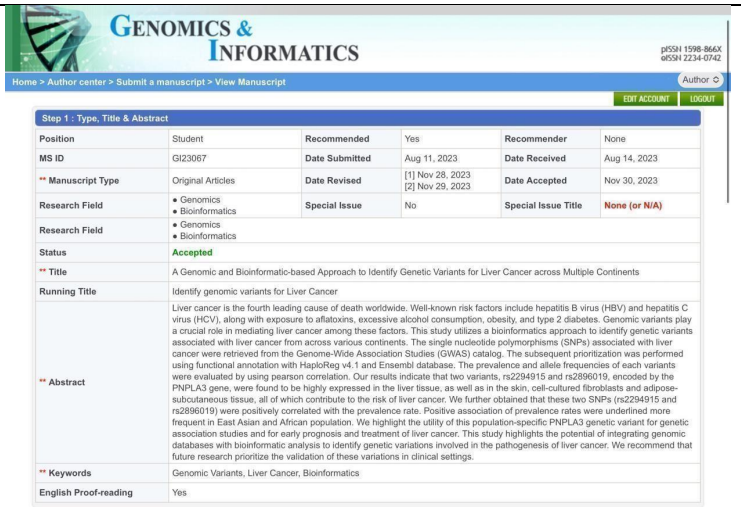
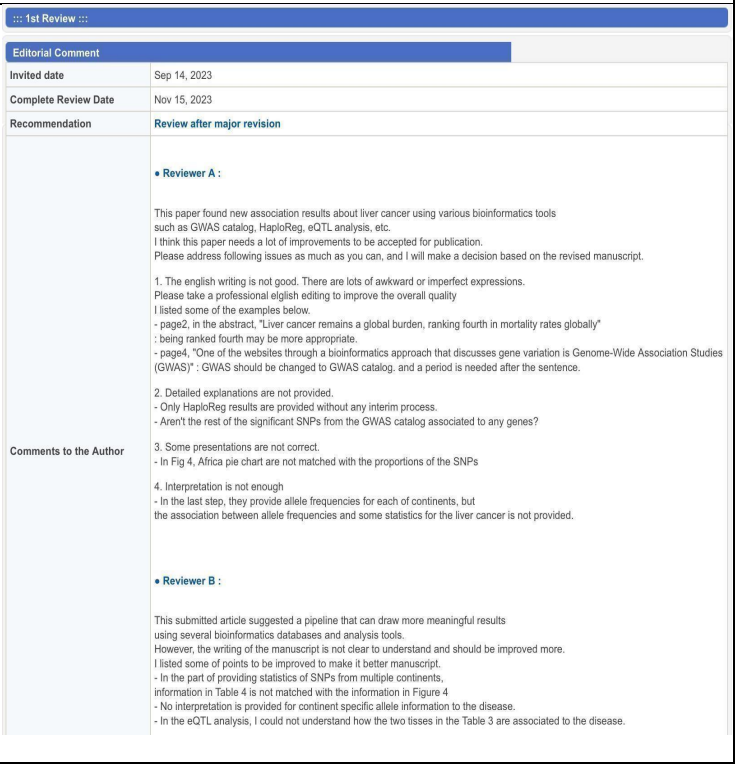
“A genomic and bioinformatic-based approach to identify genetic variants for liver cancer across multiple continents”

Genomics & Informatics Journal, Vol. 21, No. 4, 2023, pp.1-8

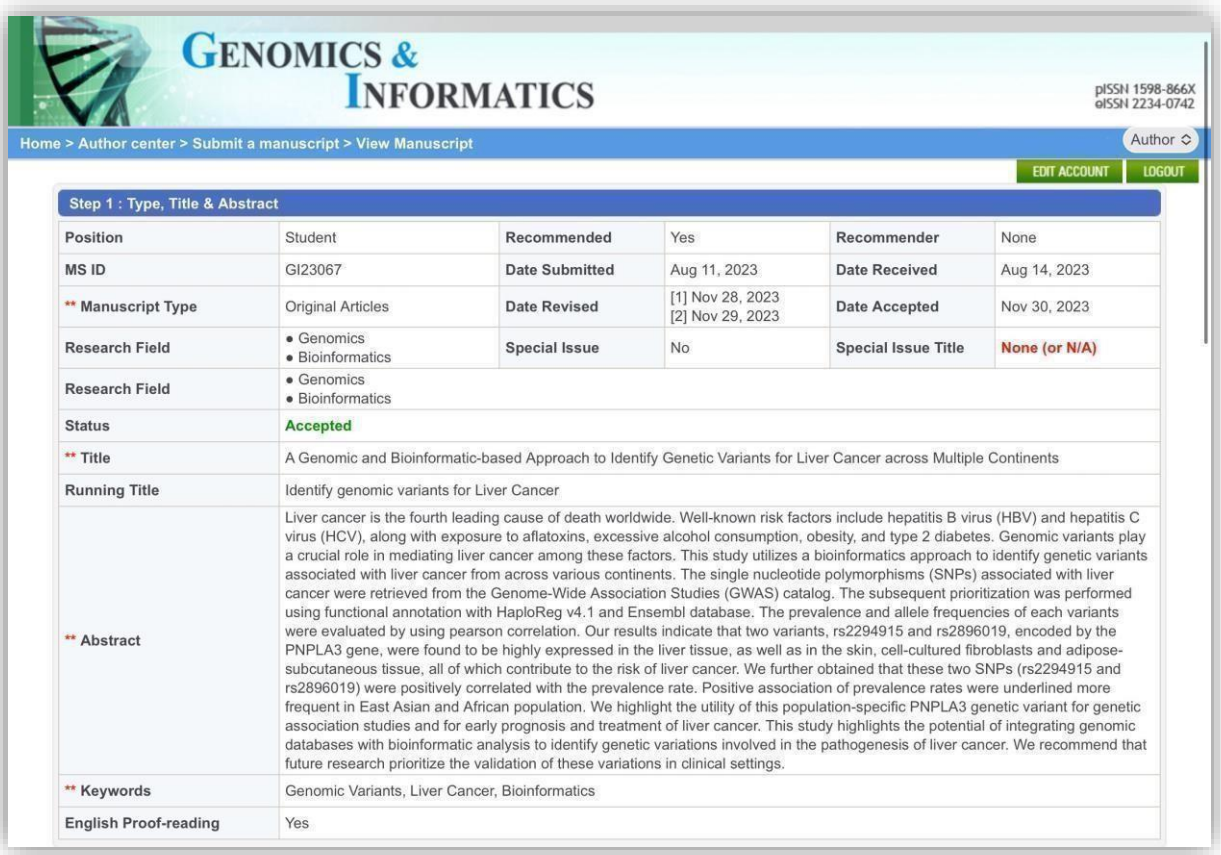
Apt. Lalu Muhammad Irham, M.Farm., Ph.D

No	Keterangan	Tanggal	Bukti Proses
1	Submit Artikel	11 Agustus 2023	 GENOMICS & INFORMATICS Home > Author center > Submit a manuscript > View Manuscript Step 1 : Type, Title & Abstract Position: Student; Recommended: Yes; Recommender: None MS ID: GI23067; Date Submitted: Aug 11, 2023; Date Received: Aug 14, 2023 ** Manuscript Type: Original Articles; Date Revised: [1] Nov 28, 2023 [2] Nov 29, 2023; Date Accepted: Nov 30, 2023 Research Field: • Genomics • Bioinformatics; Special Issue: No; Special Issue Title: None (or N/A) Research Field: • Genomics • Bioinformatics Status: Accepted ** Title: A Genomic and Bioinformatic-based Approach to Identify Genetic Variants for Liver Cancer across Multiple Continents Running Title: Identify genomic variants for Liver Cancer ** Abstract: Liver cancer is the fourth leading cause of death worldwide. Well-known risk factors include hepatitis B virus (HBV) and hepatitis C virus (HCV), along with exposure to aflatoxins, excessive alcohol consumption, obesity, and type 2 diabetes. Genomic variants play a crucial role in mediating liver cancer among these factors. This study utilizes a bioinformatics approach to identify genetic variants associated with liver cancer from across various continents. The single nucleotide polymorphisms (SNPs) associated with liver cancer were retrieved from the Genome-Wide Association Studies (GWAS) catalog. The subsequent prioritization was performed using functional annotation with HaploReg v4.1 and Ensembl database. The prevalence and allele frequencies of each variants were evaluated by using pearson correlation. Our results indicate that two variants, rs2294915 and rs2896019, encoded by the PNPLA3 gene, were found to be highly expressed in the liver tissue, as well as in the skin, cell-cultured fibroblasts and adipose-subcutaneous tissue, all of which contribute to the risk of liver cancer. We further obtained that these two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate. Positive association of prevalence rates were underlined more frequent in East Asian and African population. We highlight the utility of this population-specific PNPLA3 genetic variant for genetic association studies and for early prognosis and treatment of liver cancer. This study highlights the potential of integrating genomic databases with bioinformatic analysis to identify genetic variations involved in the pathogenesis of liver cancer. We recommend that future research prioritize the validation of these variations in clinical settings. ** Keywords: Genomic Variants, Liver Cancer, Bioinformatics English Proof-reading: Yes
2	Artikel mendapatkan hasil review dari 2 reviewer	15 November 2023	 ::: 1st Review ::: Editorial Comment Invited date: Sep 14, 2023 Complete Review Date: Nov 15, 2023 Recommendation: Review after major revision Comments to the Author: • Reviewer A : This paper found new association results about liver cancer using various bioinformatics tools such as GWAS catalog, HaploReg, eQTL analysis, etc. I think this paper needs a lot of improvements to be accepted for publication. Please address following issues as much as you can, and I will make a decision based on the revised manuscript. 1. The english writing is not good. There are lots of awkward or imperfect expressions. Please take a professional elglish editing to improve the overall quality I listed some of the examples below. - page2, in the abstract, "Liver cancer remains a global burden, ranking fourth in mortality rates globally" : being ranked fourth may be more appropriate. - page4, "One of the websites through a bioinformatics approach that discusses gene variation is Genome-Wide Association Studies (GWAS)" : GWAS should be changed to GWAS catalog. and a period is needed after the sentence. 2. Detailed explanations are not provided. - Only HaploReg results are provided without any interim process. - Aren't the rest of the significant SNPs from the GWAS catalog associated to any genes? 3. Some presentations are not correct. - In Fig 4, Africa pie chart are not matched with the proportions of the SNPs 4. Interpretation is not enough - In the last step, they provide allele frequencies for each of continents, but the association between allele frequencies and some statistics for the liver cancer is not provided. • Reviewer B : This submitted article suggested a pipeline that can draw more meaningful results using several bioinformatics databases and analysis tools. However, the writing of the manuscript is not clear to understand and should be improved more. I listed some of points to be improved to make it better manuscript. - In the part of providing statistics of SNPs from multiple continents, information in Table 4 is not matched with the information in Figure 4 - No interpretation is provided for continent specific allele information to the disease. - In the eQTL analysis, I could not understand how the two tisses in the Table 3 are associated to the disease.

History Artikel
"A genomic and bioinformatic-based approach to identify genetic variants for liver cancer across multiple continents"

Dalam penerbitan artikel ini dapat diakses melalui https://genominfo.org dengan informasi metadata artikel pada jurnal sebagai berikut.

Artikel disubmit pada tanggal 11 Agustus 2023. Selanjutnya, pada tanggal 14 Agustus 2023 artikel yang telah disubmit diterima oleh pihak Genomics & Informatics Journal.



GENOMICS & INFORMATICS

pISSN 1598-866X
eISSN 2234-0742

Home > Author center > Submit a manuscript > View Manuscript

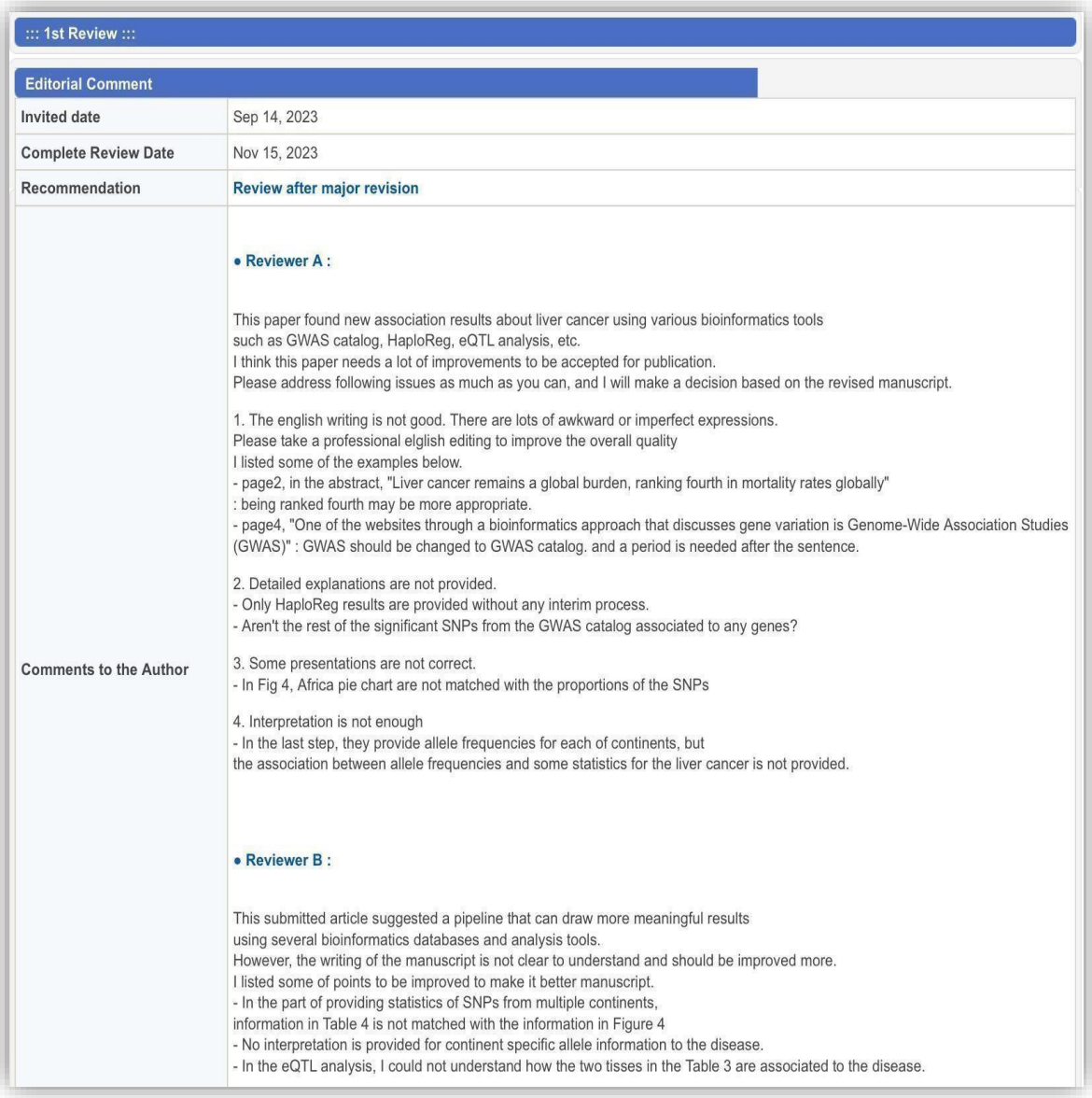
Author

EDIT ACCOUNT LOGOUT

Step 1 : Type, Title & Abstract

Position	Student	**Recommended**	Yes	**Recommender**	None
MS ID	GI23067	**Date Submitted**	Aug 11, 2023	**Date Received**	Aug 14, 2023
** **Manuscript Type**	Original Articles	**Date Revised**	[1] Nov 28, 2023 [2] Nov 29, 2023	**Date Accepted**	Nov 30, 2023
Research Field	• Genomics • Bioinformatics	**Special Issue**	No	**Special Issue Title**	None (or N/A)
Research Field	• Genomics • Bioinformatics				
Status	Accepted				
** **Title**	A Genomic and Bioinformatic-based Approach to Identify Genetic Variants for Liver Cancer across Multiple Continents				
Running Title	Identify genomic variants for Liver Cancer				
** **Abstract**	Liver cancer is the fourth leading cause of death worldwide. Well-known risk factors include hepatitis B virus (HBV) and hepatitis C virus (HCV), along with exposure to aflatoxins, excessive alcohol consumption, obesity, and type 2 diabetes. Genomic variants play a crucial role in mediating liver cancer among these factors. This study utilizes a bioinformatics approach to identify genetic variants associated with liver cancer from across various continents. The single nucleotide polymorphisms (SNPs) associated with liver cancer were retrieved from the Genome-Wide Association Studies (GWAS) catalog. The subsequent prioritization was performed using functional annotation with HaploReg v4.1 and Ensembl database. The prevalence and allele frequencies of each variants were evaluated by using pearson correlation. Our results indicate that two variants, rs2294915 and rs2896019, encoded by the PNPLA3 gene, were found to be highly expressed in the liver tissue, as well as in the skin, cell-cultured fibroblasts and adipose-subcutaneous tissue, all of which contribute to the risk of liver cancer. We further obtained that these two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate. Positive association of prevalence rates were underlined more frequent in East Asian and African population. We highlight the utility of this population-specific PNPLA3 genetic variant for genetic association studies and for early prognosis and treatment of liver cancer. This study highlights the potential of integrating genomic databases with bioinformatic analysis to identify genetic variations involved in the pathogenesis of liver cancer. We recommend that future research prioritize the validation of these variations in clinical settings.				
** **Keywords**	Genomic Variants, Liver Cancer, Bioinformatics				
English Proof-reading	Yes				

Pada tanggal 15 November 2023 artikel mendapatkan revisian yang pertama oleh 2 reviewers.

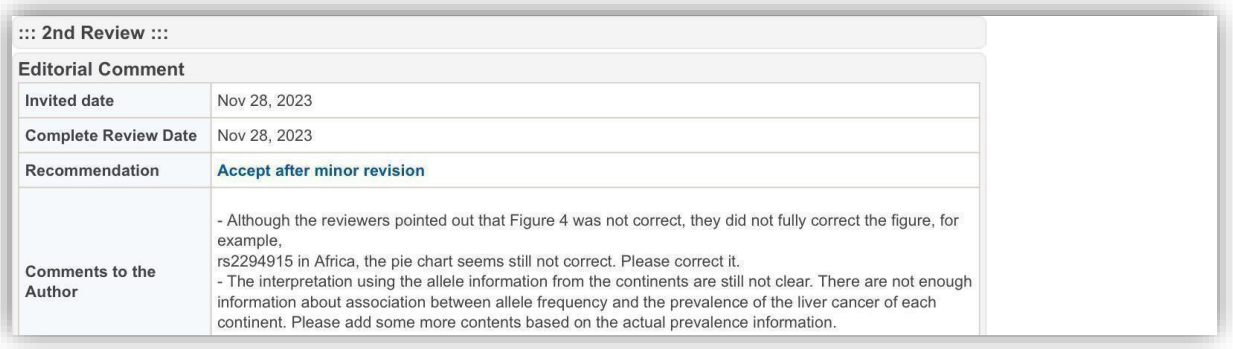


::: 1st Review :::

Editorial Comment

Invited date	Sep 14, 2023
Complete Review Date	Nov 15, 2023
Recommendation	**Review after major revision**
Comments to the Author	**• Reviewer A :** This paper found new association results about liver cancer using various bioinformatics tools such as GWAS catalog, HaploReg, eQTL analysis, etc. I think this paper needs a lot of improvements to be accepted for publication. Please address following issues as much as you can, and I will make a decision based on the revised manuscript. 1. The english writing is not good. There are lots of awkward or imperfect expressions. Please take a professional elglish editing to improve the overall quality I listed some of the examples below. - page2, in the abstract, "Liver cancer remains a global burden, ranking fourth in mortality rates globally" : being ranked fourth may be more appropriate. - page4, "One of the websites through a bioinformatics approach that discusses gene variation is Genome-Wide Association Studies (GWAS)" : GWAS should be changed to GWAS catalog. and a period is needed after the sentence. 2. Detailed explanations are not provided. - Only HaploReg results are provided without any interim process. - Aren't the rest of the significant SNPs from the GWAS catalog associated to any genes? 3. Some presentations are not correct. - In Fig 4, Africa pie chart are not matched with the proportions of the SNPs 4. Interpretation is not enough - In the last step, they provide allele frequencies for each of continents, but the association between allele frequencies and some statistics for the liver cancer is not provided. **• Reviewer B :** This submitted article suggested a pipeline that can draw more meaningful results using several bioinformatics databases and analysis tools. However, the writing of the manuscript is not clear to understand and should be improved more. I listed some of points to be improved to make it better manuscript. - In the part of providing statistics of SNPs from multiple continents, information in Table 4 is not matched with the information in Figure 4 - No interpretation is provided for continent specific allele information to the disease. - In the eQTL analysis, I could not understand how the two tisses in the Table 3 are associated to the disease.

Pada tanggal 28 November 2023 artikel mendapatkan revisian yang kedua oleh reviewers.



::: 2nd Review :::

Editorial Comment

Invited date	Nov 28, 2023
Complete Review Date	Nov 28, 2023
Recommendation	Accept after minor revision
Comments to the Author	- Although the reviewers pointed out that Figure 4 was not correct, they did not fully correct the figure, for example, rs2294915 in Africa, the pie chart seems still not correct. Please correct it. - The interpretation using the allele information from the continents are still not clear. There are not enough information about association between allele frequency and the prevalence of the liver cancer of each continent. Please add some more contents based on the actual prevalence information.

Pada tanggal 29 November 2023 hasil revisian yang kedua telah diterima oleh editor jurnal. Dan pada tanggal 30 November 2023 artikel telah diterima / accepted oleh editor.

Editorial Comment

Invited date	Nov 29, 2023
Complete Review Date	Nov 30, 2023
Recommendation	Accept as it is
Comments to the Author	None (or N/A)
Files	None (or N/A)

Close

Berikut akan dilampirkan beberapa dokumen perubahan dari artikel yang diterbitkan sejak proses submit, hasil review dari reviewer, artikel revisi, dan artikel versi terbit.

Artikel pertama kali saat disubmit
pada tanggal 11 Agustus 2023

An Insight of Genomic Variants and Bioinformatic-based Approach Mediating Liver Cancer across Multiple Continents

Muhammad Ma'ruf[1], Lalu Muhammad Irham[1]*, Wirawan Adikusuma[2], Made Ary Sarasmita[3,4] Sabiah Khairi[5], Barkah Djaka Purwanto[6,7], Rockie Chong[8], Maulida Mazaya[9]

[1]*Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia*

[2]*Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram, Indonesia*

[3]*Department of Clinical Pharmacy, College of Pharmacy, Taipei Medical University, Taipei, Taiwan*

[4]*Pharmacy Study Program, Faculty of Science and Mathematics, Udayana University, Bali, Indonesia*

[5]*School of Nursing, College of Nursing, Taipei Medical University, Taipei 11031, Taiwan*

[6]*Faculty of Medicine, University of Ahmad Dahlan, Yogyakarta 55191, Indonesia*

[7]*PKU Muhammadiyah Bantul Hospital, Bantul, Yogyakarta 55711, Indonesia*

[8]*Department of Chemistry and Biochemistry, University of California, Los Angeles, USA*

[9]*Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong Science Center, Cibinong, Indonesia*

*__Correspondence:__ Lalu Muhammad Irham; email: lalu.irham@pharm.uad.ac.id

Abstract

Liver cancer remains a global burden, ranking fourth in mortality rates globally. Risk factors such as hepatitis B virus (HBV) and hepatitis C virus (HBC) have been widely reported. Additionally, exposure to contamination with aflatoxin, alcohol consumption, obesity, type 2 diabetes, and genomic variations have been investigated as potential risk factors. Genomic variants play a crucial role in mediating liver cancer among these factors. However, specific variants involved in this process are still limitedly studied. This study uses a bioinformatics approach to identify genetic variants associated with liver cancer from across various continents. The single nucleotide polymorphisms (SNPs) associated with liver cancer were retrieved from the Genome-Wide Association Studies (GWAS) catalog. The subsequent prioritization was performed using functional annotation with HaploReg v4.1 and the Ensembl database. Our results indicate that two variants, rs2294915 and rs2896019, encoded by the PNPLA3 gene, were found to be highly expressed in the liver tissue, as well as in the skin, cell-cultured fibroblasts and adipose-subcutaneous tissue, all of which contribute to the risk of liver cancer. We highlight the utility of this population-specific *PNPLA3* genetic variant for genetic association studies and for early prognosis and treatment of liver cancer. This study emphasizes that integrating genomic databases and bioinformatic analysis is a promising approach to identify genetic variations that play a role in the pathogenesis of liver cancer. We suggest that future researchers focus on these gene variations to be validated in clinical studies.

Keywords: *Genomic Variants, Liver Cancer, Bioinformatics*

Introduction

Liver cancer is a type of carcinoma that has the highest mortality rate in the world every year (McGlynn et al., 2021). There were 841,000 cases of liver cancer in 2018, of which the death rate caused by liver cancer reached 782,000 (Bray et al., 2018). Cases of liver cancer are an average number of cases and deaths that can increase 2 to 3 times in men in parts of the world. According to the Global Cancer Statistics (GLOBOCAN), in 2020, liver cancer occupies the third position (8.3%) as a deadly disease due to cancer. Liver cancer had an incidence of 905 thousand cases in 2020 and a mortality rate of 830 thousand (Sung et al., 2021). In Indonesia, liver cancer is the second most common in men, amounting to 12.4 per 100,000 of the Indonesian population, with an average death rate of 7.6 per 100,000 (Kemenkes RI, 2019).

Factors that cause liver cancer include chronic infection with hepatitis B virus (HBV), hepatitis C virus (HBC), the result of contamination with aflatoxin, alcohol consumption, history of obesity, history of type 2 diabetes, and smoking addiction (Bray et al., 2018). According to Villanueva (2019), other risk factors are thought to exacerbate the occurrence of liver cancer, such as an unhealthy lifestyle, geographic conditions, gender, age, family history of the disease, and the severity of damage to the liver. Liver cancer is also found in areas that have cases of hepatitis B. In these areas, liver cancer is prevalent at a young age. This is because some infected with hepatitis B are obtained vertically through the delivery process (Mittal & El-Serag, 2013).

Patients often felt complaints in the form of fatigue, pain, diarrhea, skin abnormalities, and decreased appetite, all of which have affected their quality of life (Waller et al., 2015). Therefore, detecting the presence of disease symptoms in liver cancer can be done by examining deoxyribonucleic acid (DNA). Gene variation can be associated with disease progression and

pathogenesis, which includes liver cancer. One of the websites through a bioinformatics approach that discusses gene variation is Genome-Wide Association Studies (GWAS) GWAS is a database with single nucleotide polymorphism (SNP) search results that has identified several variants associated with liver fat content, circulating liver enzymes, and the development of NAFLD as well as genetic markers used in predicting a disease disorder (Wang et al., 2021).

Genetic identification in humans aims to identify inherited genetic risk factors for liver cancer. This study uses the GWAS database to map genes from genetic variations across several populations that play an essential role in the pathogenesis of liver cancer. The most significant gene variations based on their function in protein changes will be further verified.

Methods

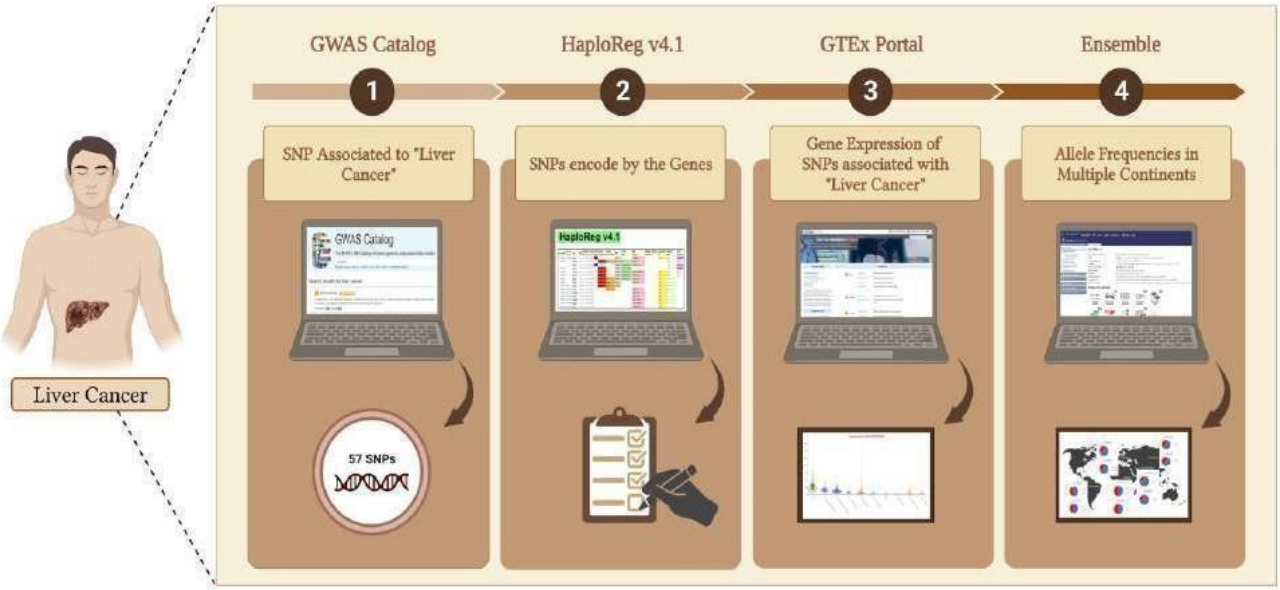


Figure 1. Analysis methodology for integrated bioinformatic, database and genomic analysis of genetic variation that affect liver cancer. The figure was created with BioRender.com under agreement number "FM25OO073C"

In this study, we adopted the method used by Ma'ruf et al (2023) to identify genomic variants associated with Stevens-Johnson syndrome (SJS). The implementation of the methodology is illustrated in Figure 1. Liver cancer-associated SNPs were obtained from the GWAS Catalog of the National Human Genome Research Institute (NHGRI) GWAS Catalog database (http://www.ebi.ac.uk/gwas) (accessed 15-02-2023). Subsequently, we performed further analysis using HaploReg (version 4.1). The p-value < 10^{-8} was applied to account for multiple tests in the GWAS Catalog, as this threshold is commonly used to identify associations between common genetic variants and traits with adjacent gene expression (Chen et al., 2021). Furthermore, to evaluate the relationships between various genetic variants and gene expression profiles, we utilized e-QTL analysis with data from the GTEx Portal database (http://www.gtexportal.org/home/) (accessed on 16-02-2023), considering gene expression across

various tissues in humans. Additionally, we confirmed the identified variants using the Ensembl Genome Browser (https://www.ensembl.org/index.html) (accessed on 17-02-2023). For this study, we considered allele frequencies in populations from Europe, Africa, America, East Asia, and Southeast Asia. Then, to understand the functions of the various gene variants, we performed evaluations using the SNP nexus database (https://www.snp-nexus.org) (accessed on 20-02-2023).

Results and Discussion

1. Identification of Genomic Variants of Liver Cancer

This study identified SNPs associated with liver cancer from the GWAS database. Among them, 29 SNPs were further confirmed through SNP duplication, as shown in Table 1. Subsequently, HaploReg version 4.1 was utilized, and a p-value $<10^{-8}$ was applied based on the number of SNPs obtained. Based on the findings presented in Table 2, we found the risk of two genes for "Liver Cancer" disease. This study analyzed tissue expression affecting liver cancer with the missense variant *PNPLA3*.

Through our integrative bioinformatics approach, two variants with a missense mutation (rs rs2294915, rs2896019) that encoded the *PNPLA3* genes were prioritized as the biological risk SNPs for Liver Cancer. Primary liver cancer, also known as hepatocellular carcinoma, is a pathological condition characterised by the development of malignant cells within the hepatic tissues. The development of cancer in extraneous anatomical sites that then metastasizes to the liver does not constitute primary liver cancer. Primary liver cancer encompasses many kinds, including hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (iCCA), and less frequent varieties such as mixed hepatocellular cholangiocarcinoma (HCC-CCA), fibrolamellar HCC (FLC), and the paediatric neoplasm hepatoblastoma (Wage et al., 2021).

Table 1. SNPs from the GWAS catalog with p-value $<10^{-8}$

No.	Variation and risk allele	p-value
1	rs2856723	$3x10^{-43}$
2	rs34675408	$1x10^{-32}$
3	rs9272105	$5x10^{-22}$
4	rs913493	$5x10^{-20}$
5	rs2294915	$2x10^{-19}$
6	rs17401966	$2x10^{-18}$
7	rs3096380	$1x10^{-17}$
8	rs9275319	$3x10^{-17}$
9	rs584368	$2x10^{-14}$
10	rs2596542	$4x10^{-13}$
11	rs1110446	$9x10^{-13}$
12	rs58489806	$3x10^{-12}$
13	rs6078460	$2x10^{-11}$
14	rs2523961	$6x10^{-11}$
15	rs7574865	$2x10^{-10}$
16	rs1110446	$3x10^{-10}$
17	rs455804	$5x10^{-10}$
18	rs58542926	$6x10^{-10}$
19	rs2523961	6x10-^{10}
20	rs8107030	$8x10^{-10}$
21	rs10272859	$9x10^{-10}$
22	rs190121281	$4x10^{-9}$
23	rs9275572	$6x10^{-9}$
24	rs2242652	$6x10^{-9}$

25	rs188273166	$1x10^{-8}$
26	rs708113	$1x10^{-8}$
27	rs2896019	$2x10^{-8}$
28	rs17047200	$3x10^{-8}$
29	rs541860626	$5x10^{-8}$

Table 2. Variants and risk alleles of liver cancer encoding prioritized SNPs

Variationandrisk alleles	Variantsnearrisk allele ($r^2 > 0.8$)	*p*-value	Gencode	Type of allele
rs2294915	rs738409	$2x10^{-19}$	*PNPLA3*	missense
rs2896019	rs3761472	$2x10^{-8}$	*PNPLA3*	missense

2. Gene expression of *PNPLA3* in 10 human tissues

The results of PNPLA3 gene expression in 10 human tissues comprise the most apparent functional consequences of genetic variation. Liver, sun-exposed skin (lower legs), non-sun-exposed skin (suprapubic), and adipose subcutaneous fibroblasts and cell cultures showed the highest *PNPLA3* gene expression in the 10 human tissues analyzed from GTEx (Figure 2). In addition, we have found that the SNP IDs rs2294915 and rs2896019 have similar gene expression variations in Sun-Exposed skin (lower legs).

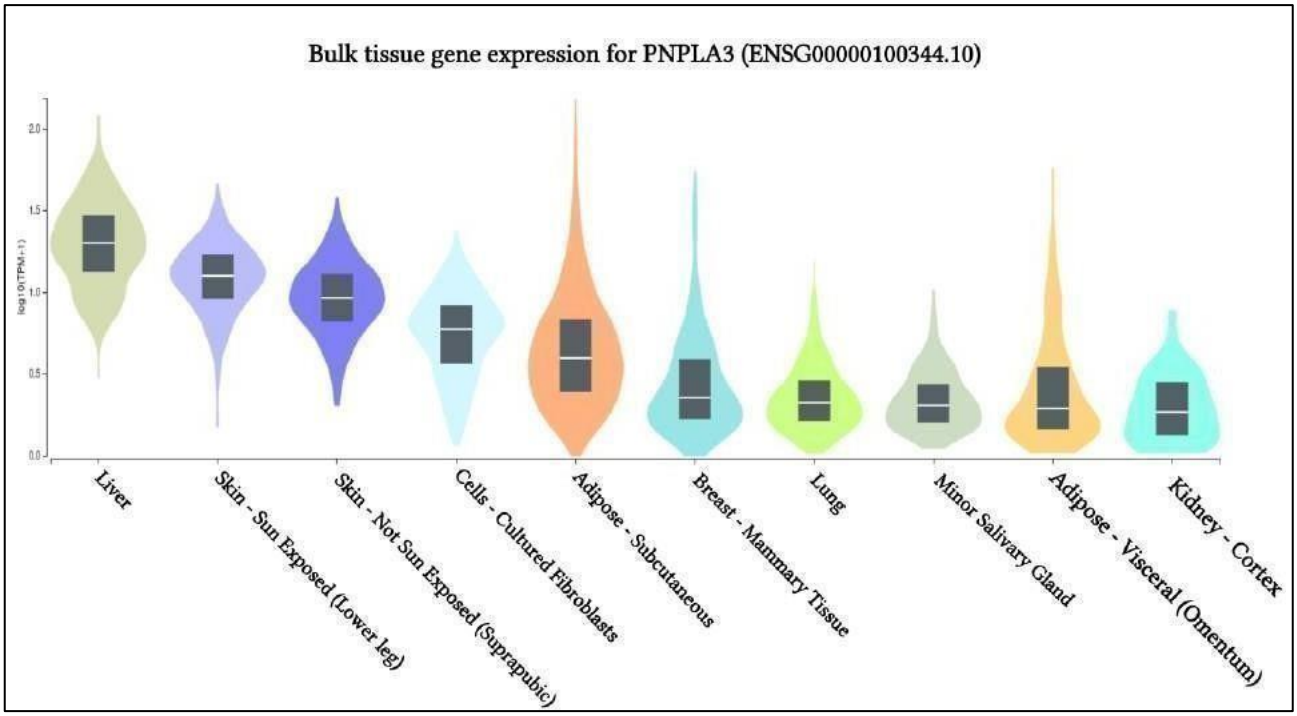


Figure 2. *PNPLA3* gene expression associated with liver cancerin several human tissues based on GTEx Portal analysis

3. Correlation between Gene Expression of *PNPLA3* and eQTL

The result in a correlation between the Gene Expression of *PNPLA3* and eQTL, we identified an allele of rs2294915 and rs2896019 in *PNPLA3* directly related to liver cancer. As shown in Table 3 and Figure 3, the CC genotypes rs2294915, and rs2896019 were associated with higher

expression of *PNPLA3* in sun-exposed (lower leg) and non-sun-exposed (suprapubic) skin tissues compared to the genotypes TT.

Table 3. Results of e-QTL in liver cancer from the GTEx portal database

SNP	Gencode ID (ENSG00000-)	Gene Symbol	*p*-value	Effect Size	Tissue	Expression Level
rs2294915	100344.10	*PNPLA3*	2.8×10^{-8}	-0.15	Skin - Sun Exposed (Lower leg)	CC>CT>TT
	100344.10	*PNPLA3*	5×10^{-8}	-0.50	Skin - Not Sun Exposed (Suprapubic)	CC>CT>TT
rs2896019	100344.10	*PNPLA3*	6.7×10^{-11}	-0.19	Skin - Sun Exposed (Lower leg)	CC>CT>TT
	100344.10	*PNPLA3*	2×10^{-9}	-0.22	Skin - Not Sun Exposed (Suprapubic)	CC>CT>TT

Source: Expression Quantitative Trait Loci (eQTL) obtained from the GTEx Portal.
Note : *PNPLA3*, patatin like phospoliphase 3; SNP, single nucleotide polymorphism; eQTL, expression quantitative trait loci; SNP, single nucleotide polymorphism.

This study found that gene expression in *PNPLA3* at rs2294915 and rs2896019 has a link with "liver cancer." Furthermore, rs2294915 and rs2896019 encoded missense mutations and the CC genotype had the highest expression of *PNPLA3* in sun-exposed (lower leg) and non-sun-exposed (suprapubic) skin tissue, and the TT genotype showed the most melancholy expression (Fig 3). According to Trepo et al (2022) at rs2294915, the *PNPLA3* locus on chromosome 22q13.31 has also been linked to alcohol-related liver cancer (p=3.71x10-7), and at rs2896019, it has a single nucleotide polymorphism relationship with liver steatosis in obese pediatric patients in children and adolescents (Stasinou et al., 2022).

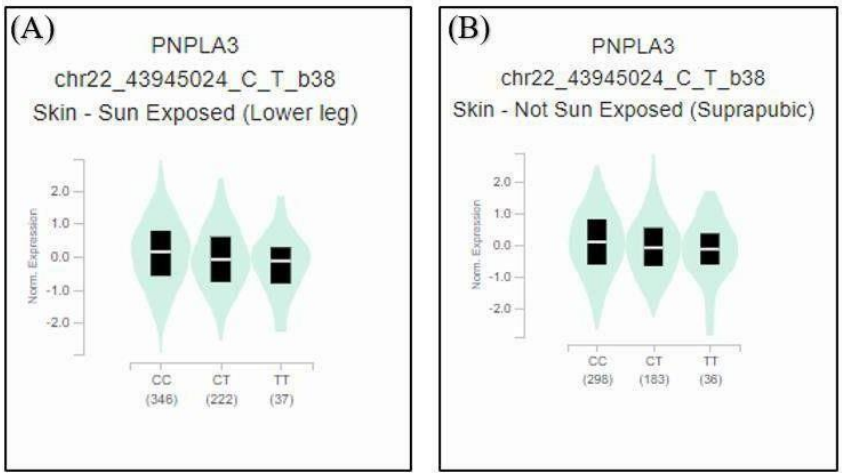


Figure 3. Patatin like phospoliphase 3 (*PNPLA3*) gene expression for each genotype of the single nucleotide polymorphism (SNP): (A) rs2294915 and (B) rs2896019.

4. Allele frequencies of candidate variants in populations in different continents

The results of the research we have done, we have identified variants associated with liver cancer gene expression and carried out allele frequency analysis in various populations. As shown in Table 4, allele variant frequencies were evaluated in multiple people from Europe, America, East Asia, South Asia, and Africa. Allele frequencies across populations varied for each SNP, as depicted in Figure 4.

Table 4. Analysis of allele frequencies for the *PNPLA3* gene from variant annotation (SNPnexus)

SNP	Gene	Location	Allele		Ilele Frequency (N				
			Ref	Eff*	AFR	AMR	EAS	EUR	SAS
rs2294915	*PNPLA3*	Missense	C	T	T: 0.163 (215)	T: 0.490 (340)	T: 0.365 (368)	T: 0.252 (254)	T: 0.246 (241)
rs2896019	*PNPLA3*	Missense	C	T	T: 0.842 (1113)	T: 0.562 (390)	T: 0.636 (641)	T: 0.801 (806)	T: 0.764 (747)

Note : *PNPLA3*, patatin like phospoliphase 3; SNP, single nucleotide polymorphism; Ref, Reference; Alt, slternate; AFR, Africa; AMR, America; EAS, East Asia; EUR, Europe; SAS, Southeast Asia.

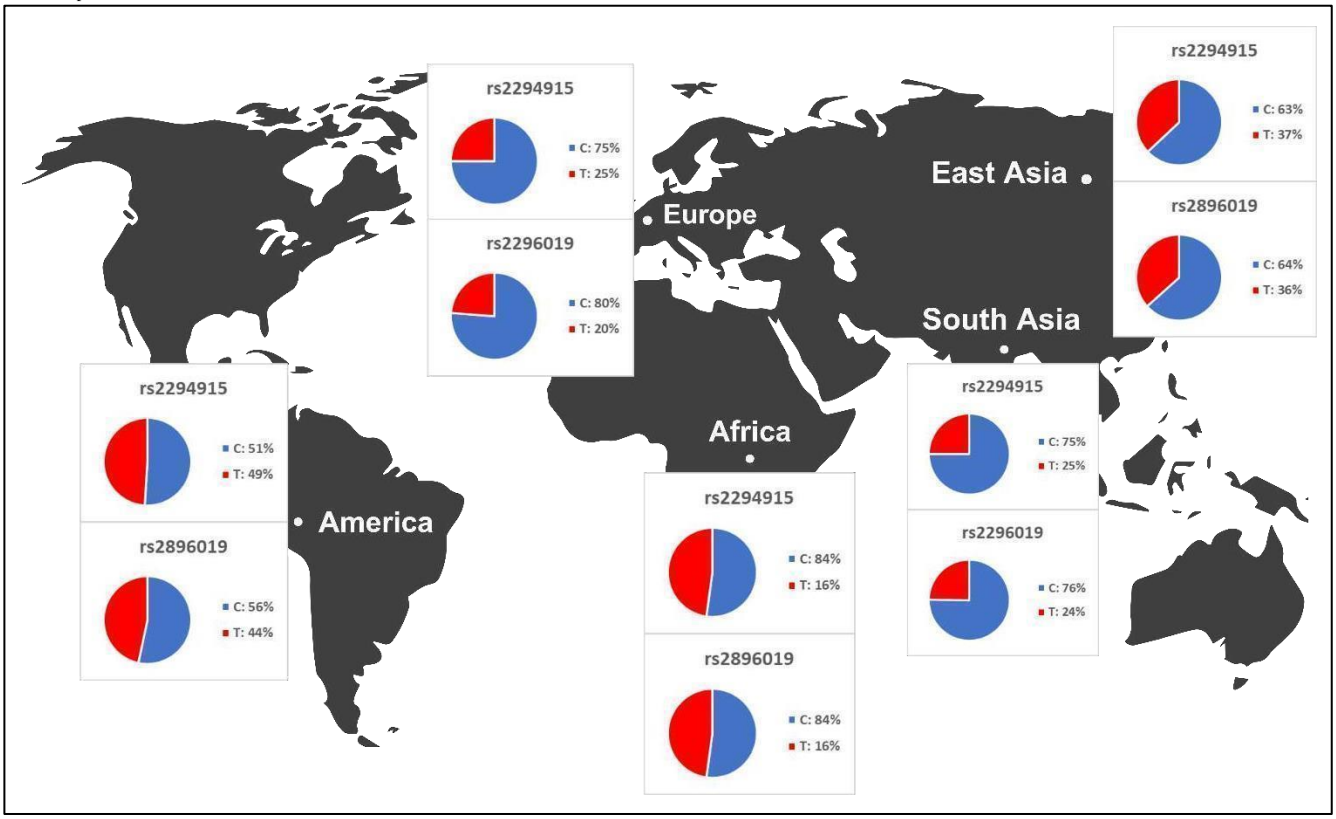


Fig 4. The results of the distribution of allele frequencies affecting *PNPLA3* across various populations. EUR, Europe; AFR, Africa; SAS, South Asia; EAS, East Asia.

Based on this finding, rs2294915 and rs2896019 are potentially related to the susceptibility to "liver cancer with effect size the highest score of -0.50 can be interpreted on the skin not exposed to sunlight (suprapubic)." According to Poggiali & Vercelli (2023), this condition is characterized by a disruption in the heme biosynthesis pathway due to reduced hepatic uroporphyrinogen decarboxylase (UROD) activity. The consequence of this phenomenon is the buildup of light-sensitive by-products, such as uroporphyrinogen, leading to the development of fragility and blistering of sun-exposed skin, as well as impairment of liver function.

Across human populations, the frequency of the T allele at rs2294915 was associated with a high expression of *PNPLA3* in liver cancer, which is much lower in African populations (16%) compared to South Asians (25%), Europeans (25%), East Asians (37%) and America (49%). In contrast, the frequency of the C allele at rs2296019 was considerably higher in African (84%), European (80%), South Asian (76%), East Asian (64%), and American (56%) populations.

In patients with liver cancer who have a history of alcohol addiction to an amount of ≥ 3 drinks per day can increase the risk of liver cancer by 16% in the general population; diabetics and people

with central obesity also increase the risk of liver cancer by 2 times (McGlynn et al., 2021). The diagnosis of liver cancer in patients often involves using serological testing in conjunction with imaging techniques, which is considered the established approach for identifying liver carcinoma. Nevertheless, the diagnostic sensitivity of the often-used serological test, specifically designed to detect alpha-fetoprotein (AFP), is at around 60%. Imaging modalities, including magnetic resonance imaging (MRI), computed tomography (CT), and ultrasonography (US), exhibit notable levels of sensitivity and specificity in the identification of liver cancer, particularly in individuals afflicted with liver cirrhosis (Huang et al., 2022).

Variant alleles (rs2294915 and rs2896019) are associated with liver cancer. Accordingly, populations from the continents of Africa, America, East Asia, Europe, and South Asia show associated *PNPLA3* expression, resulting in a higher susceptibility to liver cancer. Identifying unique and pathogenic gene variations for a disease is very interesting for research and clinical validation. Identification of these variants can not only provide clues to disease susceptibility or as a diagnostic and prognostic biomarker. (Irham et al., 2020) and but can also be used to find drug target candidates or known as drug repurposing (genomic-driven drug repurposing) (Afief et al., 2022). We hope that the discovery of candidate gene variations for *PNPLA3* can lead to successful clinical validation, which pavaes the way for this promising diagnostic and prognostic biomarker for liver cancer.

It is important to consider that this study's gene variations found to be pathogenic are still preliminary studies using genomic and bioinformatics databases. However, these results are also important information for future researchers who wish to validate these gene variations in liver cancer patients. Future research is strongly recommended to follow up with additional functional annotations to further prioritize pathogenic gene variations.

Conclusion

This study identifies genetic variants influencing "liver cancer" reveals the significance of the *PNPLA3* gene in liver tissue, as well as in skin regions exposed to the sun (lower legs), skin regions not exposed to the sun (suprapubic), cultured fibroblasts, and adipose-subcutaneous tissue, all of which contribute to an increased risk of liver cancer development. The two variants, rs2294915, and rs2896019, displayed varying allele frequencies across populations from the continents of Africa, America, East Asia, Europe, and South Asia, affecting *PNPLA3* gene expression. Consequently, these populations are more susceptible to liver cancer due to the associated *PNPLA3* expression. These findings underscore the importance of understanding and considering genomic variations in precision medicine and screening strategies for liver cancer in diverse populations across continents.

Reference

Afief, A. R., Irham, L. M., Adikusuma, W., Perwitasari, D. A., Brahmadhi, A., & Cheung, R. (2022). Integration of genomic variants and bioinformatic-based approach to drive drug repurposing for multiple sclerosis . *Biochemistry and Biophysics Reports , 101337*.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers

in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Chen, Z., Boehnke, M., Wen, X., & Mukherjee, B. (2021). Revisiting the genome-wide significance threshold for common variant GWAS. *G3: Genes, Genomes, Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkaa056

Huang, L., Sun, H., Sun, L., Shi, K., Chen, Y., Ren, X., Ge, Y., Liu, X., Knoll, W., Zhang, Q., & Wang, Y. (2022). Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning. *Nature Communications*, *14*(48), 1–14.

Irham, L. M., Chou, W. H., Calkins, M. J., Adikusuma, W., Hsieh, S. L., & Chang, W. C. (2020). Genetic variants that influence SARS-CoV-2 receptor TMPRSS2 expression among population cohorts from multiple continents. *Biochemical and Biophysical Research Communications*, *529*(2), 263–269. https://doi.org/10.1016/j.bbrc.2020.05.179

Kementerian Kesehatan Republik Indonesia. (2019). *Hari Kanker Sedunia 2019*. https://www.kemkes.go.id/article/view/19020100003/harikanker-sedunia-2019.html [Diakses pada 21 Mei 2023].

Ma'ruf, M., Fadli, J. C., Mahendra, M. R., Irham, L. M., Sulistyani, N., Adikusuma, W., Chong, R., & Septama, A. W. (2023). A bioinformatic approach to identify pathogenic variants for Stevens Johnson syndrome. *Genomics & Informatics*, *21*(2), 1–9.

McGlynn, K. A., Petrick, J. L., & El-Serag, H. B. (2021). Epidemiology of Hepatocellular Carcinoma. In *Hepatology* (Vol. 73, Issue S1, pp. 4–13). John Wiley and Sons Inc. https://doi.org/10.1002/hep.31288

Mittal, S., & El-Serag, H. B. (2013). *Epidemiology of Hepatocellular Carcinoma Consider the Population*. www.jcge.com

Poggiali, E., & Vercelli, A. (2023). Sun, alcohol, and skin lesions. *Emergency Care Journal*, *19*(11294).

Stasinou, E., Argyraki, M., Sotiriadou, F., Lambropoulos, A., & Fotoulaki, M. (2022). Association between rs738409 and rs2896019 single-nucleotide polymorphisms of phospholipase domain-containing protein 3 and susceptibility to nonalcoholic fatty liver disease in Greek children and adolescents. *Annals of Gastroenterology*, *35*(3), 297–306. https://doi.org/10.20524/aog.2022.0706

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Villanueva, A. (2019). Hepatocellular Carcinoma. *New England Journal of Medicine*, *380*(15), 1450–1462. https://doi.org/10.1056/NEJMra1713263

Wage, H., Schulze, K., Felden, J. V., Calderaro, J., & Reig, M. (2021). Rare variants of primary liver cancer: Fibrolamellar, combined, and sarcomatoid hepatocellular carcinomas. *J. Med Genet*, *64*(11).

Waller, L. P., Deshpande, V., & Pyrsopoulos, N. (2015). Hepatocellular carcinoma: A comprehensive review. In *World Journal of Hepatology* (Vol. 7, Issue 26, pp. 2648–2663). Baishideng Publishing Group Co. https://doi.org/10.4254/wjh.v7.i26.2648

Wang, J., Conti, D. V, Bogumil, D., Sheng, X., Noureddin, M., Wilkens, L. R., Le Marchand, L., Rosen, H. R., Haiman, C. A., & Setiawan, V. W. (2021). Association of Genetic Risk Score With NAFLD in An Ethnically Diverse Cohort. *Hepatology CommuniCations*, *5*(10), 1689–1703. https://doi.org/10.1002/hep4.1751/suppinfo

Hasil review yang pertama dari 2 reviewers pada tanggal 15 November 2023

::: 1st Review :::

Editorial Comment

Invited date	Sep 14, 2023
Complete Review Date	Nov 15, 2023
Recommendation	**Review after major revision**
Comments to the Author	**• Reviewer A :** This paper found new association results about liver cancer using various bioinformatics tools such as GWAS catalog, HaploReg, eQTL analysis, etc. I think this paper needs a lot of improvements to be accepted for publication. Please address following issues as much as you can, and I will make a decision based on the revised manuscript. 1. The english writing is not good. There are lots of awkward or imperfect expressions. Please take a professional elglish editing to improve the overall quality I listed some of the examples below. - page2, in the abstract, "Liver cancer remains a global burden, ranking fourth in mortality rates globally" : being ranked fourth may be more appropriate. - page4, "One of the websites through a bioinformatics approach that discusses gene variation is Genome-Wide Association Studies (GWAS)" : GWAS should be changed to GWAS catalog. and a period is needed after the sentence. 2. Detailed explanations are not provided. - Only HaploReg results are provided without any interim process. - Aren't the rest of the significant SNPs from the GWAS catalog associated to any genes? 3. Some presentations are not correct. - In Fig 4, Africa pie chart are not matched with the proportions of the SNPs 4. Interpretation is not enough - In the last step, they provide allele frequencies for each of continents, but the association between allele frequencies and some statistics for the liver cancer is not provided. **• Reviewer B :** This submitted article suggested a pipeline that can draw more meaningful results using several bioinformatics databases and analysis tools. However, the writing of the manuscript is not clear to understand and should be improved more. I listed some of points to be improved to make it better manuscript. - In the part of providing statistics of SNPs from multiple continents, information in Table 4 is not matched with the information in Figure 4 - No interpretation is provided for continent specific allele information to the disease. - In the eQTL analysis, I could not understand how the two tisses in the Table 3 are associated to the disease.

Artikel hasil dari revisi yang pertama setelah mendapatkan masukkan dan saran oleh 2 reviewers pada tanggal 15 November 2023

November 28, 2023

Dear Editors,

We are pleased to submit our revised manuscript titled "A Genomic and Bioinformatic-based Approach to Identify Genetic Variants for Liver Cancer across Multiple Continents" for consideration as an original research article in Genomics & Informatics (GI23067). We are grateful for your encouraging feedback on our manuscript. Enclosed is the revised version, addressing the comments provided by the reviewers. The revised sections of the manuscript are highlighted in yellow. We would like to express our gratitude for the opportunity to refine our manuscript and hope these revisions meet your expectations. Your review and assistance are invaluable, and we look forward to your feedback.

Sincerely yours,

Apt. Lalu Muhammad Irham M.Farm Ph.D.

Faculty of Pharmacy,

Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Jl. Prof. DR. Soepomo SH, Warungboto,

Kec. Umbulharjo, Kota Yogyakarta, Daerah Istimewa Yogyakarta

Recommendation Reviewer 1:

This paper found new association results about liver cancer using various bioinformatics tools such as GWAS catalog, HaploReg, eQTL analysis, etc. I think this paper needs a lot of improvements to be accepted for publication. Please address following issues as much as you can, and I will make a decision based on the revised manuscript.

Answer: We are grateful for the detailed review and constructive feedback. Efforts have been made to address the highlighted concerns.

Q1: The english writing is not good. There are lots of awkward or imperfect expressions. Please take a professional elglish editing to improve the overall quality
I listed some of the examples below.

- page2, in the abstract, "Liver cancer remains a global burden, ranking fourth in mortality rates globally" : being ranked fourth may be more appropriate.
- page4, "One of the websites through a bioinformatics approach that discusses gene variation is Genome-Wide Association Studies (GWAS)" : GWAS should be changed to GWAS catalog. and a period is needed after the sentence.

A1: Thank you for pointing out the areas needing improvement. We have thoroughly revised the language and corrected the specified sentences on [Page 2 and 4, Lines 38 and 79-80].

"Liver cancer is the fourth leading cause of death worldwide".

"One of the websites through a bioinformatics approach that documents genetic variants is GWAS catalog".

Q2: Detailed explanations are not provided.
- Only HaploReg results are provided without any interim process.
- Aren't the rest of the significant SNPs from the GWAS catalog associated to any genes?

A2: Thank you to the reviewers who have provided us with input. We have corrected it according to the feedback in [Page 4, lines 115-116 and 119-123]. The sentences are revised as below :

"Subsequently, HaploReg version 4.1 was utilized, and a p-value $<10^{-8}$ was applied based on the number of SNPs obtained"

"Through our integrative bioinformatics approach, two variants with a missense mutation (rs rs2294915, rs2896019) that encoded the PNPLA3 genes were prioritized as the biological risk SNPs for Liver Cancer. Primary liver cancer, also known as hepatocellular carcinoma, is a pathological condition characterised by the development of malignant cells within the hepatic tissues"

Q3: Some presentations are not correct.
- In Fig 4, Africa pie chart are not matched with the proportions of the SNPs

A3: We acknowledge this discrepancy and have revised the relevant sections, specifically Page 7, lines 179-186, to ensure accuracy in the tables and figures:

Table 4. Analysis of allele frequencies for the *PNPLA3* gene from variant annotation (SNPnexus)

SNP	Gene	Location	Allele		Allele Frequency (N)				
			Ref	Eff*	AFR	AMR	EAS	EUR	SAS
rs2294915	*PNPLA3*	Missense	C	T	T: 0.163 (215)	T: 0.490 (340)	T: 0.365 (368)	T: 0.252 (254)	T: 0.246 (241)
rs2896019	*PNPLA3*	Missense	T	G	G: 0.158 (209)	G: 0.438 (304)	G: 0.364 (367)	G: 0.199 (200)	G: 0.236 (231)

Note : *PNPLA3*, patatin like phospoliphase 3; SNP, single nucleotide polymorphism; Ref, Reference; Alt, alternate; AFR, Africa; AMR, America; EAS, East Asia; EUR, Europe; SAS, Southeast Asia.

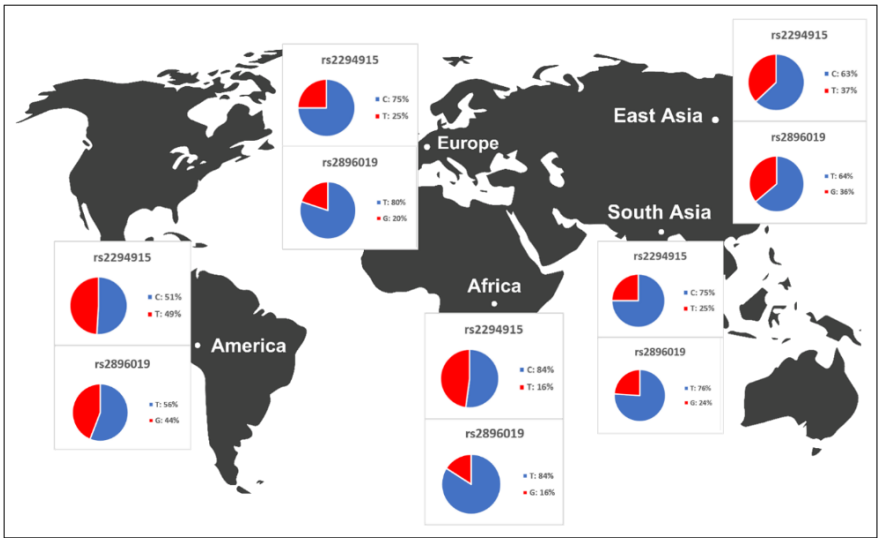


Figure 4. The results of the distribution of allele frequencies affecting *PNPLA3* across various populations. EUR, Europe; AFR, Africa; SAS, South Asia; EAS, East Asia.

Q4: Interpretation is not enough

- In the last step, they provide allele frequencies for each of continents, but the association between allele frequencies and some statistics for the liver cancer is not provided.

A4: Thank you for highlighting this oversight. We have enhanced our interpretation and revised the related sections, particularly on Pages 7 and 8, lines 169-176 and 195-199, to address this issue.

"The results of the research we have done, we have identified variants associated with liver cancer gene expression and carried out allele frequency analysis in various populations. As shown in Table 4, allele variant frequencies were evaluated in multiple people from Europe, America, East Asia, South Asia, and Africa. Allele frequencies across populations varied for each SNP, as depicted in Figure 4. Table 4 and Figure 4 show the gene expression levels at higher frequencies of the rs2294915 related allele (C) and the rs2896019 related allele (T). At the population frequency of the rs2294915 (C) allele, populations in Europe and South Asia were expressed at much higher levels than America, Africa, and East Asia".

"The allele frequencies of the T and G alleles "rs2294915" and "rs2896019" in African populations were expressed at much lower levels compared to American, European, and Southeast Asian people. Overall, the allele frequencies of the "rs2294915" and "rs2896019" variant alleles suggest a contribution to the prevalence of the variants for gene expression of PNPLA3".

Recommendation Reviewer 2:

This submitted article suggested a pipeline that can draw more meaningful results using several bioinformatics databases and analysis tools. However, the writing of the manuscript is not clear to understand and should be improved more. I listed some of points to be improved to make it better manuscript.

Answer: We appreciate the reviewer's detailed feedback and have made revisions to enhance the clarity of our manuscript.

Q1: In the part of providing statistics of SNPs from multiple continents, information in Table 4 is not matched with the information in Figure 4

A1: Thank you for highlighting this issue. We have revised the relevant sections to ensure consistency between Table 4 and Figure 4 [Page 7, lines 179-186], as below:

Table 4. Analysis of allele frequencies for the *PNPLA3* gene from variant annotation (SNPnexus)

SNP	Gene	Location	All le		Allele Frequency (N)				
			Ref	Eff*	AFR	AMR	EAS	EUR	SAS
rs2294915	*PNPLA3*	Missense	C	T	T: 0.163 (215)	T: 0.490 (340)	T: 0.365 (368)	T: 0.252 (254)	T: 0.246 (241)
rs2896019	*PNPLA3*	Missense	T	G	G: 0.158 (209)	G: 0.438 (304)	G: 0.364 (367)	G: 0.199 (200)	G: 0.236 (231)

Note : *PNPLA3*, patatin like phospoliphase 3; SNP, single nucleotide polymorphism; Ref, Reference; Alt, slternate; AFR, Africa; AMR, America; EAS, East Asia; EUR, Europe; SAS, Southeast Asia.

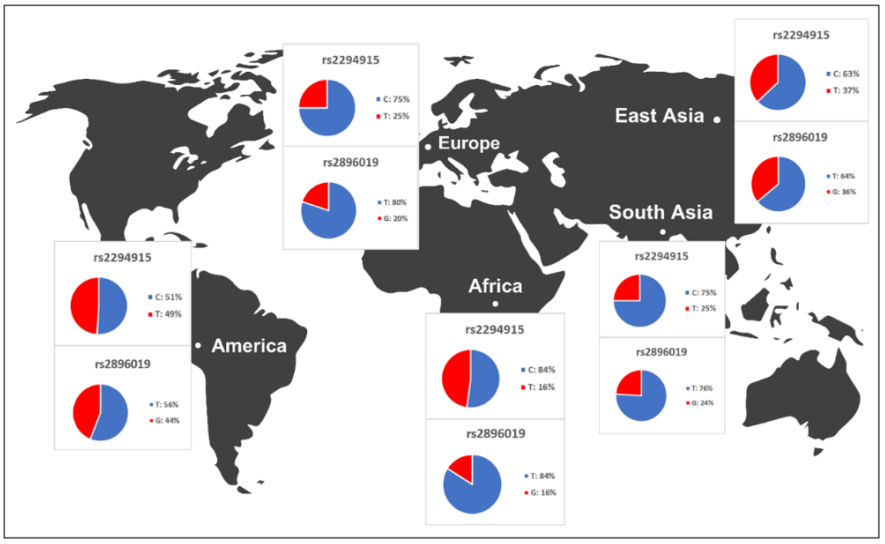


Figure 4. The results of the distribution of allele frequencies affecting *PNPLA3* across various populations. EUR, Europe; AFR, Africa; SAS, South Asia; EAS, East Asia.

Q2: No interpretation is provided for continent specific allele information to the disease

A2: We acknowledge the need for clearer interpretation and have revised the relevant sections to better articulate the association between allele frequencies and liver cancer across different continents [Page 7 and 8. Lines 169-176 and 195-199]. The sentences are revised as below:

"The results of the research we have done, we have identified variants associated with liver cancer gene expression and carried out allele frequency analysis in various populations. As shown in Table 4, allele variant frequencies were evaluated in multiple people from Europe, America, East Asia, South Asia, and Africa. Allele frequencies across populations varied for each SNP, as depicted in Figure 4. Table 4 and Figure 4 show the

gene expression levels at higher frequencies of the rs2294915 related allele (C) and the rs2896019 related allele (T). At the population frequency of the rs2294915 (C) allele, populations in Europe and South Asia were expressed at much higher levels than America, Africa, and East Asia".

"The allele frequencies of the T and G alleles "rs2294915" and "rs2896019" in African populations were expressed at much lower levels compared to American, European, and Southeast Asian people. Overall, the allele frequencies of the "rs2294915" and "rs2896019" variant alleles suggest a contribution to the prevalence of the variants for gene expression of PNPLA3".

Q3: In the eQTL analysis, I could not understand how the two tisses in the Table 3 are associated to the disease.

A3: Thank you for your valuable feedback. We have addressed your concerns in our revision, specifically in the eQTL analysis section on Pages 5 and 6, Lines 148-153 and 159-163. The relevant sentences have been revised for clarity as follows:

"The result in a correlation between the Gene Expression of PNPLA3 *and eQTL, to identify eQTLs associated with liver cancer gene expression, the GTEx database was used. We have identified minor alleles related to liver cancer, as presented in Table 3. Uniquely, several types of SNPs we found have high expression in skin tissue, namely rs2294915 and rs2896019. The CC type genotypes of rs2294915 and rs2896019 were associated with higher expression in suprapubic and underarm skin compared with the CT and TT type genotypes (Figure 3)".*

"The research results show that the genomic database can identify gene variations with the highest potential in the pathogenesis of liver cancer. Liver cancer is characterized by eyes and skin that appear yellow (Fitrianti et al., 2022). According to Nessa et. al. (2017), the severity of the liver can be gauged by the diminishing quality of liver function. The quality of liver function can be assessed from total bilirubin levels, serum albumin, and PT (partial thromboplastin time)".

A Genomic and Bioinformatic-based Approach to Identify Genetic Variants for Liver Cancer across Multiple Continents

Muhammad Ma'ruf[1], Lalu Muhammad Irham[1]*, Wirawan Adikusuma[2], Made Ary Sarasmita[3,4] Sabiah Khairi[5], Barkah Djaka Purwanto[6,7], Rockie Chong[8], Maulida Mazaya[9], Lalu Muhammad Harmain Siswanto[10]

[1]Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, 55164, Indonesia

[2]Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram, 83127, Indonesia

[3]Pharmacy Study Program, Faculty of Science and Mathematics, Udayana University, Bali, 80363, Indonesia

[4]School of Nursing, College of Nursing, Taipei Medical University, Taipei, 11031, Taiwan

[5]Faculty of Medicine, Universitas Ahmad Dahlan, Yogyakarta, 55191, Indonesia

[6]PKU Muhammadiyah Bantul Hospital, Bantul, Yogyakarta, 55711, Indonesia

[7]Department of Chemistry and Biochemistry, University of California, Los Angeles, 90095, USA

[8]Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong Science Center, Cibinong, 10340, Indonesia

[9]Mataram Training Health Center, Indonesia Ministry of Health, Mataram, 83237, Indonesia

[*]**Correspondence:** Lalu Muhammad Irham; email: lalu.irham@pharm.uad.ac.id

Abstract

Liver cancer is the fourth leading cause of death worldwide. Well-known risk factors include hepatitis B virus (HBV) and hepatitis C virus (HCV), along with exposure to aflatoxins, excessive alcohol consumption, obesity, and type 2 diabetes.. Genomic variants play a crucial role in mediating liver cancer among these factors. However, specific variants involved in this process remain under-explored. This study utilizes a bioinformatics approach to identify genetic variants associated with liver cancer from across various continents. The single nucleotide polymorphisms (SNPs) associated with liver cancer were retrieved from the Genome-Wide Association Studies (GWAS) catalog. The subsequent prioritization was performed using functional annotation with HaploReg v4.1 and the Ensembl database. Our results indicate that two variants, rs2294915 and rs2896019, encoded by the PNPLA3 gene, were found to be highly expressed in the liver tissue, as well as in the skin, cell-cultured fibroblasts and adipose-subcutaneous tissue, all of which contribute to the risk of liver cancer. We highlight the utility of this population-specific *PNPLA3* genetic variant for genetic association studies and for early prognosis and treatment of liver cancer. This study highlights the potential of integrating genomic databases with bioinformatic analysis to identify genetic variations involved in the pathogenesis of liver cancer. We recommend that future research prioritize the validation of these variations in clinical settings.

Keywords: *Genomic Variants, Liver Cancer, Bioinformatics*

Introduction

Liver cancer,a type of carcinoma, has the highest mortality rate in the world every year (McGlynn et al., 2021). In 2018, there were 841,000 cases of liver cancer, of which the death rate caused by liver cancer reached 782,000 (Bray et al., 2018).The average number of liver cancer cases and associated deaths can be two to three times higher in men in certain parts of the world. The Global Cancer Statistics (GLOBOCAN) in 2020 ranked liver cancer as the third most deadly cancer, accounting for 8.3% of cancer-related deaths. In 2020, liver cancer incidences reached 905,000, with a mortality rate of 830,000 (Sung et al., 2021). In Indonesia, liver cancer is the second most common in men, amounting to 12.4 per 100,000 of the Indonesian population, with an average death rate of 7.6 per 100,000 (Kemenkes RI, 2019).

Factors that cause liver cancer include chronic infection with hepatitis B virus (HBV), hepatitis C virus (HBC), the result of contamination with aflatoxin, alcohol consumption, history of obesity, history of type 2 diabetes, and smoking addiction (Bray et al., 2018). According to Villanueva (2019), other risk factors are thought to exacerbate the occurrence of liver cancer, such as an unhealthy lifestyle, geographic conditions, gender, age, family history of the disease, and the severity of damage to the liver. Liver cancer is also found in areas that have cases of hepatitis B. In these areas, liver cancer is prevalent at a young age. This is because some infected with hepatitis B are obtained vertically through the delivery process (Mittal & El-Serag, 2013).

Patients often felt complaints in the form of fatigue, pain, diarrhea, skin abnormalities, and decreased appetite, all of which have affected their quality of life (Waller et al., 2015). Therefore, detecting the presence of disease symptoms in liver cancer can be done by examining deoxyribonucleic acid (DNA). Gene variation can be associated with disease progression and pathogenesis, which includes liver cancer. One of the websites through a bioinformatics

approach that documents genetic variation is the GWAS catalog. GWAS catalog is a database with single nucleotide polymorphism (SNP) search results that has identified several variants associated with liver fat content, circulating liver enzymes, and the development of Non-Alcoholic Fatty Liver Disease (NAFLD) as well as genetic markers used in predicting a disease disorder (Wang et al., 2021).

Genetic identification in humans aims to identify inherited genetic risk factors for liver cancer. This study uses the GWAS catalog database to map genes from genetic variations across several populations that play an essential role in the pathogenesis of liver cancer. The most significant gene variations based on their function in protein changes will be further verified.

Methods

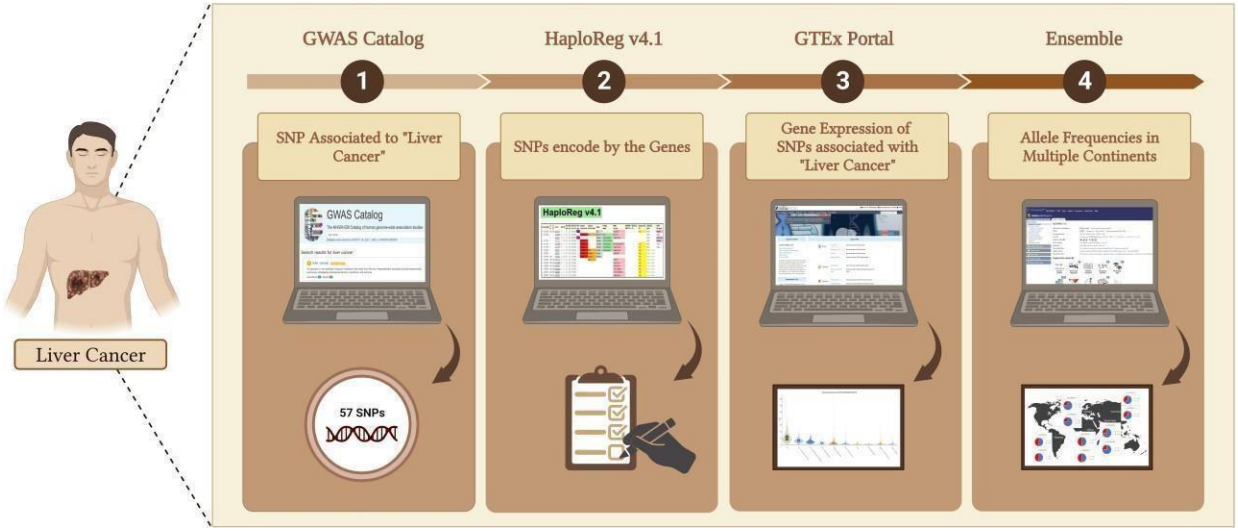


Figure 1. Analysis methodology for integrated bioinformatic, database and genomic analysis of genetic variation that affect liver cancer. The figure was created with BioRender.com under agreement number "FM25OO073C"

In this study, we adopted the method used by Ma'ruf et al (2023) and Puspitaningrum et al (2022), as depicted in Figure 1. Liver cancer-associated SNPs were obtained from the GWAS Catalog database (http://www.ebi.ac.uk/gwas; accessed on 15-02-2023). Subsequently, we performed further analysis using HaploReg (version 4.1) applying a p-value < 10^{-8} to account for multiple tests in the GWAS catalog. This threshold is commonly used to identify associations between common genetic variants and traits with adjacent gene expression (Chen et al., 2021). Furthermore, to evaluate the relationships between various genetic variants and gene expression profiles, we utilized e-QTL analysis with data sourced from the GTEx Portal database (http://www.gtexportal.org/home/; accessed on 16-02-2023), considering gene expression across various tissues in humans. Additionally, we confirmed the identified variants using the Ensembl Genome Browser (https://www.ensembl.org/index.html; accessed on 17-02-2023). Our study considered allele frequencies in populations from Europe, Africa, America, East Asia, and

Southeast Asia. To comprehend the functionalities of different gene variants, we performed evaluations using the SNP nexus database (https://www.snp-nexus.org; accessed on 20-02-2023).

Results and Discussion

1. Identification of Genomic Variants of Liver Cancer

This study identified SNPs associated with liver cancer from the GWAS catalog. Among theses SNPs, 29 of them were further confirmed through SNP duplication, as shown in Table 1. Subsequently, HaploReg version 4.1 was utilized, and a p-value $<10^{-8}$ was applied based on the number of SNPs obtained. Based on the findings presented in Table 2, we found the risk of two genes for "Liver Cancer" disease. This study analyzed tissue expression affecting liver cancer, focusing on the missense variant *PNPLA3.*

Through our integrative bioinformatics approach, two variants with a missense mutation (rs rs2294915, rs2896019) that encoded the *PNPLA3* genes were prioritized as the biological risk SNPs for Liver Cancer. Primary liver cancer, also known as hepatocellular carcinoma, is a pathological condition characterised by the development of malignant cells within the hepatic tissues. The development of cancer in extraneous anatomical sites that then metastasizes to the liver does not constitute primary liver cancer. Primary liver cancer encompasses many kinds, including hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (iCCA), and less frequent varieties such as mixed hepatocellular cholangiocarcinoma (HCC-CCA), fibrolamellar HCC (FLC), and the paediatric neoplasm hepatoblastoma (Wage et al., 2021).

Table 1. SNPs from the GWAS catalog with p-value $<10^{-8}$

No.	Variation and risk allele	p-value
1	rs2856723	$3x10^{-43}$
2	rs34675408	$1x10^{-32}$
3	rs9272105	$5x10^{-22}$
4	rs913493	$5x10^{-20}$
5	rs2294915	$2x10^{-19}$
6	rs17401966	$2x10^{-18}$
7	rs3096380	$1x10^{-17}$
8	rs9275319	$3x10^{-17}$
9	rs584368	$2x10^{-14}$
10	rs2596542	$4x10^{-13}$
11	rs1110446	$9x10^{-13}$
12	rs58489806	$3x10^{-12}$
13	rs6078460	$2x10^{-11}$
14	rs2523961	$6x10^{-11}$
15	rs7574865	$2x10^{-10}$
16	rs1110446	$3x10^{-10}$
17	rs455804	$5x10^{-10}$
18	rs58542926	$6x10^{-10}$
19	rs2523961	6x10-^{10}
20	rs8107030	$8x10^{-10}$
21	rs10272859	$9x10^{-10}$
22	rs190121281	$4x10^{-9}$
23	rs9275572	$6x10^{-9}$
24	rs2242652	$6x10^{-9}$
25	rs188273166	$1x10^{-8}$
26	rs708113	$1x10^{-8}$
27	rs2896019	$2x10^{-8}$
28	rs17047200	$3x10^{-8}$
29	rs541860626	$5x10^{-8}$

Table 2. Variants and risk alleles of liver cancer encoding prioritized SNPs

Variationandrisk alleles	Variantsnearrisk allele ($r^2 > 0.8$)	*p*-value	GENCODE	Type of allele
rs2294915	rs738409	$2x10^{-19}$	*PNPLA3*	missense
rs2896019	rs3761472	$2x10^{-8}$	*PNPLA3*	missense

2. Gene expression of *PNPLA3* across 10 human tissues

The results of *PNPLA3* gene expression across 10 human tissues comprise the most apparent functional consequences of genetic variation. Liver, sun-exposed skin (lower legs), non-sun-exposed skin (suprapubic), and adipose subcutaneous fibroblasts and cell cultures showed the highest *PNPLA3* gene expression in the 10 human tissues analyzed from GTEx (Figure 2). In addition, we have found that the SNP IDs rs2294915 and rs2896019 have similar gene expression variations in Sun-Exposed skin (lower legs). The exciting thing about these findings is that liver cancer patients often experience complaints that their skin appears yellow. Further results showed that the *PNPLA3* gene has high expression in suprapubic and underarm skin.

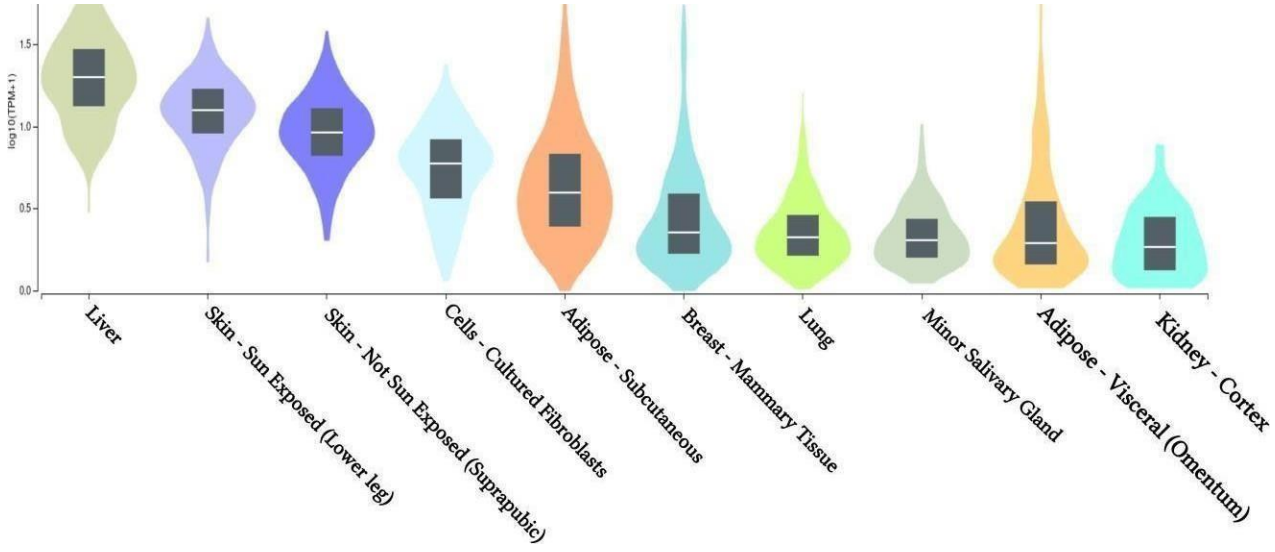


Figure 2. *PNPLA3* gene expression associated with liver cancer across human tissues based on GTEx Portal analysis

3. Correlation between Gene Expression of *PNPLA3* and eQTL

The result in a correlation between the Gene Expression of *PNPLA3* and eQTL, to identify eQTLs associated with liver cancer gene expression, the GTEx database was used. We have identified minor alleles related to liver cancer, as presented in Table 3. Uniquely, several types of SNPs we found have high expression in skin tissue, namely rs2294915 and rs2896019. The CC type genotypes of rs2294915 and rs2896019 were associated with higher expression in suprapubic and underarm skin compared with the CT and TT type genotypes (Figure 3).

Table 3. Results of e-QTL in liver cancer from the GTEx portal database

SNP	Gencode ID (ENSG00000-)	Gene Symbol	*p*-value	Effect Size	Tissue	Expression Level
rs2294915	100344.10	*PNPLA3*	2.8×10^{-8}	-0.15	Skin - Sun Exposed (Lower leg)	CC>CT>TT
	100344.10	*PNPLA3*	5×10^{-8}	-0.50	Skin - Not Sun Exposed (Suprapubic)	CC>CT>TT
rs2896019	100344.10	*PNPLA3*	6.7×10^{-11}	-0.19	Skin - Sun Exposed (Lower leg)	CC>CT>TT
	100344.10	*PNPLA3*	2×10^{-9}	-0.22	Skin - Not Sun Exposed (Suprapubic)	CC>CT>TT

Source: Expression Quantitative Trait Loci (eQTL) obtained from the GTEx Portal.
Note : *PNPLA3*, patatin like phospoliphase 3; SNP, single nucleotide polymorphism; eQTL, expression quantitative trait loci; SNP, single nucleotide polymorphism.

The research results show that the genomic database can identify gene variations with the most potential in the pathogenesis of liver cancer. Liver cancer is characterized by eyes and skin that appear yellow (Fitrianti et al., 2022). According to Nessa et al (2017), the severity of the liver can be seen from the decreasing quality of the liver. The quality of the liver can be assessed from total bilirubin levels, serum albumin, and PT (partial thromboplastin time).

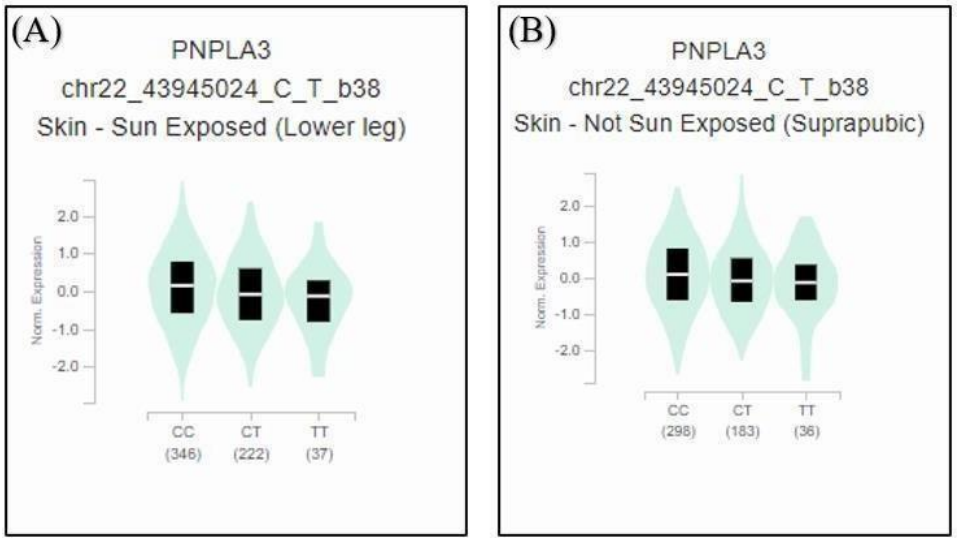


Figure 3. Patatin like phospolipase domain containing protein 3 (*PNPLA3*) gene expression for each genotype of the single nucleotide polymorphism (SNP): (A) rs2294915 and (B) rs2896019.

4. Allele frequencies of candidate variants in populations in different continents

The results of the research we have done, we have identified variants associated with liver cancer gene expression and carried out allele frequency analysis in various populations. As shown in Table 4, allele variant frequencies were evaluated in multiple people from Europe, America, East Asia, South Asia, and Africa. Allele frequencies across populations varied for each SNP, as depicted in Figure 4. Table 4 and Figure 4 show the gene expression levels at higher frequencies of the rs2294915 related allele (C) and the rs2896019 related allele (T). At the population frequency of the rs2294915 (C) allele, populations in Europe and South Asia were expressed at much higher levels than America, Africa, and East Asia.

Table 4. Analysis of allele frequencies for the *PNPLA3* gene from variant annotation (SNPnexus)

SNP	Gene	Location	Allele		Allele Frequency (N)			
			Ref	Eff*	AFR	AMR	EAS	EUR
rs2294915	*PNPLA3*	Missense	C	T	T: 0.163 (215)	T: 0.490 (340)	T: 0.365 (368)	T: 0.252 (254)
rs2896019	*PNPLA3*	Missense	T	G	G: 0.158 (209)	G: 0.438 (304)	G: 0.364 (367)	G: 0.199 (200)

Note : *PNPLA3*, patatin like phospolipase domain containting protein 3; SNP, single nucleotide polymorphism; Ref, Reference; Alt, Alternate; AFR, Africa; AMR, America; EAS, East Asia; EUR, Europe; SAS, Southeast Asia.

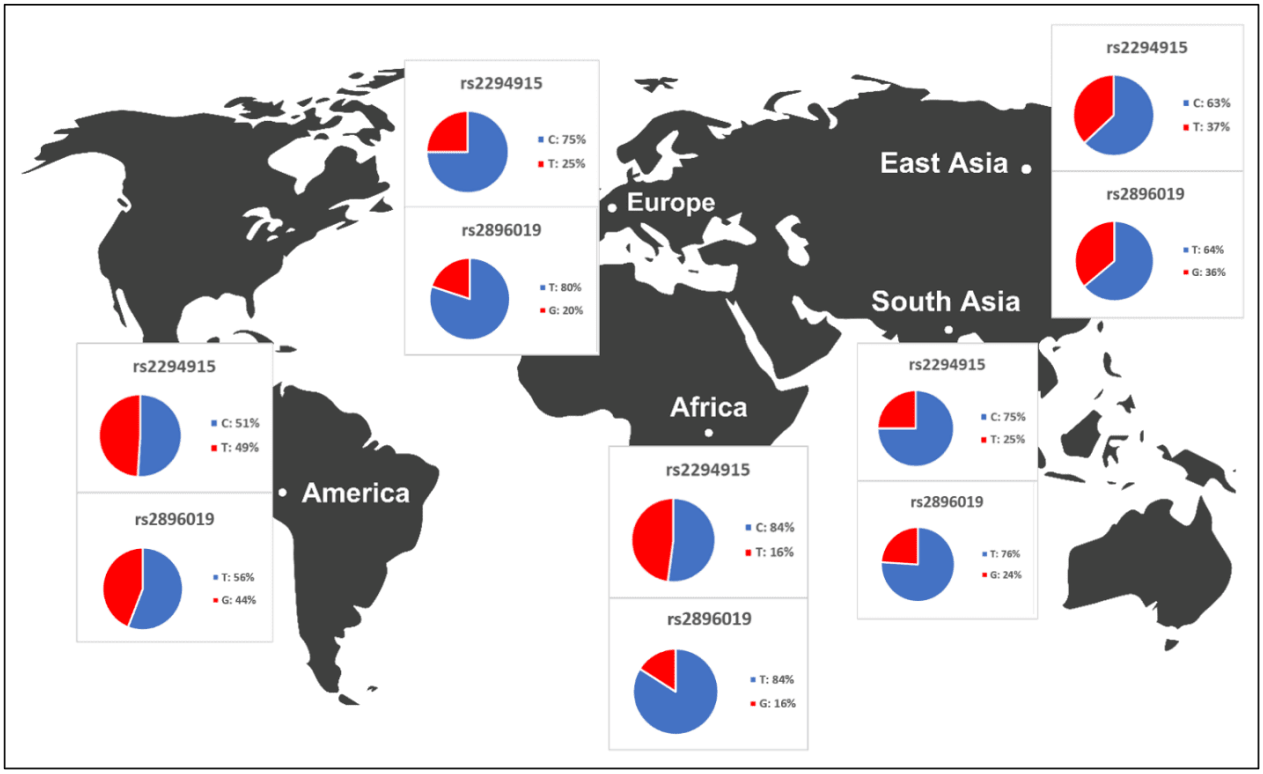


Figure 4. The results of the distribution of allele frequencies affecting *PNPLA3* across various populations. EUR, Europe; AFR, Africa; SAS, South Asia; EAS, East Asia.

Based on this finding, rs2294915 and rs2896019 are potentially related to the susceptibility to "liver cancer with effect size the highest score of -0.50 can be interpreted on the skin not exposed to sunlight (suprapubic)." According to Poggiali & Vercelli (2023), this condition is characterized by a disruption in the heme biosynthesis pathway due to reduced hepatic uroporphyrinogen decarboxylase (UROD) activity. The consequence of this phenomenon is the buildup of light-sensitive by-products, such as uroporphyrinogen, leading to the development of fragility and blistering of sun-exposed skin, as well as impairment of liver function.

The allele frequencies of the T and G alleles "rs2294915" and "rs2896019" in African populations were expressed at much lower levels compared to American, European, and Southeast Asian people. Overall, the allele frequencies of the "rs2294915" and "rs2896019" variant alleles suggest a contribution to the prevalence of the variants for gene expression of *PNPLA3*.

Across human populations, the frequency of the T allele at rs2294915 was associated with a high expression of *PNPLA3* in liver cancer, which is much lower in African populations (16%)

compared to South Asians (25%), Europeans (25%), East Asians (37%) and America (49%). In contrast, the frequency of the C allele at rs2296019 was considerably higher in African (84%), European (80%), South Asian (76%), East Asian (64%), and American (56%) populations.

Patients with liver cancer who have a history of alcohol addiction to an amount of ≥ 3 drinks per day can increase the risk of liver cancer by 16% in the general population; diabetics and people with central obesity also increase the risk of liver cancer by 2 times (McGlynn et al., 2021). The diagnosis of liver cancer in patients often involves using serological testing in conjunction with imaging techniques, which is considered the established approach for identifying liver carcinoma. Nevertheless, the sensitivity of the often-used serological test, specifically designed to detect alpha-fetoprotein (AFP), is at around 60%. Imaging modalities, including magnetic resonance imaging (MRI), computed tomography (CT), and ultrasonography (US), exhibit notable levels of sensitivity and specificity in the identification of liver cancer, particularly in individuals afflicted with liver cirrhosis (Huang et al., 2022).

Variant alleles (rs2294915 and rs2896019) are associated with liver cancer. Accordingly, populations from the continents of Africa, America, East Asia, Europe, and South Asia show associated *PNPLA3* expression, resulting in a higher susceptibility to liver cancer. Identifying unique and pathogenic gene variations for a disease is very interesting for both research and clinical validation. These variants not only offer insights into disease susceptibility but also serve as potential diagnostic and prognostic biomarkers (Irham et al., 2020). Additionally, they can facilitate the identification of drug target candidates, a concept known as genomic-driven drug repurposing (Afief et al., 2022). We anticipate that the discovery of candidate gene variations in *PNPLA3* will pave the way for successful clinical validation, potentially establishing this as a promising diagnostic and prognostic biomarker for liver cancer.

It is important to acknowledge that the genetic variants identified in this study as potentially pathogenic are based on preliminary investigations using genomic and bioinformatics databases. However, these findings offer crucial insights for future researchers intending to validate these genetic variants in liver cancer patients. We strongly recommend future research t incorporate further functional annotations, which would further aid in prioritizing these pathogenic genetic variants.

Conclusion

This study identifies genetic variants influencing “liver cancer” and reveals the significance of the *PNPLA3* gene in liver tissue, as well as in skin regions exposed to the sun (lower legs), skin regions not exposed to the sun (suprapubic), cultured fibroblasts, and adipose-subcutaneous tissue. These findings collectively contribute to an increased risk of liver cancer development. The observed variations in allele frequencies of the two identified variants, rs2294915, and rs2896019, across populations from Africa, America, East Asia, Europe, and South Asia, significantly impact *PNPLA3* gene expression. Consequently, these population groups exhibit varying susceptibilities to liver cancer based on associated *PNPLA3* expression levels. These discoveries highlight the critical relevance of understanding genomic variations in precision medicine and designing screening strategies for liver cancer across diverse populations on different continents.

Reference

Afief, A. R., Irham, L. M., Adikusuma, W., Perwitasari, D. A., Brahmadhi, A., & Cheung, R. (2022). Integration of genomic variants and bioinformatic-based approach to drive drug repurposing for multiple sclerosis . *Biochemistry and Biophysics Reports , 101337*.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Chen, Z., Boehnke, M., Wen, X., & Mukherjee, B. (2021). Revisiting the genome-wide significance threshold for common variant GWAS. *G3: Genes, Genomes, Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkaa056

Fitrianti, F., Shatri, H., Faisal, E., Putranto, R., Agung, R. A., Nainggolan, L., Kalista, K. F., & Irawan, C. (2022). Pain Management in Advanced Hepatocellular Carcinoma: A Case Study. *Jurnal Penyakit Dalam Indonesia*, *9*(1), 53–58.

Huang, L., Sun, H., Sun, L., Shi, K., Chen, Y., Ren, X., Ge, Y., Liu, X., Knoll, W., Zhang, Q., & Wang, Y. (2022). Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning. *Nature Communications*, *14*(48), 1–14.

Irham, L. M., Chou, W. H., Calkins, M. J., Adikusuma, W., Hsieh, S. L., & Chang, W. C. (2020). Genetic variants that influence SARS-CoV-2 receptor TMPRSS2 expression among population cohorts from multiple continents. *Biochemical and Biophysical Research Communications*, *529*(2), 263–269. https://doi.org/10.1016/j.bbrc.2020.05.179

Kementerian Kesehatan Republik Indonesia. (2019). *Hari Kanker Sedunia 2019*. https://www.kemkes.go.id/article/view/19020100003/harikanker-sedunia-2019.html [Diakses pada 21 Mei 2023].

Ma'ruf, M., Fadli, J. C., Mahendra, M. R., Irham, L. M., Sulistyani, N., Adikusuma, W., Chong, R., & Septama, A. W. (2023). A bioinformatic approach to identify pathogenic variants for Stevens Johnson syndrome. *Genomics & Informatics*, *21*(2), 1–9.

McGlynn, K. A., Petrick, J. L., & El-Serag, H. B. (2021). Epidemiology of Hepatocellular Carcinoma. In *Hepatology* (Vol. 73, Issue S1, pp. 4–13). John Wiley and Sons Inc. https://doi.org/10.1002/hep.31288

Mittal, S., & El-Serag, H. B. (2013). *Epidemiology of Hepatocellular Carcinoma Consider the Population*. www.jcge.com

Nessa, M., Yusuf, F., & Mudatsir. (2017). Hubungan Child Pugh Score dengan Kualitas Hidup pada Pasien Sirosis Hati di Bagian Penyakit Dalam RSUDZA Banda Aceh. *Jurnal Ilmiah Mahasiswa Kedokteran Biomedis*, *2*(4), 24–34.

Poggiali, E., & Vercelli, A. (2023). Sun, alcohol, and skin lesions. *Emergency Care Journal*, *19*(11294).

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), 71–81. https://doi.org/10.12928/mf.v19i2.23706

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

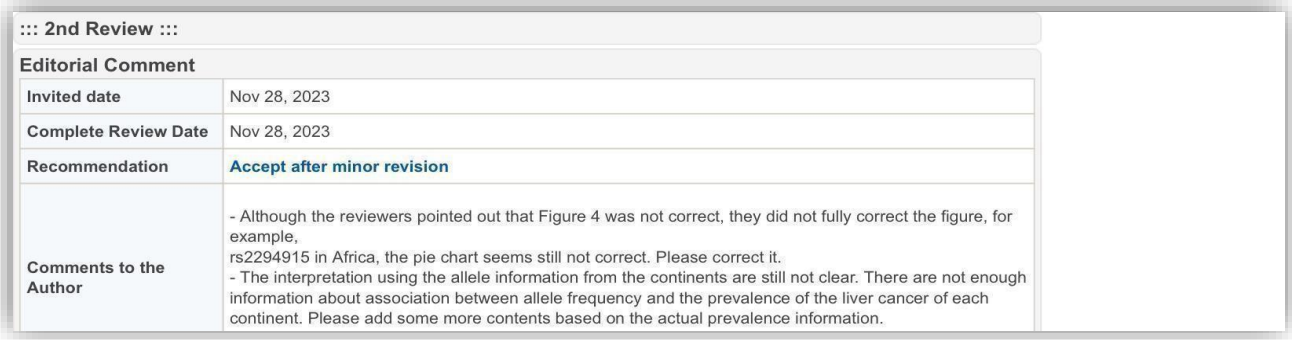
Villanueva, A. (2019). Hepatocellular Carcinoma. *New England Journal of Medicine*, *380*(15), 1450–1462. https://doi.org/10.1056/NEJMra1713263

Wage, H., Schulze, K., Felden, J. V., Calderaro, J., & Reig, M. (2021). Rare variants of primary liver cancer: Fibrolamellar, combined, and sarcomatoid hepatocellular carcinomas. *J. Med Genet*, *64*(11).

Waller, L. P., Deshpande, V., & Pyrsopoulos, N. (2015). Hepatocellular carcinoma: A comprehensive review. In *World Journal of Hepatology* (Vol. 7, Issue 26, pp. 2648–2663). Baishideng Publishing Group Co. https://doi.org/10.4254/wjh.v7.i26.2648

Wang, J., Conti, D. V, Bogumil, D., Sheng, X., Noureddin, M., Wilkens, L. R., Le Marchand, L., Rosen, H. R., Haiman, C. A., & Setiawan, V. W. (2021). Association of Genetic Risk Score With NAFLD in An Ethnically Diverse Cohort. *Hepatology CommuniCations*, *5*(10), 1689–1703. https://doi.org/10.1002/hep4.1751/suppinfo

Pada tanggal 28 November 2023 artikel mendapatkan revisian yang kedua oleh reviewers.



::: 2nd Review :::

Editorial Comment

Invited date	Nov 28, 2023
Complete Review Date	Nov 28, 2023
Recommendation	Accept after minor revision
Comments to the Author	- Although the reviewers pointed out that Figure 4 was not correct, they did not fully correct the figure, for example, rs2294915 in Africa, the pie chart seems still not correct. Please correct it. - The interpretation using the allele information from the continents are still not clear. There are not enough information about association between allele frequency and the prevalence of the liver cancer of each continent. Please add some more contents based on the actual prevalence information.

Artikel hasil dari revisi yang kedua setelah mendapatkan masukan dan saran oleh 2 reviewers pada tanggal 28 November 2023

November 29, 2023

Dear Editors,

We are pleased to submit our revised manuscript titled "A Genomic and Bioinformatic-based Approach to Identify Genetic Variants for Liver Cancer across Multiple Continents" for consideration as an original research article in Genomics & Informatics (GI23067). We are grateful for your encouraging feedback on our manuscript. Enclosed is the revised version, addressing the comments provided by the reviewers. The revised sections of the manuscript are highlighted in yellow. We would like to express our gratitude for the opportunity to refine our manuscript and hope these revisions meet your expectations. Your review and assistance are invaluable, and we look forward to your feedback.

Sincerely yours,

Apt. Lalu Muhammad Irham M.Farm Ph.D.

Faculty of Pharmacy,

Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Jl. Prof. DR. Soepomo SH, Warungboto,

Kec. Umbulharjo, Kota Yogyakarta, Daerah Istimewa Yogyakarta

Recommendation Reviewer 2:
Accept after minor revision

Answer: We are grateful for the detailed review and constructive feedback. Efforts have been made to address the highlighted concerns.

Q1: Although the reviewers pointed out that Figure 4 was not correct, they did not fully correct the figure, for example, rs2294915 in Africa, the pie chart seems still not correct. Please correct it.

A1: We acknowledge this discrepancy and have revised the relevant sections, specifically Page 7, lines 184, to ensure accuracy in the figures:

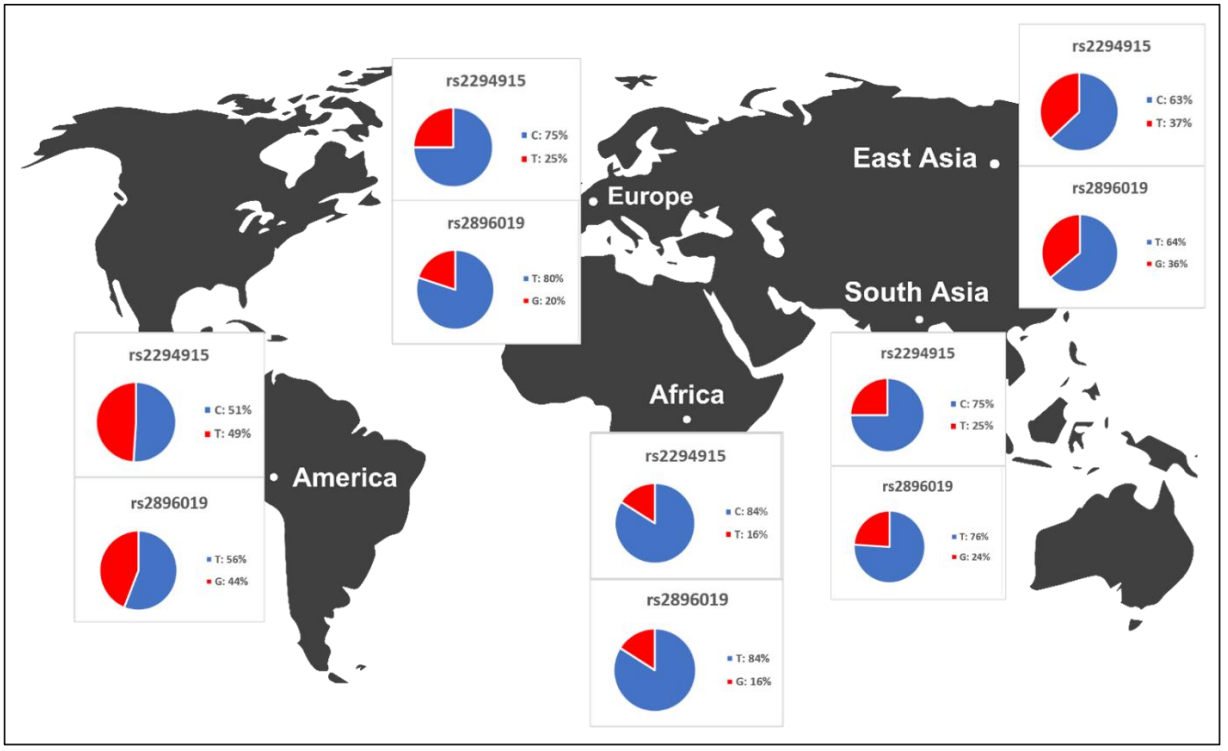


Figure 4. The results of the distribution of allele frequencies affecting *PNPLA3* across various populations. EUR, Europe; AFR, Africa; SAS, South Asia; EAS, East Asia.

Q2: The interpretation using the allele information from the continents are still not clear. There are not enough information about association between allele frequency and the prevalence of the liver cancer of each continent. Please add some more contents based on the actual prevalence information.

A2: Thank you for highlighting this oversight. We have enhanced our interpretation and revised the association between allele frequency and the prevalence of the liver cancer of each continent, in each of the following continents:

Abstract [Page 2, lines 51-54]

*"*We further obtained that these *two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate. Positive association of prevalence rates were underlined more frequent in East Asian and African population."*

Methods [Page 3, Lines 116-121]

"Furthermore, epidemiological and genomic data of the prevalence of liver cancer rates were obtained from Li et al (2022). The prevalence rates and allele frequencies of the variants in multiple continents were evaluated using IBM SPSS Statistics 25.0 with the Pearson Correlation test. After the preocedure was evaluated, the values of p-value 95% CI were obtained. All plots were created using line charts. A p-value less than 0.05 (P≤0.05) was considered as statistically significant in current study"

Result and Disscusion [Page 8, Lines 216-227]

"Next, the association between allele frequency and the prevalence of the liver cancer of each continent was evaluated. Data on the prevalence of liver cancer in the continents were obtained from Li et al (2022) (Li et al., 2023). Herein, two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate of liver cancer in multiple continents (Africa, America, East Asia, Europe, South Asia) based on the pearson's correlation analysis (P-value<0.011) (Figure 5). Populations with higher frequencies of variant alleles of these polymorphisms are thought to have a higher prevalence. We highlighted that these two variants (rs2294915 and rs2896019) were more frequent in the East Asian and Africa, which performed the higher aggressiveness of liver cancer in East Asian and African compared to America, Europe and South Asia. This study might give an insight that East Asian and African with carriers variants rs2294915 and rs2896019 might be more highly susceptible to suffer the liver cancer".

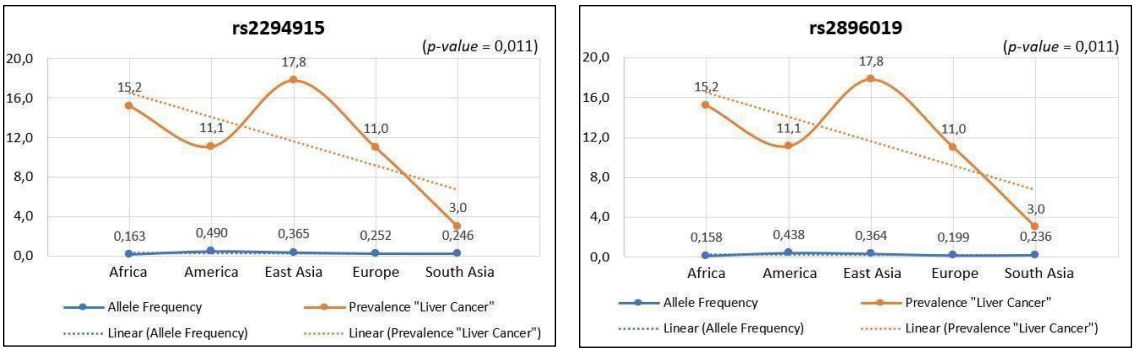


Figure 5. The association between allele frequency and the prevalence of liver cancer of each continent

Conclusion [Page 10, Lines 266-271]

"Our study also demonstrated that these *two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate. Positive association of prevalence rates were underlined more frequent in East Asian and African population. The higher frequency of the variants allele of these polymorphisms in population, the higher the estimated prevalence rates. The variants investigated in this study were likely to predispose to liver cancer and could play a role in its progression and aggressiveness"*

A Genomic and Bioinformatic-based Approach to Identify Genetic Variants for Liver Cancer across Multiple Continents

Muhammad Ma'ruf[1], Lalu Muhammad Irham[1]*, Wirawan Adikusuma[2], Made Ary Sarasmita[3,4] Sabiah Khairi[5], Barkah Djaka Purwanto[6,7], Rockie Chong[8], Maulida Mazaya[9], Lalu Muhammad Harmain Siswanto[10]

[1]Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

[2]Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram, Indonesia

[3]Department of Clinical Pharmacy, College of Pharmacy, Taipei Medical University, Taipei, Taiwan

[4]Pharmacy Study Program, Faculty of Science and Mathematics, Udayana University, Bali, Indonesia

[5]School of Nursing, College of Nursing, Taipei Medical University, Taipei 11031, Taiwan

[6]Faculty of Medicine, Universitas Ahmad Dahlan, Yogyakarta 55191, Indonesia

[7]PKU Muhammadiyah Bantul Hospital, Bantul, Yogyakarta 55711, Indonesia

[8]Department of Chemistry and Biochemistry, University of California, Los Angeles, USA

[9]Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong Science Center, Cibinong, Indonesia

[10]Mataram Training Health Center, Indonesia Ministry of Health, Indonesia

[*]**Correspondence:** Lalu Muhammad Irham; email: lalu.irham@pharm.uad.ac.id

Abstract

Liver cancer is the fourth leading cause of death worldwide. Well-known risk factors include hepatitis B virus (HBV) and hepatitis C virus (HCV), along with exposure to aflatoxins, excessive alcohol consumption, obesity, and type 2 diabetes. Genomic variants play a crucial role in mediating liver cancer among these factors. However, specific variants involved in this process remain under-explored. This study utilizes a bioinformatics approach to identify genetic variants associated with liver cancer from across various continents. The single nucleotide polymorphisms (SNPs) associated with liver cancer were retrieved from the Genome-Wide Association Studies (GWAS) catalog. The subsequent prioritization was performed using functional annotation with HaploReg v4.1 and the Ensembl database. The prevalence and allele frequencies of each variants were evaluated by using pearson correlation. Our results indicate that two variants, rs2294915 and rs2896019, encoded by the *PNPLA3* gene, were found to be highly expressed in the liver tissue, as well as in the skin, cell-cultured fibroblasts and adipose-subcutaneous tissue, all of which contribute to the risk of liver cancer. We further obtained that these two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate. Positive association of prevalence rates were underlined more frequent in East Asian and African population. We highlight the utility of this population-specific *PNPLA3* genetic variant for genetic association studies and for early prognosis and treatment of liver cancer. This study highlights the potential of integrating genomic databases with bioinformatic analysis to identify genetic variations involved in the pathogenesis of liver cancer. The genetic variants investigated in this study were likely to predispose to liver cancer and could affect its progression and aggressiveness. We recommend that future research prioritize the validation of these variations in clinical settings.

Keywords: *Genomic Variants, Liver Cancer, Bioinformatics*

Introduction

Liver cancer,a type of carcinoma, has the highest mortality rate in the world every year (McGlynn et al., 2021). In 2018, there were 841,000 cases of liver cancer, of which the death rate caused by liver cancer reached 782,000 (Bray et al., 2018).The average number of liver cancer cases and associated deaths can be two to three times higher in men in certain parts of the world. The Global Cancer Statistics (GLOBOCAN) in 2020 ranked liver cancer as the third most deadly cancer, accounting for 8.3% of cancer-related deaths. In 2020, liver cancer incidences reached 905,000, with a mortality rate of 830,000 (Sung et al., 2021). In Indonesia, liver cancer is the second most common in men, amounting to 12.4 per 100,000 of the Indonesian population, with an average death rate of 7.6 per 100,000 (Kemenkes RI, 2019).

Factors that cause liver cancer include chronic infection with hepatitis B virus (HBV), hepatitis C virus (HBC), the result of contamination with aflatoxin, alcohol consumption, history of obesity, history of type 2 diabetes, and smoking addiction (Bray et al., 2018). According to Villanueva (2019), other risk factors are thought to exacerbate the occurrence of liver cancer, such as an unhealthy lifestyle, geographic conditions, gender, age, family history of the disease, and the severity of damage to the liver. Liver cancer is also found in areas that have cases of hepatitis B. In these areas, liver cancer is prevalent at a young age. This is because some infected with hepatitis B are obtained vertically through the delivery process (Mittal & El-Serag, 2013).

Patients often felt complaints in the form of fatigue, pain, diarrhea, skin abnormalities, and decreased appetite, all of which have affected their quality of life (Waller et al., 2015). Therefore, detecting the presence of disease symptoms in liver cancer can be done by examining deoxyribonucleic acid (DNA). Gene variation can be associated with disease progression and pathogenesis, which includes liver cancer. One of the websites through a bioinformatics approach that documents genetic variation is the GWAS catalog. GWAS catalog is a database with single nucleotide polymorphism (SNP) search results that has identified several variants associated with liver fat content, circulating liver enzymes, and the development of Non-Alcoholic Fatty Liver Disease (NAFLD) as well as genetic markers used in predicting a disease disorder (Wang et al., 2021).

Genetic identification in humans aims to identify inherited genetic risk factors for liver cancer. This study uses the GWAS catalog database to map genes from genetic variations across several populations that play an essential role in the pathogenesis of liver cancer. The most significant gene variations based on their function in protein changes will be further verified.

Methods

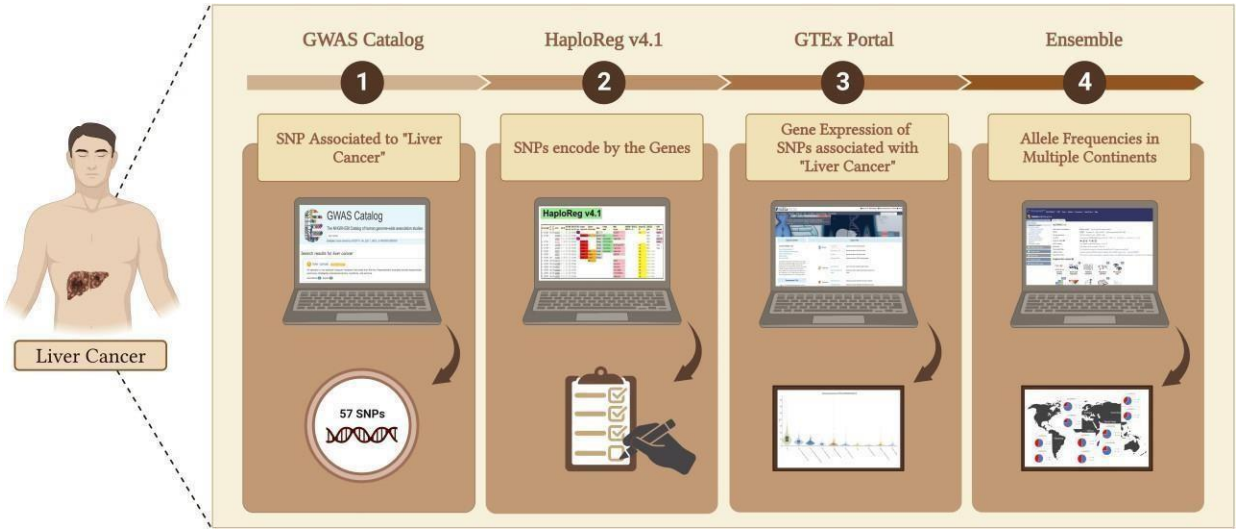


Figure 1. Analysis methodology for integrated bioinformatic, database and genomic analysis of genetic variation that affect liver cancer. The figure was created with BioRender.com under agreement number "FM25OO073C"

In this study, we adopted the method used by Ma'ruf et al (2023) and Puspitaningrum et al (2022), as depicted in Figure 1. Liver cancer-associated SNPs were obtained from the GWAS Catalog database (http://www.ebi.ac.uk/gwas; accessed on 15-02-2023). Subsequently, we performed further analysis using HaploReg (version 4.1) applying a p-value < 10^{-8} to account for multiple tests in the GWAS catalog. This threshold is commonly used to identify associations between common genetic variants and traits with adjacent gene expression (Chen et al., 2021). Furthermore, to evaluate the relationships between various genetic variants and gene expression profiles, we utilized e-QTL analysis with data sourced from the GTEx Portal database

(http://www.gtexportal.org/home/; accessed on 16-02-2023), considering gene expression across various tissues in humans. Additionally, we confirmed the identified variants using the Ensembl Genome Browser (https://www.ensembl.org/index.html; accessed on 17-02-2023). Our study considered allele frequencies in populations from Europe, Africa, America, East Asia, and Southeast Asia. To comprehend the functionalities of different gene variants, we performed evaluations using the SNP nexus database (https://www.snp-nexus.org; accessed on 20-02-2023). Furthermore, epidemiological and genomic data of the prevalence of liver cancer rates were obtained from Li et al (2022). The prevalence rates and allele frequencies of the variants in multiple continents were evaluated using IBM SPSS Statistics 25.0 with the Pearson Correlation test. After the preocedure was evaluated, the values of *p-value* 95% CI were obtained. All plots were created using line charts. A *p-value* less than 0.05 ($P \leq 0.05$) was considered as statistically significant in current study.

Results and Discussion

1. Identification of Genomic Variants of Liver Cancer

This study identified SNPs associated with liver cancer from the GWAS catalog. Among theses SNPs, 29 of them were further confirmed through SNP duplication, as shown in Table 1. Subsequently, HaploReg version 4.1 was utilized, and a p-value $<10^{-8}$ was applied based on the number of SNPs obtained. Based on the findings presented in Table 2, we found the risk of two genes for "Liver Cancer" disease. This study analyzed tissue expression affecting liver cancer, focusing on the missense variant *PNPLA3*.

Through our integrative bioinformatics approach, two variants with a missense mutation (rs rs2294915, rs2896019) that encoded the *PNPLA3* genes were prioritized as the biological risk SNPs for Liver Cancer. Primary liver cancer, also known as hepatocellular carcinoma, is a pathological condition characterised by the development of malignant cells within the hepatic tissues. The development of cancer in extraneous anatomical sites that then metastasizes to the liver does not constitute primary liver cancer. Primary liver cancer encompasses many kinds, including hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (iCCA), and less frequent varieties such as mixed hepatocellular cholangiocarcinoma (HCC-CCA), fibrolamellar HCC (FLC), and the paediatric neoplasm hepatoblastoma (Wage et al., 2021).

Table 1. SNPs from the GWAS catalog with p-value $<10^{-8}$

No.	Variation and risk allele	p-value
1	rs2856723	$3x10^{-43}$
2	rs34675408	$1x10^{-32}$
3	rs9272105	$5x10^{-22}$
4	rs913493	$5x10^{-20}$
5	rs2294915	$2x10^{-19}$
6	rs17401966	$2x10^{-18}$
7	rs3096380	$1x10^{-17}$
8	rs9275319	$3x10^{-17}$
9	rs584368	$2x10^{-14}$
10	rs2596542	$4x10^{-13}$
11	rs1110446	$9x10^{-13}$
12	rs58489806	$3x10^{-12}$
13	rs6078460	$2x10^{-11}$
14	rs2523961	$6x10^{-11}$
15	rs7574865	$2x10^{-10}$

16	rs1110446	$3x10^{-10}$
17	rs455804	$5x10^{-10}$
18	rs58542926	$6x10^{-10}$
19	rs2523961	$6x10-^{10}$
20	rs8107030	$8x10^{-10}$
21	rs10272859	$9x10^{-10}$
22	rs190121281	$4x10^{-9}$
23	rs9275572	$6x10^{-9}$
24	rs2242652	$6x10^{-9}$
25	rs188273166	$1x10^{-8}$
26	rs708113	$1x10^{-8}$
27	rs2896019	$2x10^{-8}$
28	rs17047200	$3x10^{-8}$
29	rs541860626	$5x10^{-8}$

Table 2. Variants and risk alleles of liver cancer encoding prioritized SNPs

Variation and risk alleles	Variants near risk allele ($r^2 > 0.8$)	*p*-value	GENCODE	Type of allele
rs2294915	rs738409	$2x10^{-19}$	*PNPLA3*	missense
rs2896019	rs3761472	$2x10^{-8}$	*PNPLA3*	missense

2. Gene expression of *PNPLA3* across 10 human tissues

The results of *PNPLA3* gene expression across 10 human tissues comprise the most apparent functional consequences of genetic variation. Liver, sun-exposed skin (lower legs), non-sun-exposed skin (suprapubic), and adipose subcutaneous fibroblasts and cell cultures showed the highest *PNPLA3* gene expression in the 10 human tissues analyzed from GTEx (Figure 2). In addition, we have found that the SNP IDs rs2294915 and rs2896019 have similar gene expression variations in Sun-Exposed skin (lower legs). The exciting thing about these findings is that liver cancer patients often experience complaints that their skin appears yellow. Further results showed that the *PNPLA3* gene has high expression in suprapubic and underarm skin.

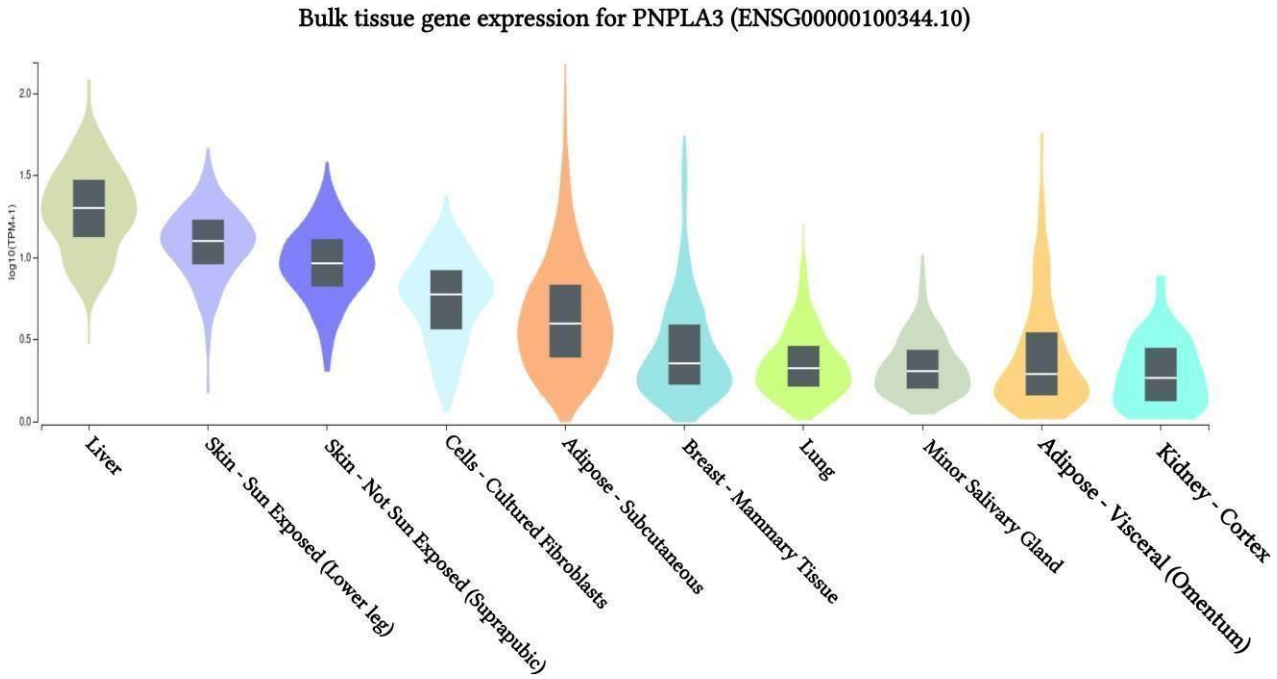


Figure 2. *PNPLA3* gene expression associated with liver cancer across human tissues based on GTEx Portal analysis

3. Correlation between Gene Expression of *PNPLA3* and eQTL

The result in a correlation between the Gene Expression of PNPLA3 and eQTL, to identify eQTLs associated with liver cancer gene expression, the GTEx database was used. We have identified minor alleles related to liver cancer, as presented in Table 3. Uniquely, several types of SNPs we found have high expression in skin tissue, namely rs2294915 and rs2896019. The CC type genotypes of rs2294915 and rs2896019 were associated with higher expression in suprapubic and underarm skin compared with the CT and TT type genotypes (Figure 3).

Table 3. Results of e-QTL in liver cancer from the GTEx portal database

SNP	Gencode ID (ENSG00000-)	Gene Symbol	*p*-value	Effect Size	Tissue	Expression Level
rs2294915	100344.10	*PNPLA3*	2.8×10^{-8}	-0.15	Skin - Sun Exposed (Lower leg)	CC>CT>TT
	100344.10	*PNPLA3*	5×10^{-8}	-0.50	Skin - Not Sun Exposed (Suprapubic)	CC>CT>TT
rs2896019	100344.10	*PNPLA3*	6.7×10^{-11}	-0.19	Skin - Sun Exposed (Lower leg)	CC>CT>TT
	100344.10	*PNPLA3*	2×10^{-9}	-0.22	Skin - Not Sun Exposed (Suprapubic)	CC>CT>TT

Source: Expression Quantitative Trait Loci (eQTL) obtained from the GTEx Portal.
Note : *PNPLA3*, patatin like phospoliphase 3; SNP, single nucleotide polymorphism; eQTL, expression quantitative trait loci; SNP, single nucleotide polymorphism.

The research results show that the genomic database can identify gene variations with the most potential in the pathogenesis of liver cancer. Liver cancer is characterized by eyes and skin that appear yellow (Fitrianti et al., 2022). According to Nessa et al (2017), the severity of the liver can be seen from the decreasing quality of the liver. The quality of the liver can be assessed from total bilirubin levels, serum albumin, and PT (partial thromboplastin time).

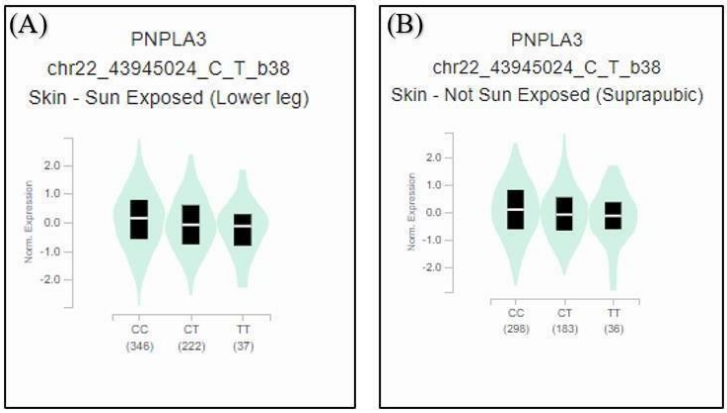


Figure 3. Patatin like phospolipase domain containing protein 3 (*PNPLA3*) gene expression for each genotype of the single nucleotide polymorphism (SNP): (A) rs2294915 and (B) rs2896019.

4. Allele frequencies of candidate variants in populations in different continents

The results of the research we have done, we have identified variants associated with liver cancer gene expression and carried out allele frequency analysis in various populations. As shown in Table 4, allele variant frequencies were evaluated in multiple people from Europe, America, East Asia, South Asia, and Africa. Allele frequencies across populations varied for each SNP, as depicted in Figure 4. Table 4 and Figure 4 show the gene expression levels at higher frequencies of the rs2294915 related allele (C) and the rs2896019 related allele (T). At the population frequency of the rs2294915 (C) allele, populations in Europe and South Asia were expressed at much higher levels than America, Africa, and East Asia.

Table 4. Analysis of allele frequencies for the *PNPLA3* gene from variant annotation (SNPnexus)

SNP	Gene	Location	Allele		Allele Frequency (N)				
			Ref	Eff*	AFR	AMR	EAS	EUR	SAS
rs2294915	*PNPLA3*	Missense	C	T	T: 0.163 (215)	T: 0.490 (340)	T: 0.365 (368)	T: 0.252 (254)	T: 0.246 (241)
rs2896019	*PNPLA3*	Missense	T	G	G: 0.158 (209)	G: 0.438 (304)	G: 0.364 (367)	G: 0.199 (200)	G: 0.236 (231)

Note : *PNPLA3*, patatin like phospolipase domain containting protein 3; SNP, single nucleotide polymorphism; Ref, Reference; Alt, Alternate; AFR, Africa; AMR, America; EAS, East Asia; EUR, Europe; SAS, Southeast Asia.

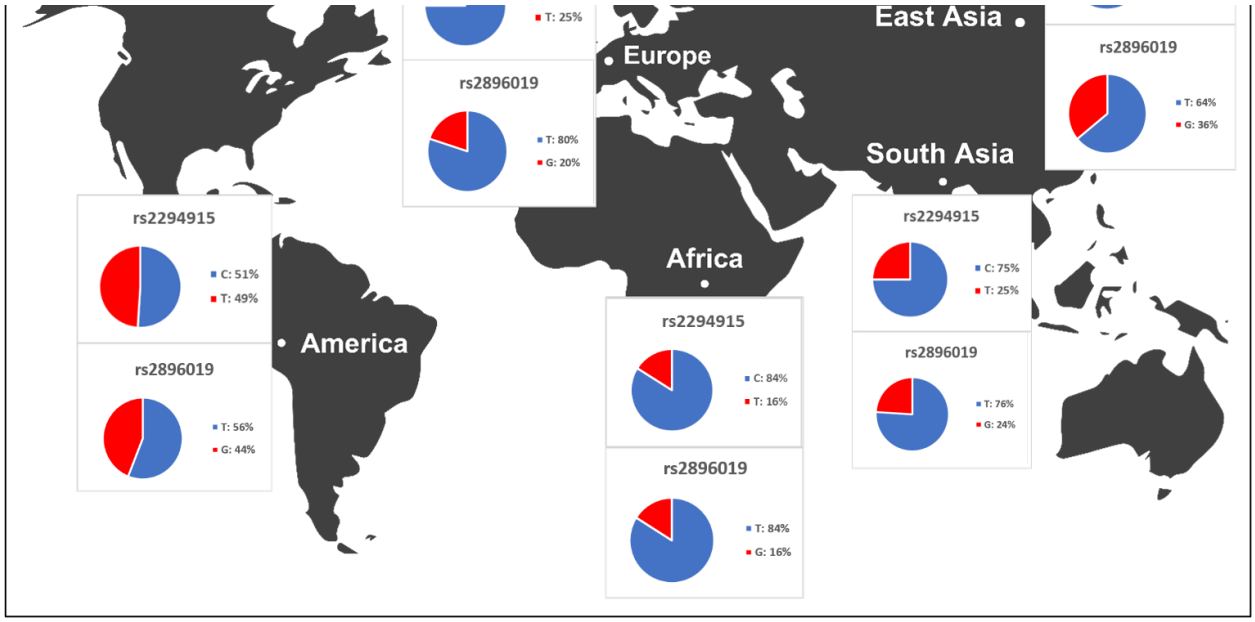


Figure 4. The results of the distribution of allele frequencies affecting *PNPLA3* across various populations. EUR, Europe; AFR, Africa; SAS, South Asia; EAS, East Asia.

Based on this finding, rs2294915 and rs2896019 are potentially related to the susceptibility to "liver cancer with effect size the highest score of -0.50 can be interpreted on the skin not exposed to sunlight (suprapubic)." According to Poggiali & Vercelli (2023), this condition is characterized by a disruption in the heme biosynthesis pathway due to reduced hepatic uroporphyrinogen decarboxylase (UROD) activity. The consequence of this phenomenon is the buildup of light-sensitive by-products, such as uroporphyrinogen, leading to the development of fragility and blistering of sun-exposed skin, as well as impairment of liver function.

The allele frequencies of the T and G alleles "rs2294915" and "rs2896019" in African populations were expressed at much lower levels compared to American, European, and Southeast Asian people. Overall, the allele frequencies of the "rs2294915" and "rs2896019" variant alleles suggest a contribution to the prevalence of the variants for gene expression of *PNPLA3*.

Across human populations, the frequency of the T allele at rs2294915 was associated with a high expression of *PNPLA3* in liver cancer, which is much lower in African populations (16%) compared to South Asians (25%), Europeans (25%), East Asians (37%) and America (49%). In contrast, the frequency of the C allele at rs2296019 was considerably higher in African (84%), European (80%), South Asian (76%), East Asian (64%), and American (56%) populations. Next, the association between allele frequency and the prevalence of the liver cancer of each continent was evaluated. Data on the prevalence of liver cancer in the continents were obtained from Li et al (2022) (Li et al., 2023). Herein, two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate of liver cancer in multiple continents (Africa, America, East

Asia, Europe, South Asia) based on the pearson's correlation analysis (*P-value*<0.011) (Figure 5). Populations with higher frequencies of variant alleles of these polymorphisms are thought to have a higher prevalence. We highlighted that these two variants (rs2294915 and rs2896019) were more frequent in the East Asian and Africa, which performed the higher aggressiveness of liver cancer in East Asian and African compared to America, Europe and South Asia. This study might give an insight that East Asian and African with carriers variants rs2294915 and rs2896019 might be more highly susceptible to suffer the liver cancer.

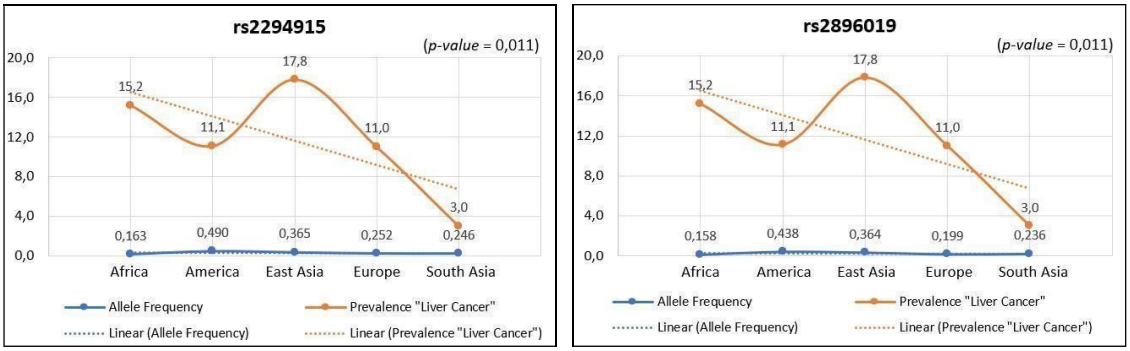


Figure 5. The association between allele frequency and the prevalence of liver cancer of each continent

Patients with liver cancer who have a history of alcohol addiction to an amount of ≥ 3 drinks per day can increase the risk of liver cancer by 16% in the general population; diabetics and people with central obesity also increase the risk of liver cancer by 2 times (McGlynn et al., 2021). The diagnosis of liver cancer in patients often involves using serological testing in conjunction with imaging techniques, which is considered the established approach for identifying liver carcinoma. Nevertheless, the sensitivity of the often-used serological test, specifically designed to detect alpha-fetoprotein (AFP), is at around 60%. Imaging modalities, including magnetic resonance imaging (MRI), computed tomography (CT), and ultrasonography (US), exhibit notable levels of sensitivity and specificity in the identification of liver cancer, particularly in individuals afflicted with liver cirrhosis (Huang et al., 2022).

Variant alleles (rs2294915 and rs2896019) are associated with liver cancer. Accordingly, populations from the continents of Africa, America, East Asia, Europe, and South Asia show associated *PNPLA3* expression, resulting in a higher susceptibility to liver cancer. Identifying unique and pathogenic gene variations for a disease is very interesting for both research and clinical validation. These variants not only offer insights into disease susceptibility but also serve as potential diagnostic and prognostic biomarkers (Irham et al., 2020). Additionally, they can facilitate the identification of drug target candidates, a concept known as genomic-driven drug repurposing (Afief et al., 2022). We anticipate that the discovery of candidate gene variations in *PNPLA3* will pave the way for successful clinical validation, potentially establishing this as a promising diagnostic and prognostic biomarker for liver cancer.

It is important to acknowledge that the genetic variants identified in this study as potentially pathogenic are based on preliminary investigations using genomic and bioinformatics databases. However, these findings offer crucial insights for future researchers intending to validate these genetic variants in liver cancer patients. We strongly recommend future research incorporate

further functional annotations, which would further aid in prioritizing these pathogenic genetic variants.

Conclusion

This study identifies genetic variants influencing "liver cancer" and reveals the significance of the *PNPLA3* gene in liver tissue.consequently, these population groups exhibit varying susceptibilities to liver cancer based on associated *PNPLA3* expression levels. . The observed variations in allele frequencies of the two identified variants, rs2294915, and rs2896019, across populations from Africa, America, East Asia, Europe, and South Asia, significantly impact *PNPLA3* gene expression. Our study also demonstrated that these two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate. Positive association of prevalence rates were underlined more frequent in East Asian and African population. The higher frequency of the variants allele of these polymorphisms in population, the higher the estimated prevalence rates. The variants investigated in this study were likely to predispose to liver cancer and could play a role in its progression and aggressiveness. These discoveries highlight the critical relevance of understanding genomic variations in precision medicine and designing screening strategies for liver cancer across diverse populations on different continents.

Reference

Afief, A. R., Irham, L. M., Adikusuma, W., Perwitasari, D. A., Brahmadhi, A., & Cheung, R. (2022). Integration of genomic variants and bioinformatic-based approach to drive drug repurposing for multiple sclerosis . *Biochemistry and Biophysics Reports , 101337*.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Chen, Z., Boehnke, M., Wen, X., & Mukherjee, B. (2021). Revisiting the genome-wide significance threshold for common variant GWAS. *G3: Genes, Genomes, Genetics*, *11*(2). https://doi.org/10.1093/g3journal/jkaa056

Fitrianti, F., Shatri, H., Faisal, E., Putranto, R., Agung, R. A., Nainggolan, L., Kalista, K. F., & Irawan, C. (2022). Pain Management in Advanced Hepatocellular Carcinoma: A Case Study. *Jurnal Penyakit Dalam Indonesia*, *9*(1), 53–58.

Huang, L., Sun, H., Sun, L., Shi, K., Chen, Y., Ren, X., Ge, Y., Liu, X., Knoll, W., Zhang, Q., & Wang, Y. (2022). Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning. *Nature Communications*, *14*(48), 1–14.

Irham, L. M., Chou, W. H., Calkins, M. J., Adikusuma, W., Hsieh, S. L., & Chang, W. C. (2020). Genetic variants that influence SARS-CoV-2 receptor TMPRSS2 expression among population cohorts from multiple continents. *Biochemical and Biophysical Research Communications*, *529*(2), 263–269. https://doi.org/10.1016/j.bbrc.2020.05.179

Kementerian Kesehatan Republik Indonesia. (2019). *Hari Kanker Sedunia 2019*. https://www.kemkes.go.id/article/view/19020100003/harikanker-sedunia-2019.html [Diakses pada 21 Mei 2023].

Li Q, Cao M, Lei L, Yang F, Li H, Yan X, He S, Zhang S, Teng Y, Xia C, Chen W. (2022). Burden of liver cancer: From epidemiology to prevention. *Chin J Cancer Res*, 34(6):554-566. https://doi.org/10.21147/j.issn.1000-9604. 2022.06.02

Ma'ruf, M., Fadli, J. C., Mahendra, M. R., Irham, L. M., Sulistyani, N., Adikusuma, W., Chong, R., & Septama, A. W. (2023). A bioinformatic approach to identify pathogenic variants for Stevens Johnson syndrome. *Genomics & Informatics*, *21*(2), 1–9.

McGlynn, K. A., Petrick, J. L., & El-Serag, H. B. (2021). Epidemiology of Hepatocellular Carcinoma. In *Hepatology* (Vol. 73, Issue S1, pp. 4–13). John Wiley and Sons Inc. https://doi.org/10.1002/hep.31288

Mittal, S., & El-Serag, H. B. (2013). *Epidemiology of Hepatocellular Carcinoma Consider the Population*. www.jcge.com

Nessa, M., Yusuf, F., & Mudatsir. (2017). Hubungan Child Pugh Score dengan Kualitas Hidup pada Pasien Sirosis Hati di Bagian Penyakit Dalam RSUDZA Banda Aceh. *Jurnal Ilmiah Mahasiswa Kedokteran Biomedis*, *2*(4), 24–34.

Poggiali, E., & Vercelli, A. (2023). Sun, alcohol, and skin lesions. *Emergency Care Journal*, *19*(11294).

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), 71–81. https://doi.org/10.12928/mf.v19i2.23706

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Villanueva, A. (2019). Hepatocellular Carcinoma. *New England Journal of Medicine*, *380*(15), 1450–1462. https://doi.org/10.1056/NEJMra1713263

Wage, H., Schulze, K., Felden, J. V., Calderaro, J., & Reig, M. (2021). Rare variants of primary liver cancer: Fibrolamellar, combined, and sarcomatoid hepatocellular carcinomas. *J. Med Genet*, *64*(11).

Waller, L. P., Deshpande, V., & Pyrsopoulos, N. (2015). Hepatocellular carcinoma: A comprehensive review. In *World Journal of Hepatology* (Vol. 7, Issue 26, pp. 2648–2663). Baishideng Publishing Group Co. https://doi.org/10.4254/wjh.v7.i26.2648

Wang, J., Conti, D. V, Bogumil, D., Sheng, X., Noureddin, M., Wilkens, L. R., Le Marchand, L., Rosen, H. R., Haiman, C. A., & Setiawan, V. W. (2021). Association of Genetic Risk Score With NAFLD in An Ethnically Diverse Cohort. *Hepatology CommuniCations*, *5*(10), 1689–1703. https://doi.org/10.1002/hep4.1751/suppinfo

Pada tanggal 29 November 2023 hasil revisian yang kedua telah diterima oleh editor jurnal. Dan pada tanggal 30 November 2023 artikel telah diterima / accepted oleh editor.

Editorial Comment

Invited date	Nov 29, 2023
Complete Review Date	Nov 30, 2023
Recommendation	Accept as it is
Comments to the Author	None (or N/A)
Files	None (or N/A)

Close

Pada tanggal 18 Desember 2023 mendapatkan proofs artikel sebelum di terbitkan

GENOMICS & INFORMATICS
OPEN ACCESS
eISSN : 2234-0742

Article Title: A genomic and bioinformatic-based approach to identify genetic variants for liver cancer across multiple continents
DOI: 10.5808/gi.23067

Dear Dr. Lalu Muhammad Irham,

Here are the proofs of your article.

- You can submit your corrections **online** or **e-mail**.
- For online submission please insert your corrections in the online correction form.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- Check the questions that may have arisen during copy editing and insert your answers/corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable.
- The publication of inaccurate data such as dosages and units can have serious consequences.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.

Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.

* We would appreciate your correction **within 48 hours (including holidays)** of the publishing schedule.

Should you encounter difficulties with the proofs, please contact me.
Thank you very much.

Sincerely yours,

M2PI
E-mail: support@m2-pi.com

Comments to Authors:

Please check the author's proof.

Author Query Form

Query	Details Required	Author's Response
Q1	1) 1Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia 2Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram , Indonesia 9Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong Science Center, Cibinong , Indonesia =>Please provide postal code. 2) 10Mataram Training Health Center, Indonesia Ministry of Health, Indonesia =>Please provide city name and postal code. 3) Please provide ORCID for all authors.	1)Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta 55164, Indonesia 2) Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram 83127, Indonesia 9) Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong, 16911, Indonesia 10) Mataram Training Health Center, Indonesia Ministry of Health, Mataram, 83237, Indonesia ORCID : Muhammad Ma'ruf: https://orcid.org/0009-0006-5088-8089 Lalu Muhammad Irham: https://orcid.org/0000-0002-0091-4887 Wirawan Adikusuma: https://orcid.org/0000-0001-9165-690X
Q2	with hepatitis B virus and hepatitis C virus =>Following standard guidance on abbreviation usage, the acronyms were deleted here because they were not repeated within the text itself. However, if you prefer to re-include them, please note that the standard abbreviation for hepatitis C is HCV, not HBC as in the original text here.	Yes, We Follow the standard guidance HCV. Thank you for your feedback and guidance on the use of acronyms for hepatitis B virus (HBV) and hepatitis C virus (HCV) in the manuscript.
Q3	DNA =>As a minor note, biomedical style guides generally state that DNA is not necessary to define using the full term.	Thank you for your insightful feedback regarding the use of the term "DNA" in the manuscript. We appreciate your guidance on the conventions within biomedical style guides that suggest the full term for DNA might not be necessary
Q4	p-values were obtained. =>This wording was simplified because a 95% CI is a distinct statistical concept from a "95% level of certainty" in the sense of using a P<0.05 cutoff for statistical significance; furthermore, since the next sentences state that a P<0.05 cutoff was used, this information does not need to be repeated.	Thank you for your thoughtful insights regarding the clarification of statistical concepts in the manuscript, specifically regarding the description of p-values and confidence intervals (CI).
Q5	Authors' Contribution Conceptualization: Data curation: Formal analysis: Funding acquisition: Methodology:	Conceptualization: MM, LMI, WA. Data curation: MM, LMI Formal analysis: MM, LMI, WA Methodology: MM, LMI, WA. Writing – original draft: MM, LMI Writing – review & editing: MM, LMI, WA, BDP, MAS, SK, RC, MM, LMHS.

	Writing - original draft: Writing - review & editing: =>Please refer to the following format and provide.	
Q6	References are to be numbered in the order of citation within the article in brackets. Please check correction.	Yes, it is correct
Q7	Please provide reference.	Li Q, Cao M, Lei L, Yang F, Li H, Ya n X, He S, Zhang S, Teng Y, Xia C, Che n W. (2022). Burden of liver cancer: From epidemiology to prevention. Chin J
Q8	If possible, please change the corresponding text as follows: SNPs associated with liver cancer SNPs encoded by the genes Gene expression of SNPs associated with liver cancer Allele frequencies on multiple continents	These terms were correct

G&I
Genomics & Informatics

Original article

eISSN 2234-0742
Genomics Inform [Epub ahead of print]
https://doi.org/10.5808/gi.23067

Received: August 14, 2023
Revised: November 29, 2023
Accepted: November 30, 2023

*Corresponding author:
E-mail: lalu.irham@pharm.uad.ac.id

A genomic and bioinformatic-based approach to identify genetic variants for liver cancer across multiple continents

Muhammad Ma'ruf[1], Lalu Muhammad Irham[1*], Wirawan Adikusuma[2], Made Ary Sarasmita[3,4], Sabiah Khairi[5], Barkah Djaka Bar,kah Djaka Purwanto[6,7], Rockie Chong[8,10], Maulida Mazaya , Lalu Muhammad Harmasier Siswanto

[1]Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2]Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram, Indonesia
[3]Department of Clinical Pharmacy, College of Pharmacy, Taipei Medical University, Taipei 11031, Taiwan
[4]Pharmacy Study Program, Faculty of Science and Mathematics, Udayana University, Bali, Indonesia
[5]School of Nursing, College of Nursing, Taipei Medical University, Taipei 11031, Taiwan
[6]Faculty of Medicine, Universitas Ahmad Dahlan, Yogyakarta 55191, Indonesia
7PKU Muhammadiyah Bantul Hospital, Bantul, Yogyakarta 55711, Indonesia
[8]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA
[9]Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong Science Center, Cibinong, Indonesia
[10]Mataram Training Health Center, Indonesia Ministry of Health, Indonesia

Liver cancer is the fourth leading cause of death worldwide. Well-known risk factors include hepatitis B virus and hepatitis C virus, along with exposure to aflatoxins, excessive alcohol consumption, obesity, and type 2 diabetes. Genomic variants play a crucial role in mediating the associations between these risk factors and liver cancer. However, the specific variants involved in this process remain under-explored. This study utilized a bioinformatics approach to identify genetic variants associated with liver cancer from various continents. Single-nucleotide polymorphisms associated with liver cancer were retrieved from the genome-wide association studies catalog. Prioritization was then performed using functional annotation with HaploReg v4.1 and the Ensembl database. The prevalence and allele frequencies of each variant were evaluated using Pearson correlation coefficients. Two variants, rs2294915 and rs2896019, encoded by the *PNPLA3* gene, were found to be highly expressed in the liver tissue, as well as in the skin, cell-cultured fibroblasts, and adipose-subcutaneous tissue, all of which contribute to the risk of liver cancer. We further found that these two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate. Positive associations with the prevalence rate were more frequent in East Asian and African populations. We highlight the utility of this population-specific *PNPLA3* genetic variant for genetic association studies and for the early prognosis and treatment of liver cancer. This study highlights the potential of integrating genomic databases with bioinformatic analysis to identify genetic variations involved in the pathogenesis of liver cancer. The genetic variants investigated in this study are likely to predispose to liver cancer and could affect its progression and aggressiveness. We recommend future research prioritizing the validation of these variations in clinical settings.

Keywords: bioinformatics, genomic variants, liver neoplasms

2023-12-19 12:43:36
Here the name of Barkah Djaka was repeated two times. Please revise it to be correct one Barkah Djaka Purwanto

2023-12-19 13:01:43
Please revise the keywords as follows: bioinformatics, genomic variants, liver cancer, PNPLA3, rs2294915, rs2896019

Introduction

Patients often report symptoms such as fatigue, pain, diarrhea, skin abnormalities, and decreased appetite, all of which can ad-

Fig. 1. Analytical methodology for integrated bioinformatic, database, and genomic analysis of genetic variations that affect liver cancer. The figure was created with BioRender.com under agreement number "FM25OO073C".

https://doi.org/10.5808/gi.23067

ciated SNPs were obtained from the GWAS Catalog database (http://www.ebi.ac.uk/gwas; accessed on 15-02-2023). Subsetwo variants with missense mutations (rs2294915 and rs2896019) that encode the *PNPLA3* gene as biological risk SNPs for liver can-

rs2896019	rs3761472	2×10^{-8}	*PNPLA3*	Missense

SNP, single-nucleotide polymorphism

cellular carcinoma (HCC), intrahepatic cholangiocarcinoma, and less common varieties like mixed hepatocellular cholangiocarcinolated to these findings. Further analysis indicated that the *PNPLA3* gene is also highly expressed in suprapubicandunderarmskin.

Source: expression quantitative trait loci (eQTLs) obtained from the GTEx Portal [14].
SNP, single-nucleotide polymorphism.

https://doi.org/10.5808/gi.23067

ing of the eyes and skin [15]. Nessa et al. [16] note that the severity of liver disease can be gauged by the declining quality of liver func-

Based on these findings, rs2294915 and rs2896019 may be associated with an increased susceptibility to liver cancer, with the high-

rs2896019	PNPLA3	Missense	T	G	G: 0.158 (209)	G: 0.458 (304)	G: 0.364 (307)	G: 0.199 (200)	G: 0.256 (251)

SNP, single-nucleotide polymorphism; Ref, reference; Alt, alternate; AFR, Africa; AMR, America; EAS, East Asia; EUR, Europe; SAS, Southeast Asia.

allele at rs2296019 is considerably higher in African (84%), European (80%), South Asian (76%), East Asian (64%), and American

[12,18]. In this context, two SNPs (rs2294915 and rs2896019) were found to be positively correlated with the prevalence rate of

https://doi.org/10.5808/gi.23067

liver cancer across multiple continents (Africa, America, East Asia, Europe, South Asia), as determined by Pearson's correlation analysis ($p = 0.011$) (Fig. 5). Populations with higher frequencies of variant alleles of these polymorphisms are thought to have a higher prevalence of liver cancer. We highlighted that these two variants (rs2294915 and rs2896019) are more frequent in East Asian and African populations, which exhibit higher aggressiveness of liver cancer compared to America, Europe, and South Asia. This study suggests that individuals in East Asian and African populations carrying the variant alleles rs2294915 and rs2896019 may be more susceptible to liver cancer.

Patients with liver cancer who also have a history of alcohol abuse, consuming ≥3 drinks per day, have a 16% increased risk of developing liver cancer compared to the general population. Additionally, individuals with diabetes and those with central obesity are at twice the risk of developing liver cancer [1]. The diagnosis of liver cancer typically involves serological testing combined with imaging techniques, which is the standard approach for detecting liver carcinoma. However, the sensitivity of the commonly used serological test, which is designed to detect alpha-fetoprotein, is only about 60%. Imaging modalities such as magnetic resonance imaging, computed tomography, and ultrasonography demonstrate high levels of sensitivity and specificity in detecting liver cancer, especially in patients with liver cirrhosis [19].

Variant alleles (rs2294915 and rs2896019) are associated with liver cancer. Populations from Africa, America, East Asia, Europe, and South Asia exhibit associated *PNPLA3* expression, which leads to an increased susceptibility to liver cancer. The identification of unique and pathogenic gene variations for a disease is of great interest for both research and clinical validation. These variants provide insights into disease susceptibility and also act as potential diagnostic and prognostic biomarkers [20]. Furthermore, they can aid in the identification of drug target candidates, an approach referred to as genomic-driven drug repurposing [21]. We expect that the discovery of candidate gene variations in *PNPLA3* will facilitate successful clinical validation, potentially establishing it as a promising diagnosticandprognosticbiomarkerforlivercancer.

It is important to acknowledge that the genetic variants identified in this study as potentially pathogenic are based on preliminary investigations using genomic and bioinformatics databases. While these findings provide crucial insights for future researchers aiming to validate these genetic variants in liver cancer patients, it is important to proceed with caution. We strongly recommend that future research includes additional functional annotations to aid in theprioritization of thesepathogenicgeneticvariants.

This study identified genetic variants that influence liver cancer, highlighting the importance of the *PNPLA3* gene in liver tissue. Consequently, these population groups exhibit varying susceptibilities to liver cancer based on the associated *PNPLA3* expression levels. The observed variations in allele frequencies of the two identified variants, rs2294915 and rs2896019, across populations from Africa, America, East Asia, Europe, and South Asia, significantly impact *PNPLA3* gene expression. Our study also demonstrated that these two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate. The positive association of prevalence rates was more frequently observed in East Asian and African populations. The higher the frequency of the variant alleles of these polymorphisms in a population, the higher the estimated prevalence rates. The variants investigated in this study are likely to predispose individuals to liver cancer and could play a role in its progression and aggressiveness. These findings highlight the critical importance of understanding genomic variations for precision medicine and for designing targeted screening strategies for liver cancer acrossdiversepopulations on differentcontinents.

Authors' Contribution

Conceptualization:
Datacuration:
Formal analysis:
Funding acquisition:
Methodology:
Writing - original draft:
Writing - review & editing:

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

References

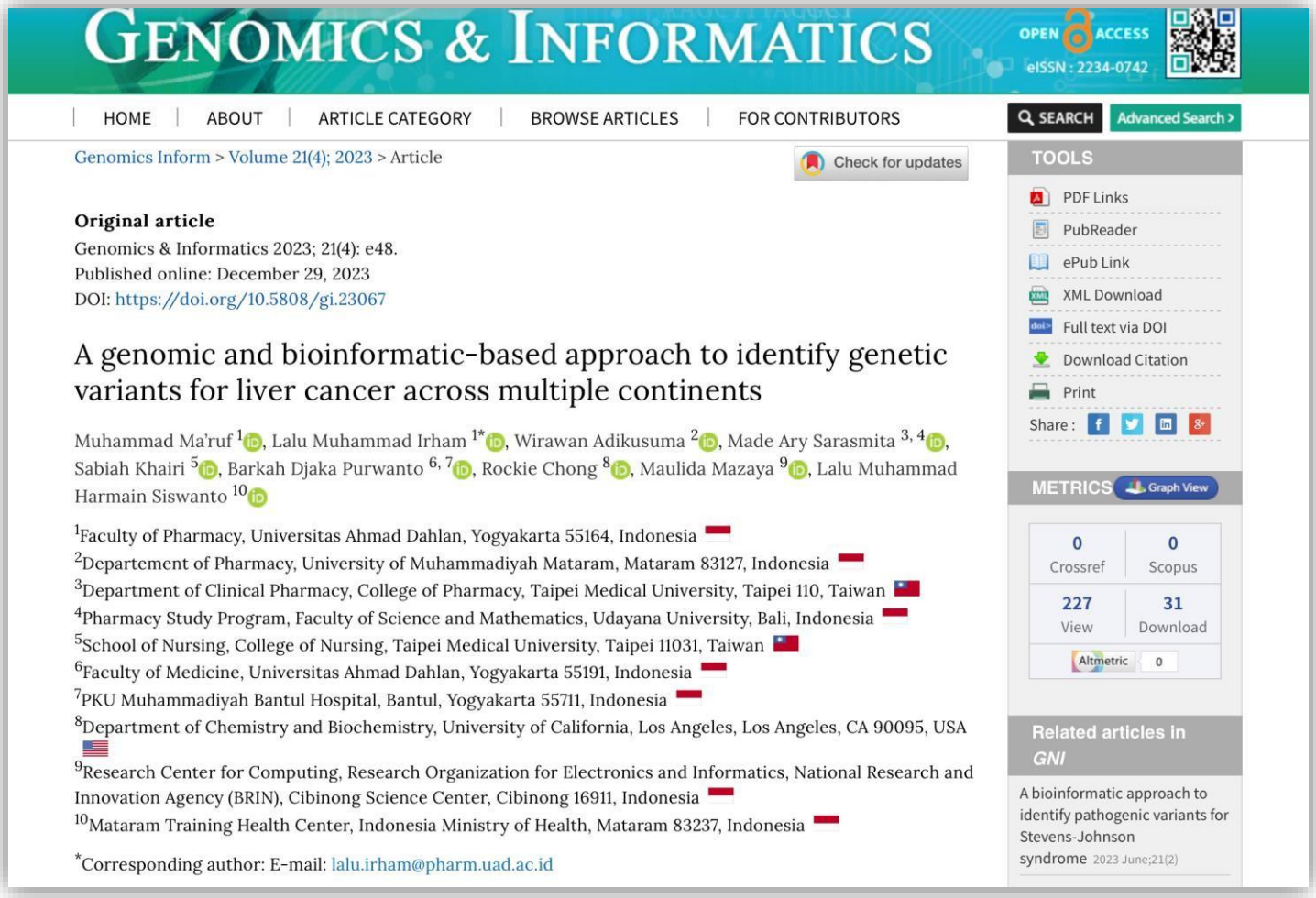
1. McGlynn KA, Petrick JL, El-Serag HB. Epidemiology of hepatocellularcarcinoma.Hepatology2021;73Suppl1:4-13.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CACancerJClin2018;68:394-424.
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.CACancerJ Clin2021;71:209-249.
4. Hari Kanker Sedunia 2019. Jakarta Selatan: Kementerian Kese

hatan Republik Indonesia, 2019. Accessed 2023 May 21. Available from: https://www.kemkes.go.id/article/view/19020100003/harikanker-sedunia-2019.html.
5. Villanueva A. Hepatocellular carcinoma. N Engl J Med 2019;380:1450-1462.
6. Mittal S, El-Serag HB. Epidemiology of hepatocellular carcinoma: consider the population. J Clin Gastroenterol 2013;47 Suppl:S2-S6.
7. Waller LP, Deshpande V, Pyrsopoulos N. Hepatocellular carcinoma: a comprehensive review. World J Hepatol 2015;7:2648-2663.
8. Wang J, Conti DV, Bogumil D, Sheng X, Noureddin M, Wilkens LR, et al. Association of genetic risk score with NAFLD in an ethnicallydiversecohort.HepatolCommun2021;5:1689-1703.
9. Ma'ruf M, Fadli JC, Mahendra MR, Irham LM, Sulistyani N, Adikusuma W, et al. A bioinformatic approach to identify pathogenic variants for Stevens-Johnson syndrome. Genomics Inform 2023;21:e26.
10. Puspitaningrum AN, Perwitasari DA, Adikusuma W, Djalilah GN, Dania H, Maliza R, et al. Integration of genomic databases and bioinformatic approach to identify genomic variants for Sjogren's syndrome on multiple continents. J Pharm Sci 2022;19:71-81.
11. Chen Z, Boehnke M, Wen X, Mukherjee B. Revisiting the genome-wide significance threshold for common variant GWAS. G3(Bethesda)2021;11:jkaa056.
12. Li Q, Cao M, Lei L, Yang F, Li H, Yan X, et al. Burden of liver cancer: From epidemiology to prevention. Chin J Cancer Res 2022;34:554-566.
13. Wege H, Schulze K, von Felden J, Calderaro J, Reig M, Rare Liver Tumors Working Group of the European Reference Network on Hepatological Diseases (ERN RARE-LIVER). Rare variants of primary liver cancer: fibrolamellar, combined, and sarcomatoid hepatocellular carcinomas. Eur JMed Genet 2021;64:104313.
14. GTEx Portal database. Cambridge, MA: Broad Institute of MIT and Harvard, 2023. Accessed 2023 Feb 16. Available from: http://www.gtexportal.org/home/.
15. Fitrianti F, Shatri H, Faisal E, Putranto R, Agung RA, Nainggolan L, Kalista KF, et al. Pain management in advanced hepatocellular carcinoma:acasestudy.JPenyakitDalamIndones2022;9:6.
16. Nessa ML, Yusuf F, Mudatsir M. Relationship of Child Pugh score and quality of life of liver cirrhosis at Internal Medicine RSUDZA Banda Aceh. J Ilmiah Mahasiswa Kedokteran Biomed 2017;2:24-34.
17. Poggiali E, Vercelli A. Sun, alcohol, and skin lesions. Emerg Care J 2023;19:11294.
18. Li et al. (2023).
19. Huang [illegible] Sun H, Sun L, Shi K, Chen Y, Ren X, et al. Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning. Commun Biol 2023;6:1148.
20. Irham LM, Chou WH, Adikusuma W, Hsieh SL, Chang WC. Genetic variants that influence SARS-CoV-2 receptor TMPRSS2 expression among population cohorts from multiple continents. Biochem Biophys Res Commun 2021;538:262-269.
21. Afief AR, Irham LM, Adikusuma W, Perwitasari DA, Brahmadhi A, Chong R. Integration of genomic variants and bioinformatic-based approach to drive drug repurposing for multiple sclerosis.BiochemBiophysRep2022;32:101337.

user
2023-12-19 16:30:33
Li Q, Cao M, Lei L, Yang F, Li H, Yan X, He S, Zhang S, Teng Y, Xia C, Chen W (2022). Burden of liver cancer: From epidemiology to prevention. Chin J Cancer Res 34(6):554-566. https://doi.org/10.21147/j.issn.1000-9604.2022.06.02

Pada tanggal 31 Desember 2023 artikel telah terbit di

Genomics & Informatics Journal



GENOMICS & INFORMATICS
OPEN ACCESS
eISSN : 2234-0742

Original article
Genomics & Informatics 2023; 21(4): e48.
Published online: December 29, 2023
DOI: https://doi.org/10.5808/gi.23067

A genomic and bioinformatic-based approach to identify genetic variants for liver cancer across multiple continents

Muhammad Ma'ruf [1], Lalu Muhammad Irham [1*], Wirawan Adikusuma [2], Made Ary Sarasmita [3,4], Sabiah Khairi [5], Barkah Djaka Purwanto [6,7], Rockie Chong [8], Maulida Mazaya [9], Lalu Muhammad Harmain Siswanto [10]

[1]Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta 55164, Indonesia
[2]Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram 83127, Indonesia
[3]Department of Clinical Pharmacy, College of Pharmacy, Taipei Medical University, Taipei 110, Taiwan
[4]Pharmacy Study Program, Faculty of Science and Mathematics, Udayana University, Bali, Indonesia
[5]School of Nursing, College of Nursing, Taipei Medical University, Taipei 11031, Taiwan
[6]Faculty of Medicine, Universitas Ahmad Dahlan, Yogyakarta 55191, Indonesia
[7]PKU Muhammadiyah Bantul Hospital, Bantul, Yogyakarta 55711, Indonesia
[8]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA
[9]Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong Science Center, Cibinong 16911, Indonesia
[10]Mataram Training Health Center, Indonesia Ministry of Health, Mataram 83237, Indonesia

*Corresponding author: E-mail: lalu.irham@pharm.uad.ac.id

G&I
Genomics & Informatics

Original article

eISSN 2234-0742
Genomics Inform 2023;21(4):e48
https://doi.org/10.5808/gi.23067

Received: August 14, 2023
Revised: November 29, 2023
Accepted: November 30, 2023

*Corresponding author:
E-mail: lalu.irham@pharm.uad.ac.id

A genomic and bioinformatic-based approach to identify genetic variants for liver cancer across multiple continents

Muhammad Ma'ruf[1], Lalu Muhammad Irham[1]*, Wirawan Adikusuma[2], Made Ary Sarasmita[3,4], Sabiah Khairi[5], Barkah Djaka Purwanto[6,7], Rockie Chong[8], Maulida Mazaya[9], Lalu Muhammad Harmain Siswanto[10]

[1]Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta 55164, Indonesia
[2]Departement of Pharmacy, University of Muhammadiyah Mataram, Mataram 83127, Indonesia
[3]Department of Clinical Pharmacy, College of Pharmacy, Taipei Medical University, Taipei 110, Taiwan
[4]Pharmacy Study Program, Faculty of Science and Mathematics, Udayana University, Bali, Indonesia
[5]School of Nursing, College of Nursing, Taipei Medical University, Taipei 11031, Taiwan
[6]Faculty of Medicine, Universitas Ahmad Dahlan, Yogyakarta 55191, Indonesia
[7]PKU Muhammadiyah Bantul Hospital, Bantul, Yogyakarta 55711, Indonesia
[8]Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA
[9]Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong Science Center, Cibinong 16911, Indonesia
[10]Mataram Training Health Center, Indonesia Ministry of Health, Mataram 83237, Indonesia

Liver cancer is the fourth leading cause of death worldwide. Well-known risk factors include hepatitis B virus and hepatitis C virus, along with exposure to aflatoxins, excessive alcohol consumption, obesity, and type 2 diabetes. Genomic variants play a crucial role in mediating the associations between these risk factors and liver cancer. However, the specific variants involved in this process remain under-explored. This study utilized a bioinformatics approach to identify genetic variants associated with liver cancer from various continents. Single-nucleotide polymorphisms associated with liver cancer were retrieved from the genome-wide association studies catalog. Prioritization was then performed using functional annotation with HaploReg v4.1 and the Ensembl database. The prevalence and allele frequencies of each variant were evaluated using Pearson correlation coefficients. Two variants, rs2294915 and rs2896019, encoded by the *PNPLA3* gene, were found to be highly expressed in the liver tissue, as well as in the skin, cell-cultured fibroblasts, and adipose-subcutaneous tissue, all of which contribute to the risk of liver cancer. We further found that these two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate. Positive associations with the prevalence rate were more frequent in East Asian and African populations. We highlight the utility of this population-specific *PNPLA3* genetic variant for genetic association studies and for the early prognosis and treatment of liver cancer. This study highlights the potential of integrating genomic databases with bioinformatic analysis to identify genetic variations involved in the pathogenesis of liver cancer. The genetic variants investigated in this study are likely to predispose to liver cancer and could affect its progression and aggressiveness. We recommend future research prioritizing the validation of these variations in clinical settings.

Keywords: bioinformatics, genomic variants, liver cancer, PNPLA3, rs2294915, rs2896019

Introduction

Liver cancer, a type of carcinoma, has the highest mortality rate in the world each year [1]. In 2018, there were 841,000 new cases of liver cancer, and the death toll reached 782,000 [2]. The average incidence of liver cancer and the associated mortality rate can be two to three times higher in men, particularly in certain regions of the world. According to the Global Cancer Statistics (GLOBOCAN) in 2020, liver cancer was ranked as the third most deadly cancer, responsible for 8.3% of all cancer-related deaths. In that year, there were 905,000 new cases of liver cancer, with a mortality rate of 830,000 [3]. In Indonesia, liver cancer is the second most common cancer among men, with an incidence rate of 12.4 per 100,000 of the population and an average mortality rate of 7.6 per 100,000 [4]. Factors that contribute to liver cancer include chronic infection with hepatitis B virus and hepatitis C virus, exposure to aflatoxin contamination, alcohol consumption, a history of obesity, type 2 diabetes, and smoking addiction [2]. Villanueva [5] notes that additional risk factors may exacerbate the incidence of liver cancer, including an unhealthy lifestyle, geographic conditions, gender, age, family history of the disease, and the extent of liver damage. Liver cancer is also prevalent in regions with high rates of hepatitis B infection. In these areas, the disease often manifests at a younger age, partly because hepatitis B can be transmitted vertically from mother to child during childbirth [6].

Patients often report symptoms such as fatigue, pain, diarrhea, skin abnormalities, and decreased appetite, all of which can adversely affect their quality of life [7]. Consequently, the detection of disease symptoms in liver cancer can involve examining DNA. Variations in genes may be linked to the progression and pathogenesis of diseases, including liver cancer. The genome-wide association studies (GWAS) Catalog is a resource that employs a bioinformatics approach to document genetic variations. This database contains search results for single-nucleotide polymorphisms (SNPs) and has identified several variants associated with liver fat content, circulating liver enzymes, and the development of non-alcoholic fatty liver disease, as well as genetic markers useful in predicting disease disorders [8].

Genetic identification studies in humans aim to identify inherited genetic risk factors for various conditions, including liver cancer. This study used the GWAS catalog database to map genes from genetic variations across several populations that play an essential role in the pathogenesis of liver cancer. The most significant gene variations based on their function in protein changes were further verified.

Methods

In this study, we adopted the method used by Ma'ruf et al. [9] and Puspitaningrum et al. [10], as depicted in Fig. 1. Liver cancer-asso-

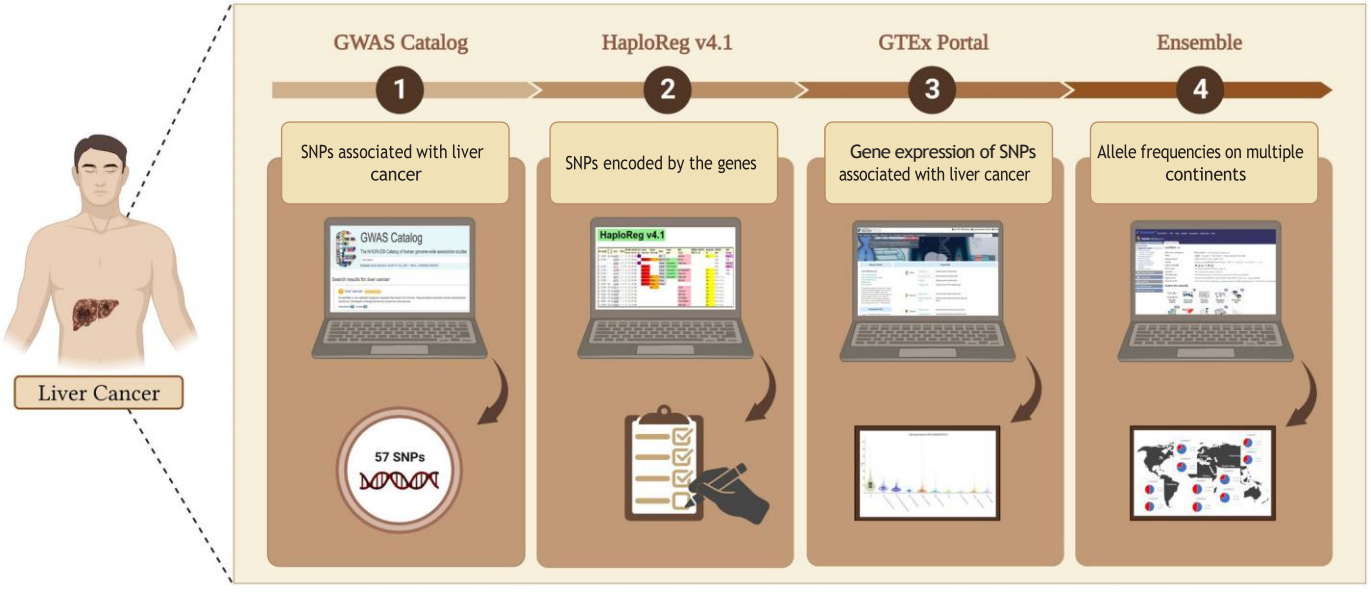


Fig. 1. Analytical methodology for integrated bioinformatic, database, and genomic analysis of genetic variations that affect liver cancer. The figure was created with BioRender.com under agreement number "FM25OO073C". SNP, single-nucleotide polymorphism; GWAS, genome-wide association study.

https://doi.org/10.5808/gi.23067

ciated SNPs were obtained from the GWAS Catalog database (http://www.ebi.ac.uk/gwas; accessed on 15-02-2023). Subsequently, we performed further analysis using HaploReg (version 4.1) applying a $p < 10^{-8}$ to account for multiple tests in the GWAS catalog. This threshold is commonly used to identify associations between common genetic variants and traits with adjacent gene expression [11]. Furthermore, to evaluate the relationships between various genetic variants and gene expression profiles, we conducted an analysis of expression quantitative trait loci (eQTLs) with data sourced from the GTEx Portal database (http://www.gtexportal.org/home/; accessed on 16 Feb 2023), considering gene expression across various tissues in humans. Additionally, we confirmed the identified variants using the Ensembl Genome Browser (https://www.ensembl.org/index.html; accessed on 17 Feb 2023). Our study considered allele frequencies in populations from Europe, Africa, America, East Asia, and Southeast Asia. To explore the functionalities of different gene variants, we performed evaluations using the SNP nexus database (https://www.snp-nexus.org; accessed on 20 Feb 2023). Furthermore, epidemiological and genomic data on the prevalence of liver cancer rates were obtained from Li et al. [12]. The prevalence rates and allele frequencies of the variants in multiple continents were evaluated using IBM SPSS Statistics 25.0 (IBM Corp., Armonk, NY, USA) with the Pearson correlation test. After the procedure was evaluated, the p-values were obtained. All plots were created using line charts. A $p < 0.05$ was considered statistically significant in the current study.

Results and Discussion

Identification of genomic variants of liver cancer

This study identified SNPs associated with liver cancer from the GWAS catalog. Of these SNPs, 29 were further confirmed through SNP genotyping, as shown in Table 1. Subsequently, HaploReg version 4.1 was utilized, applying a p-value threshold of $<10^{-8}$ based on the number of SNPs obtained. The findings presented in Table 2 indicate an increased risk associated with two genes for liver cancer. The study also analyzed tissue expression impacting liver cancer, with a focus on missense variants of *PNPLA3* (patatin-like phospholipase domain-containing 3).

Through our integrative bioinformatics approach, we prioritized two variants with missense mutations (rs2294915 and rs2896019) that encode the *PNPLA3* gene as biological risk SNPs for liver cancer. Primary liver cancer is a pathological condition characterized by the development of malignant cells within the hepatic tissues. The development of cancer at extraneous anatomical sites that subsequently metastasizes to the liver does not constitute primary liver

Table 1. SNPs from the GWAS catalog with a $p < 10^{-8}$

No.	Variation and risk allele	p-value
1	rs2856723	3×10^{-43}
2	rs34675408	1×10^{-32}
3	rs9272105	5×10^{-22}
4	rs913493	5×10^{-20}
5	rs2294915	2×10^{-19}
6	rs17401966	2×10^{-18}
7	rs3096380	1×10^{-17}
8	rs9275319	3×10^{-17}
9	rs584368	2×10^{-14}
10	rs2596542	4×10^{-13}
11	rs1110446	9×10^{-13}
12	rs58489806	3×10^{-12}
13	rs6078460	2×10^{-11}
14	rs2523961	6×10^{-11}
15	rs7574865	2×10^{-10}
16	rs1110446	3×10^{-10}
17	rs455804	5×10^{-10}
18	rs58542926	6×10^{-10}
19	rs2523961	6×10^{-10}
20	rs8107030	8×10^{-10}
21	rs10272859	9×10^{-10}
22	rs190121281	4×10^{-9}
23	rs9275572	6×10^{-9}
24	rs2242652	6×10^{-9}
25	rs188273166	1×10^{-8}
26	rs708113	1×10^{-8}
27	rs2896019	2×10^{-8}
28	rs17047200	3×10^{-8}
29	rs541860626	5×10^{-8}

SNP, single-nucleotide polymorphism; GWAS, genome-wide association study.

Table 2. Variants and risk alleles of the prioritized SNPs for liver cancer

Variation and risk alleles	Variants near risk allele ($r^2 > 0.8$)	p-value	GENCODE	Type of allele
rs2294915	rs738409	2×10^{-19}	*PNPLA3*	Missense
rs2896019	rs3761472	2×10^{-8}	*PNPLA3*	Missense

SNP, single-nucleotide polymorphism.

cancer. Primary liver cancer includes several types, such as hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma, and less common varieties like mixed hepatocellular cholangiocarcinoma, fibrolamellar HCC, and the pediatric neoplasm hepatoblastoma [13].

Gene expression of *PNPLA3* across 10 human tissues

The results of *PNPLA3* gene expression across 10 human tissues revealed significant functional consequences of genetic variation. The highest levels of *PNPLA3* gene expression were observed in the liver, sun-exposed skin (lower legs), non-sun-exposed skin (suprapubic), and adipose-subcutaneous fibroblasts and cell cultures, according to analyses of the 10 human tissues from the GTEx database (Fig. 2). Additionally, we found that the SNP IDs rs2294915 and rs2896019 exhibited similar patterns of gene expression variation in sun-exposed skin (lower legs). Notably, patients with liver cancer often report that their skin appears yellow, which may be related to these findings. Further analysis indicated that the *PNPLA3* gene is also highly expressed in suprapubic and underarm skin.

Correlation between gene expression of *PNPLA3* and eQTLs

The study revealed a correlation between the gene expression of *PNPLA3* and eQTLs. To identify eQTLs associated with liver cancer gene expression, we utilized the GTEx database. We identified minor alleles that are related to liver cancer, as detailed in Table 3 [14]. Notably, we discovered that several SNPs, specifically rs2294915 and rs2896019, exhibit high expression in skin tissue. The CC genotype of both rs2294915 and rs2896019 was associated with increased expression in suprapubic and underarm skin compared to the CT and TT genotypes, as shown in Fig. 3.

The research results show that the genomic database could be used to identify gene variations with significant potential in the

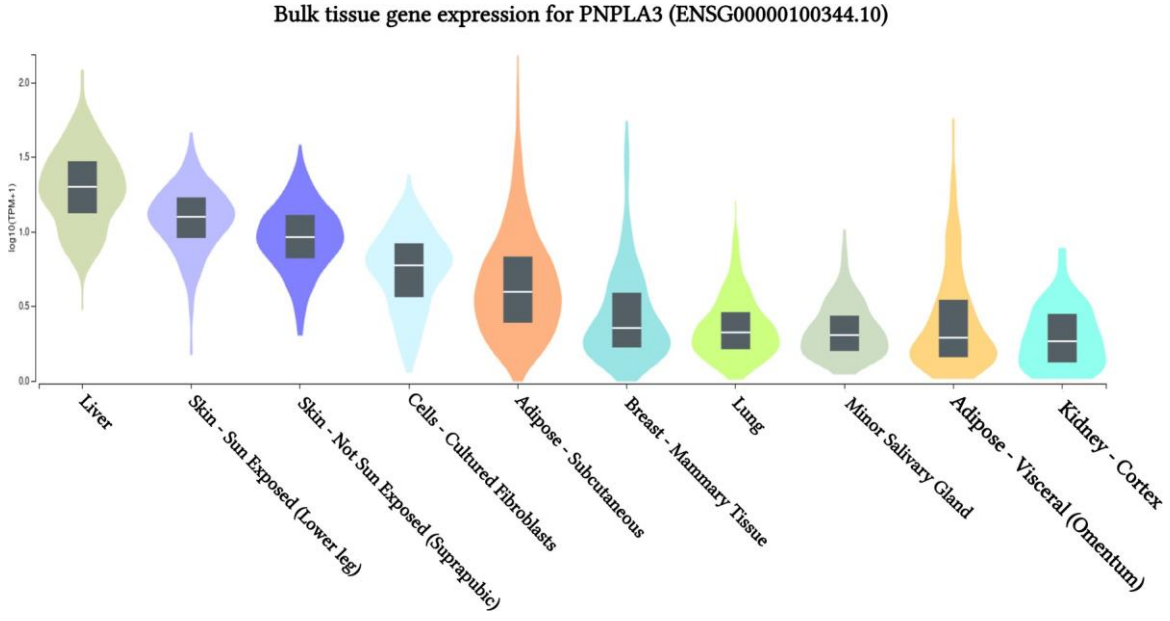


Fig. 2. *PNPLA3* gene expression associated with liver cancer across human tissues based on GTEx Portal analysis.

Table 3. Results of eQTLs in liver cancer from the GTEx Portal database

SNP	Gencode ID (ENSG00000-)	Gene symbol	p-value	Effect size	Tissue	Expression level
rs2294915	100344.1	*PNPLA3*	2.8×10^{-8}	-0.15	Skin - sun exposed (lower leg)	CC > CT > TT
	100344.1	*PNPLA3*	5×10^{-8}	-0.50	Skin - not sun exposed (suprapubic)	CC > CT > TT
rs2896019	100344.1	*PNPLA3*	6.7×10^{-11}	-0.19	Skin - sun exposed (lower leg)	CC > CT > TT
	100344.1	*PNPLA3*	2×10^{-9}	-0.22	Skin - not sun exposed (suprapubic)	CC > CT > TT

Source: expression quantitative trait loci (eQTLs) obtained from the GTEx Portal [14].
eQTL, expression quantitative trait loci; SNP, single-nucleotide polymorphism.

https://doi.org/10.5808/gi.23067

pathogenesis of liver cancer. Liver cancer is marked by the yellowing of the eyes and skin [15]. Nessa et al. [16] note that the severity of liver disease can be gauged by the declining quality of liver function. This quality can be evaluated by measuring total bilirubin levels, serum albumin, and prothrombin time.

Allele frequencies of candidate variants in populations in different continents

We identified variants associated with liver cancer gene expression and conducted allele frequency analysis across various populations. As indicated in Table 4, we evaluated the frequency of allele variants in individuals from Europe, America, East Asia, South Asia, and Africa. The allele frequencies for each SNP differed among these populations, as illustrated in Fig. 4. Both Table 4 and Fig. 4 demonstrate that gene expression levels are higher for populations with increased frequencies of the rs2294915 (C) allele and the rs2896019 (T) allele. Specifically, the gene expression associated with the rs2294915 (C) allele was significantly higher in European and South Asian populations compared to those in America, Africa, and East Asia.

Based on these findings, rs2294915 and rs2896019 may be associated with an increased susceptibility to liver cancer, with the highest effect size of -0.50 observed on skin not exposed to sunlight, such as the suprapubic area. Poggiali and Vercelli [17] describe this condition as being characterized by a disruption in the heme biosynthesis pathway, which is due to decreased activity of hepatic uroporphyrinogen decarboxylase. This disruption leads to an accumulation of light-sensitive by-products, including uroporphyrinogen, resulting in the development of skin fragility and blistering in areas exposed to the sun, as well as impaired liver function.

The allele frequencies of the T and G alleles at loci rs2294915 and rs2896019 were significantly lower in African populations compared to those in American, European, and Southeast Asian populations. Overall, the allele frequencies of the variant alleles rs2294915 and rs2896019 suggest they may contribute to the prevalence of variants affecting the gene expression of *PNPLA3*.

Across human populations, the frequency of the T allele at rs2294915 is associated with high expression of *PNPLA3* in liver cancer. This frequency is much lower in African populations (16%) compared to South Asians (25%), Europeans (25%), East Asians (37%), and Americans (49%). Conversely, the frequency of the C

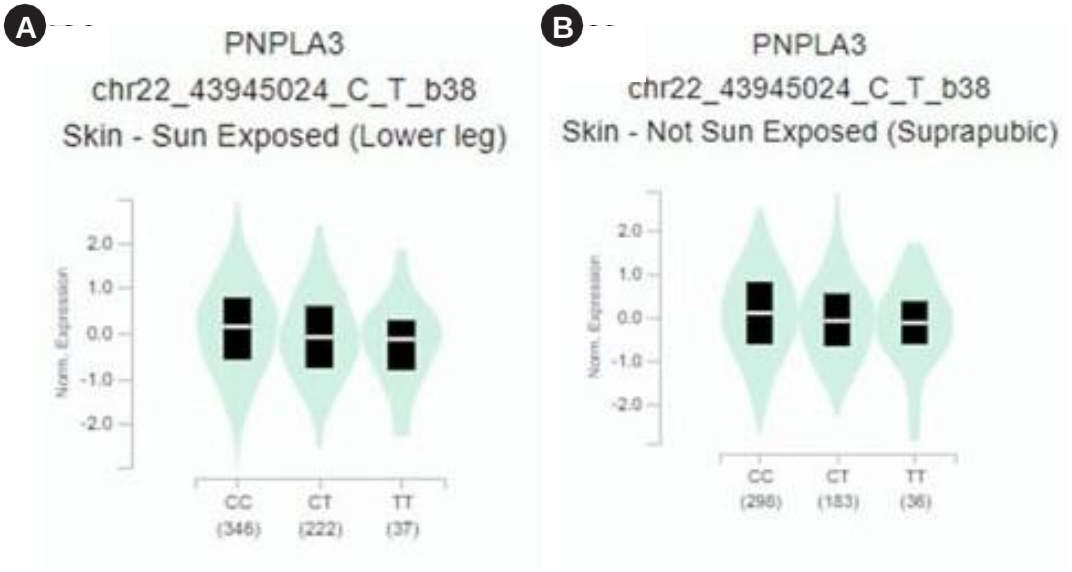


Fig. 3. *PNPLA3* gene expression for each genotype of the single-nucleotide polymorphisms: (A) rs2294915 and (B) rs2896019.

Table 4. Analysis of allele frequencies for the *PNPLA3* gene from variant annotation (SNPnexus)

SNP	Gene	Location	Allele		Allele frequency (n)				
			Ref	Eff*	AFR	AMR	EAS	EUR	SAS
rs2294915	*PNPLA3*	Missense	C	T	T: 0.163 (215)	T: 0.490 (340)	T: 0.365 (368)	T: 0.252 (254)	T: 0.246 (241)
rs2896019	*PNPLA3*	Missense	T	G	G: 0.158 (209)	G: 0.438 (304)	G: 0.364 (367)	G: 0.199 (200)	G: 0.236 (231)

SNP, single-nucleotide polymorphism; Ref, reference; Eff, alternate; AFR, Africa; AMR, America; EAS, East Asia; EUR, Europe; SAS, Southeast Asia.

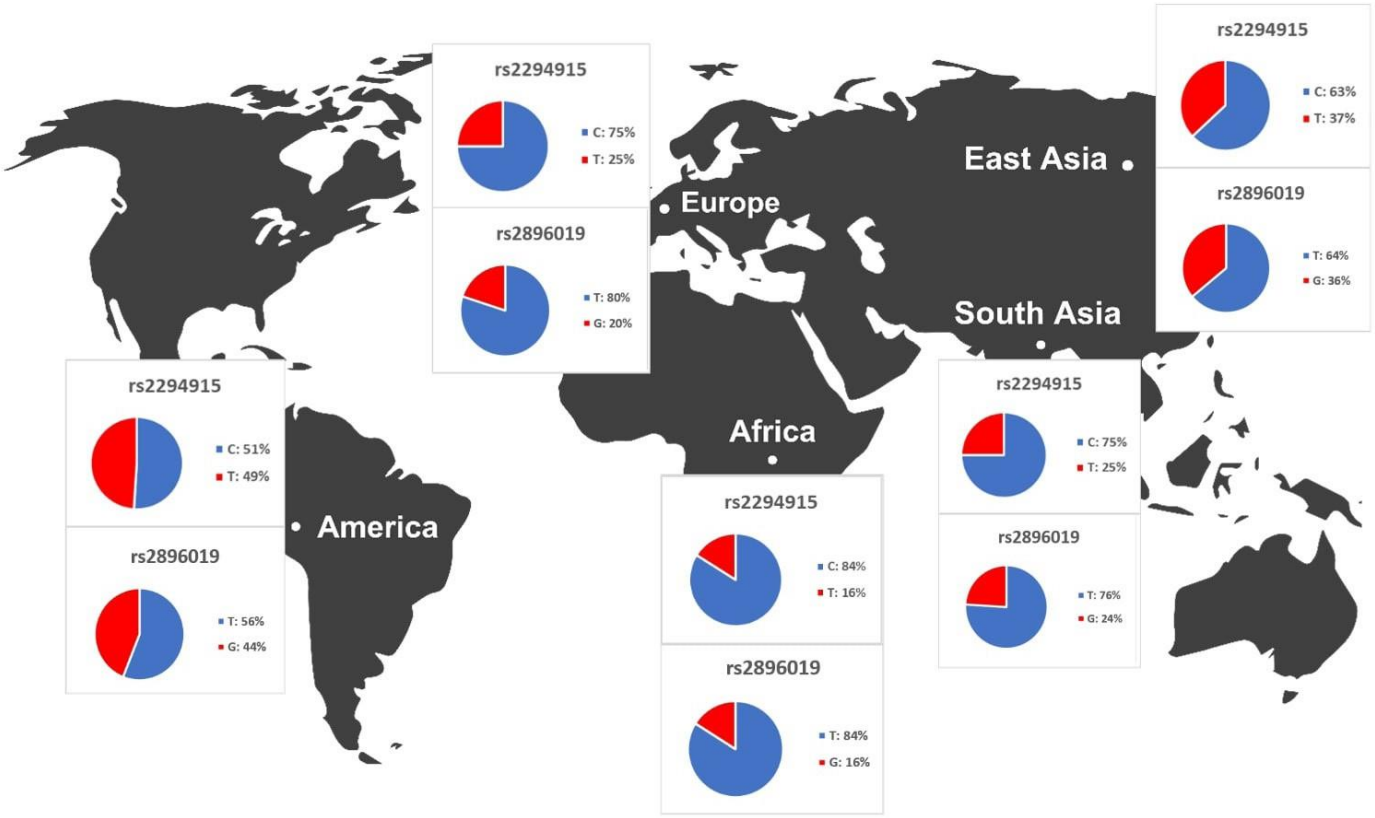


Fig. 4. The results of the distribution of *PNPLA3* allele frequencies across various populations.

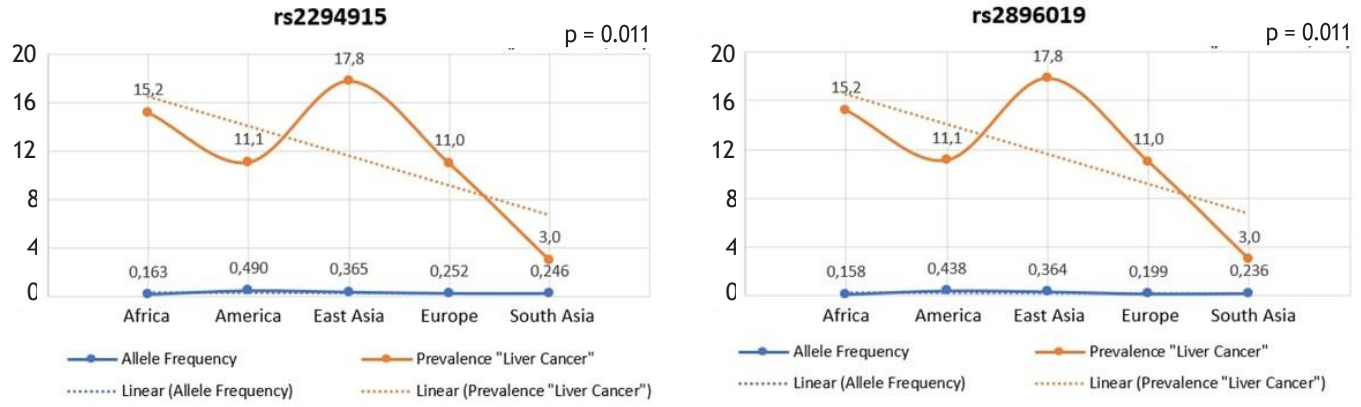


Fig. 5. The association between allele frequency and the prevalence of liver cancer on each continent.

allele at rs2296019 is considerably higher in African (84%), European (80%), South Asian (76%), East Asian (64%), and American (56%) populations. Next, we evaluated the association between allele frequency and the prevalence of liver cancer on each continent. Data on liver cancer prevalence were obtained from Li et al. [12,18]. In this context, two SNPs (rs2294915 and rs2896019) were found to be positively correlated with the prevalence rate of liver cancer across multiple continents (Africa, America, East Asia, Europe, South Asia), as determined by Pearson's correlation analysis ($p = 0.011$) (Fig. 5). Populations with higher frequencies of

https://doi.org/10.5808/gi.23067

variant alleles of these polymorphisms are thought to have a higher prevalence of liver cancer. We highlighted that these two variants (rs2294915 and rs2896019) are more frequent in East Asian and African populations, which exhibit higher aggressiveness of liver cancer compared to America, Europe, and South Asia. This study suggests that individuals in East Asian and African populations carrying the variant alleles rs2294915 and rs2896019 may be more susceptible to liver cancer.

Patients with liver cancer who also have a history of alcohol abuse, consuming ≥3 drinks per day, have a 16% increased risk of developing liver cancer compared to the general population. Additionally, individuals with diabetes and those with central obesity are at twice the risk of developing liver cancer [1]. The diagnosis of liver cancer typically involves serological testing combined with imaging techniques, which is the standard approach for detecting liver carcinoma. However, the sensitivity of the commonly used serological test, which is designed to detect alpha-fetoprotein, is only about 60%. Imaging modalities such as magnetic resonance imaging, computed tomography, and ultrasonography demonstrate high levels of sensitivity and specificity in detecting liver cancer, especially in patients with liver cirrhosis [19].

Variant alleles (rs2294915 and rs2896019) are associated with liver cancer. Populations from Africa, America, East Asia, Europe, and South Asia exhibit associated *PNPLA3* expression, which leads to an increased susceptibility to liver cancer. The identification of unique and pathogenic gene variations for a disease is of great interest for both research and clinical validation. These variants provide insights into disease susceptibility and also act as potential diagnostic and prognostic biomarkers [20]. Furthermore, they can aid in the identification of drug target candidates, an approach referred to as genomic-driven drug repurposing [21]. We expect that the discovery of candidate gene variations in *PNPLA3* will facilitate successful clinical validation, potentially establishing it as a promising diagnostic and prognostic biomarker for liver cancer.

It is important to acknowledge that the genetic variants identified in this study as potentially pathogenic are based on preliminary investigations using genomic and bioinformatics databases. While these findings provide crucial insights for future researchers aiming to validate these genetic variants in liver cancer patients, it is important to proceed with caution. We strongly recommend that future research includes additional functional annotations to aid in the prioritization of these pathogenic genetic variants.

This study identified genetic variants that influence liver cancer, highlighting the importance of the *PNPLA3* gene in liver tissue. Consequently, these population groups exhibit varying susceptibilities to liver cancer based on the associated *PNPLA3* expression levels. The observed variations in allele frequencies of the two identified variants, rs2294915 and rs2896019, across populations from Africa, America, East Asia, Europe, and South Asia, significantly impact *PNPLA3* gene expression. Our study also demonstrated that these two SNPs (rs2294915 and rs2896019) were positively correlated with the prevalence rate. The positive association of prevalence rates was more frequently observed in East Asian and African populations. The higher the frequency of the variant alleles of these polymorphisms in a population, the higher the estimated prevalence rates. The variants investigated in this study are likely to predispose individuals to liver cancer and could play a role in its progression and aggressiveness. These findings highlight the critical importance of understanding genomic variations for precision medicine and for designing targeted screening strategies for liver cancer across diverse populations on different continents.

ORCID

Muhammad Ma'ruf: https://orcid.org/0009-0006-5088-8089
Lalu Muhammad Irham: https://orcid.org/0000-0002-0091-4887
Wirawan Adikusuma: https://orcid.org/0000-0001-9165-690X
Made Ary Samita: https://orcid.org/0000-0002-2846-8786
Sabiah Khairi: https://orcid.org/0000-0002-4412-1176
Barkah Djaka Purwanto: https://orcid.org/0000-0003-1133-1158
Rockie Chong: https://orcid.org/0000-0001-6736-9687
Maulida Mazaya: https://orcid.org/0000-0003-2197-8449
Lalu Muhmammad Harmain Siswanto: https://orcid.org/0009-0009-4030-1012

Authors' Contribution

Conceptualization: MM (Muhammad Ma'ruf), LMI, WA. Data curation: MM (Muhammad Ma'ruf), LMI. Formal analysis: MM (Muhammad Ma'ruf), LMI, WA. Methodology: MM (Muhammad Ma'ruf), LMI, WA. Writing – original draft: MM (Muhammad Ma'ruf), LMI. Writing – review & editing: MM (Muhammad Ma'ruf), LMI, WA, BDP, MAS, SK, RC, MM (Maulida Mazaya), LMHS.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

References

1. McGlynn KA, Petrick JL, El-Serag HB. Epidemiology of hepato-

cellular carcinoma. Hepatology 2021;73 Suppl 1:4-13.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209-249.
4. Hari Kanker Sedunia 2019. Jakarta Selatan: Kementerian Kesehatan Republik Indonesia, 2019. Accessed 2023 May 21. Available from: https://www.kemkes.go.id/article/view/19020100003/harikanker-sedunia-2019.html.
5. Villanueva A. Hepatocellular carcinoma. N Engl J Med 2019;380:1450-1462.
6. Mittal S, El-Serag HB. Epidemiology of hepatocellular carcinoma: consider the population. J Clin Gastroenterol 2013;47 Suppl:S2-S6.
7. Waller LP, Deshpande V, Pyrsopoulos N. Hepatocellular carcinoma: a comprehensive review. World J Hepatol 2015;7:2648-2663.
8. Wang J, Conti DV, Bogumil D, Sheng X, Noureddin M, Wilkens LR, et al. Association of genetic risk score with NAFLD in an ethnically diverse cohort. Hepatol Commun 2021;5:1689-1703.
9. Ma'ruf M, Fadli JC, Mahendra MR, Irham LM, Sulistyani N, Adikusuma W, et al. A bioinformatic approach to identify pathogenic variants for Stevens-Johnson syndrome. Genomics Inform 2023;21:e26.
10. Puspitaningrum AN, Perwitasari DA, Adikusuma W, Djalilah GN, Dania H, Maliza R, et al. Integration of genomic databases and bioinformatic approach to identify genomic variants for Sjogren's syndrome on multiple continents. J Pharm Sci 2022;19:71-81.
11. Chen Z, Boehnke M, Wen X, Mukherjee B. Revisiting the genome-wide significance threshold for common variant GWAS. G3 (Bethesda) 2021;11:jkaa056.
12. Li Q, Cao M, Lei L, Yang F, Li H, Yan X, et al. Burden of liver cancer: From epidemiology to prevention. Chin J Cancer Res 2022;34:554-566.
13. Wege H, Schulze K, von Felden J, Calderaro J, Reig M; Rare Liver Tumors Working Group of the European Reference Network on Hepatological Diseases (ERN RARE-LIVER). Rare variants of primary liver cancer: fibrolamellar, combined, and sarcomatoid hepatocellular carcinomas. Eur J Med Genet 2021;64:104313.
14. GTEx Portal database. Cambridge, MA: Broad Institute of MIT and Harvard, 2023. Accessed 2023 Feb 16. Available from: http://www.gtexportal.org/home/.
15. Fitrianti F, Shatri H, Faisal E, Putranto R, Agung RA, Nainggolan L, Kalista KF, et al. Pain management in advanced hepatocellular carcinoma: a case study. J Penyakit Dalam Indones 2022;9:6.
16. Nessa ML, Yusuf F, Mudatsir M. Relationship of Child Pugh score and quality of life of liver cirrhosis at Internal Medicine RSUDZA Banda Aceh. J Ilmiah Mahasiswa Kedokteran Biomed 2017;2:24-34.
17. Poggiali E, Vercelli A. Sun, alcohol, and skin lesions. Emerg Care J 2023;19:11294.
18. Li Q, Cao M, Lei L, Yang F, Li H, Yan X, et al. Burden of liver cancer: from epidemiology to prevention. Chin J Cancer Res 2022;34:554-566.
19. Huang L, Sun H, Sun L, Shi K, Chen Y, Ren X, et al. Rapid, label-free histopathological diagnosis of liver cancer based on Raman spectroscopy and deep learning. Nat Commun 2023;14:48.
20. Irham LM, Chou WH, Calkins MJ, Adikusuma W, Hsieh SL, Chang WC. Genetic variants that influence SARS-CoV-2 receptor TMPRSS2 expression among population cohorts from multiple continents. Biochem Biophys Res Commun 2020;529:263-269.
21. Afief AR, Irham LM, Adikusuma W, Perwitasari DA, Brahmadhi A, Chong R. Integration of genomic variants and bioinformatic-based approach to drive drug repurposing for multiple sclerosis. Biochem Biophys Rep 2022;32:101337.

https://doi.org/10.5808/gi.23067

Profil Genomics & Informatics Journal

(Sumber: Resurchify.com)

Impact Score 2022-2023	h-Index 2022-2023	Rank 2022-2023	SJR 2022-2023
1.03	10	11649	0.437

Note: The **impact score** or **impact index** shown here is equivalent to the average number of times documents published in a journal/conference in the past two years have been cited in the current year (i.e., Cites / Doc. (2 years)). It is based on Scopus data and can be a little higher or different compared to the impact factor (IF) produced by Journal Citation Report. Please refer to the Web of Science data source to check the exact journal impact factor ™ (Thomson Reuters) metric.

Important Metrics and Factor

Title	Genomics and Informatics
Abbreviation	Genom. Inform.
Publication Type	Journal
Subject Area, Categories, Scope	Ecology, Evolution, Behavior and Systematics (Q2); Genetics (Q3); Health Informatics (Q3)
h-index	10
Overall Rank/Ranking	11649
SCImago Journal Rank (SJR)	0.437
Impact Score	1.03
Publisher	Korea Genome Organization
Country	South Korea
ISSN	22340742
Best Quartile	Q2
Coverage History	2019-2022

Genomics and Informatics Impact IF 2023 Prediction

Impact IF 2022 of **Genomics and Informatics** is **1.03**. If the same downward trend persists, Impact IF may **fall** in 2023 as well.

Impact IF Trend

Year wise Impact IF of Genomics and Informatics. Based on Scopus data.

Year	Impact IF
2023/2024	Coming Soon
2022	1.03
2021	1.73
2020	1.47
2019	0.00

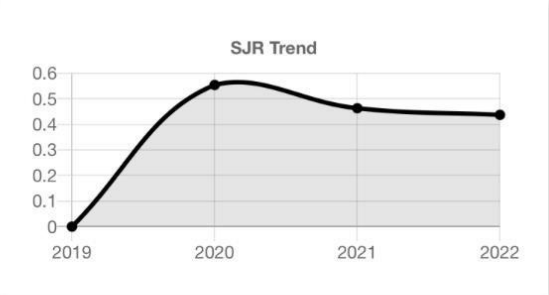
Genomics and Informatics h-index

Genomics and Informatics has an h-index of **10**. It means **10** articles of this **journal** have more than **10** number of citations. The h-index is a way of measuring the productivity and citation impact of the publications. The h-index is defined as the maximum value of h such that the given journal/author has published h papers that have each been cited at least h number of times.

Genomics and Informatics Rank and SCImago Journal Rank (SJR)

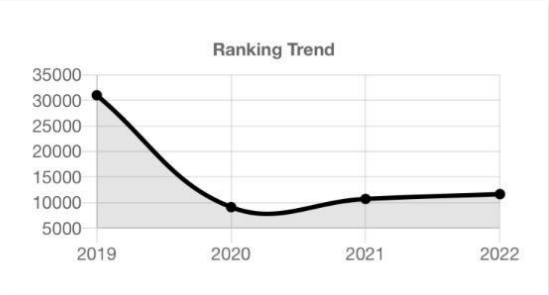
The overall rank of **Genomics and Informatics** is **11649**. According to SCImago Journal Rank (SJR), this journal is ranked **0.437**. SCImago Journal Rank is an indicator, which measures the scientific influence of journals. It considers the number of citations received by a journal and the importance of the journals from where these citations come.

SJR of Genomics and Informatics by Year



Year	SJR
2023/2024	Coming Soon
2022	0.437
2021	0.463
2020	0.554
2019	

Ranking of Genomics and Informatics by Year



Year	Ranking
2023/2024	Coming Soon
2022	11649
2021	10713
2020	9112
2019	30992