentifikasi subjek yang rentan terhadap stroke, dan jalur ALDH2 dapat menjadi target intervensi terapeutik yang potensial (Guo *et al.*, 2013).

- against stroke by clearing 4-HNE. *Cell Research*, 23(7), 915–930. <https://doi.org/10.1038/cr.2013.69>
- GWAS. (n.d.). *GWAS Catalog*. National Human Genome Research Institute. Retrieved October 15, 2023, from <https://www.ebi.ac.uk/gwas/>
- HaploReg, V. 2. (n.d.). *Single-Nucleotide Polymorphisms (SNP)*. Broad Institute. Retrieved October 15, 2023, from <https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php>
- Institute, E. B. (n.d.). *Ensembl Genom*. Retrieved December 19, 2023, from <https://asia.ensembl.org/index.html>
- Irham, L. M., Adikusuma, W., Lolita, L., Puspitaningrum, A. N., Afief, A. R., Sarasmita, M. A., Dania, H., Khairi, S., Djalilah, G. N., Purwanto, B. D., & Chong, R. (2023). Investigation of susceptibility genes for chickenpox disease across multiple continents. *Biochemistry and Biophysics Reports*, 33(December 2022), 101419. <https://doi.org/10.1016/j.bbrep.2022.101419>
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Research*, 49(D1), D1388–D1395. <https://doi.org/10.1093/nar/gkaa971>
- Li, C., Yang, Y., Liu, X., Li, Z., Liu, H., & Tan, Q. (2020). Glucose metabolism-related gene polymorphisms as the risk predictors of type 2 diabetes. *Diabetology and Metabolic Syndrome*, 12(1), 4–9. <https://doi.org/10.1186/s13098-020-00604-5>
- Lifia, A., Kartikasari, N., Devi, A., Wibowo, K., & Budiawan, H. (2023). Identifikasi Variasi Gen dan Ekspresi Gen Yang Berhubungan Dengan Anemia Aplastik Menggunakan Pendekatan Genomik Dan Bioinformatika. *Lambung Farmasi: Jurnal Ilmu Kefarmasian*, 4(2), 300–306.
- NCBI. (2018). *GCKR glucokinase regulator Homo sapiens (human)*. National Center for Biotechnology Information. <https://www.ncbi.nlm.nih.gov/sire.ub.edu/gene/2646>
- Pal, L. R., & Moulton, J. (2015). Genetic Basis of Common Human Disease: Insight into the Role of Missense SNPs from Genome-Wide Association Studies. *Journal of Molecular Biology*, 427(13), 2271–2289. <https://doi.org/doi.org/10.1016/j.jmb.2015.04.014>
- Portal, G. (n.d.). *GTX Portal*. Broad Institute of MIT and Harvard. Retrieved December 16, 2023, from <https://www.gtexportal.org/home/>
- Puspitaningrum, A. N., Perwitasari, D. A., Adikusuma, W., Djalilah, G. N., Dania, H., Maliza, R., Faridah, I. N., Sarasmita, M. A., Rezadhini, M., Cheung, R., & Irham, L. M. (2022). Integration of genomic databases and bioinformatic approach to identify genomic variants for sjogren's syndrome on multiple continents. *Media Farmasi: Jurnal Ilmu Farmasi*, 19(2), 71.

<https://doi.org/10.12928/mf.v19i2.23706>

- Shallis, R. M., Ahmad, R., & Zeidan, A. M. (2018). Aplastic anemia: Etiology, molecular pathogenesis, and emerging concepts. *European Journal of Haematology*, *101*(6), 711–720. <https://doi.org/10.1111/ejh.13153>
- Stanzione, R., Cotugno, M., Bianchi, F., Marchitti, S., Forte, M., Volpe, M., & Rubattu, S. (2020). Pathogenesis of ischemic stroke: Role of epigenetic mechanisms. *Genes*, *11*(1). <https://doi.org/10.3390/genes11010089>
- Zhang, K., Loong, S. S. E., Yuen, L. Z. H., Venketasubramanian, N., Chin, H. L., Lai, P. S., & Tan, B. Y. Q. (2023). Genetics in Ischemic Stroke: Current Perspectives and Future Directions. *Journal of Cardiovascular Development and Disease*, *10*(12). <https://doi.org/10.3390/jcdd10120495>
- Zhou, Y. J., Hong, S. C., Yin, R. X., Yang, Q., Cao, X. L., & Chen, W. X. (2015). Polymorphisms in the GCKR are associated with serum lipid traits, the risk of coronary artery disease and ischemic stroke. *International Journal of Clinical and Experimental Medicine*, *8*(7), 10678–10686.