

Identification of Diagnostic and Prognostic Biomarkers in Nasopharyngeal Carcinoma Using Integrated Transcriptomics and Elastic Net Survival Analysis



Nur Aziz^{1,*} , Laily Rahmawati² and Jae Youl Cho³

¹Department of Histology and Cell Biology, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Indonesia

²Department of Molecular Biology, Faculty of Medicine, Universitas Negeri Yogyakarta, Indonesia

³Department of Integrative Biotechnology, and Biomedical Institute for Convergence at SKKU, Sungkyunkwan University, Suwon16419, Republic of Korea

© 2025 The Author(s). Published by Bentham Open.

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



CrossMark

*Address correspondence to this author at the Department of Histology and Cell Biology, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Indonesia; nur.aziz20@ugm.ac.id

Published: August 23, 2025

Cite as: Aziz N, Rahmawati L, Cho J. Identification of Diagnostic and Prognostic Biomarkers in Nasopharyngeal Carcinoma Using Integrated Transcriptomics and Elastic Net Survival Analysis. Open Bioinform J. 2025; 18: e18750362408821. <http://dx.doi.org/10.2174/0118750362408821250821062122>



Send Orders for Reprints to
reprints@benthamscience.net

Supplementary Table 1. GEO Series Used in This Study.

No	GEO Accession	N	NPC	Original Study	Prognostic Data	Transcriptomic Type	Platform	Remarks
1	GSE218847	4	4	Not published	Not available	Microarray	GPL29371	Used as validation dataset
2	GSE62328	0	30	Available	Not available	Microarray	GPL13693	Excluded (no normal sample)
3	GSE132112	0	95	Available	Not available	Microarray	GPL17585	Excluded (no normal sample)
4	GSE61218	6	10	Available	Not available	Microarray	GPL19061	Used as validation dataset
5	GSE103611	0	48	Available	Not available	Microarray	GPL19251	Excluded (no normal sample)
6	GSE40290	8	25	Not published	Not available	Microarray	GPL8380	Used as validation dataset
7	GSE12452	10	31	Available	Not available	Microarray	GPL570	Used as diagnostic training dataset
8	GSE53819	18	18	Available	Not available	Microarray	GPL6480	Used as diagnostic training dataset
9	GSE32389	0	20	Available	Not available	Microarray	GPL4381	Excluded (no normal sample)
10	GSE13597	3	25	Available	Not available	Microarray	GPL96	Excluded (imbalanced sample)
11	GSE34573	3	15	Available	Not available	Microarray	GPL570	Used as validation dataset
12	GSE64634	4	12	Available	Not available	Microarray	GPL570	Used as validation dataset
13	GSE227541	6	6	Available	Not available	RNA-seq	GPL16791	Used as validation dataset
14	GSE153125	0	22	Not published	Not available	RNA-seq	GPL24676	Excluded (no normal sample)
15	GSE134886	3	3	Available	Not available	RNA-seq	GPL20795	Used as validation dataset
16	GSE165181	0	2	Available	Not available	RNA-seq	GPL11154	Excluded (no normal sample)
17	GSE118719	4	7	Available	Not available	RNA-seq	GPL20301	Used as validation dataset
18	GSE68799	4	42	Not published	Not available	RNA-seq	GPL11154	Excluded (imbalanced sample)
19	GSE63381	0	4	Available	Not available	RNA-seq	GPL11154	Excluded (no normal sample)
20	GSE102349	0	113	Available	Available	RNA-seq	GPL11154	Used as prognostic dataset

Supplementary Table 2. Statistical Analysis of Diagnostic Validation Datasets.

No	Gene	p_value	Group_Higher	Significance	GSE	Group
1	BUB1B	0.029758	NPC	*	GSE118719	Validation Dataset
2	GAS2L3	0.156376	NPC	ns	GSE118719	Validation Dataset
3	NFE2L3	0.149591	NPC	ns	GSE118719	Validation Dataset
4	OIP5	0.072601	NPC	ns	GSE118719	Validation Dataset
5	PDGFRL	0.072601	NPC	ns	GSE118719	Validation Dataset
6	CD1D	1.48E-04	N	***	GSE118719	Validation Dataset
7	CYP4B1	0.694486	NPC	ns	GSE118719	Validation Dataset
8	IL33	0.072601	N	ns	GSE118719	Validation Dataset
9	KLF2	4.93E-05	N	***	GSE118719	Validation Dataset
10	NAPSB	0.010733	N	*	GSE118719	Validation Dataset
11	VILL	6.69E-04	N	***	GSE118719	Validation Dataset
12	BUB1B	0.351997	N	ns	GSE134886	Validation Dataset
13	GAS2L3	0.918094	N	ns	GSE134886	Validation Dataset
14	NFE2L3	0.136663	N	ns	GSE134886	Validation Dataset
15	OIP5	0.257984	N	ns	GSE134886	Validation Dataset
16	PDGFRL	0.82909	NPC	ns	GSE134886	Validation Dataset
17	CD1D	0.706765	NPC	ns	GSE134886	Validation Dataset
18	CYP4B1	0.240524	NPC	ns	GSE134886	Validation Dataset
19	IL33	0.080856	NPC	ns	GSE134886	Validation Dataset
20	KLF2	0.280287	NPC	ns	GSE134886	Validation Dataset
21	NAPSB	0.915994	NPC	ns	GSE134886	Validation Dataset
22	VILL	0.124656	NPC	ns	GSE134886	Validation Dataset
23	BUB1B	0.339853	NPC	ns	GSE218847	Validation Dataset
24	GAS2L3	0.068229	NPC	ns	GSE218847	Validation Dataset
25	NFE2L3	0.009215	NPC	**	GSE218847	Validation Dataset
26	OIP5	0.359455	NPC	ns	GSE218847	Validation Dataset
27	PDGFRL	0.053508	NPC	ns	GSE218847	Validation Dataset
28	CD1D	0.347649	N	ns	GSE218847	Validation Dataset
29	CYP4B1	0.568812	N	ns	GSE218847	Validation Dataset
30	IL33	0.030383	N	*	GSE218847	Validation Dataset
31	KLF2	0.007885	N	**	GSE218847	Validation Dataset
32	NAPSB	0.060602	N	ns	GSE218847	Validation Dataset
33	VILL	0.030383	N	*	GSE218847	Validation Dataset
34	BUB1B	0.023012	N	*	GSE227541	Validation Dataset
35	GAS2L3	0.47117	N	ns	GSE227541	Validation Dataset
36	NFE2L3	1.03E-04	N	***	GSE227541	Validation Dataset
37	OIP5	0.030639	N	*	GSE227541	Validation Dataset
38	PDGFRL	0.688921	N	ns	GSE227541	Validation Dataset
39	CD1D	0.020241	NPC	*	GSE227541	Validation Dataset
40	CYP4B1	0.030639	NPC	*	GSE227541	Validation Dataset
41	IL33	0.065552	NPC	ns	GSE227541	Validation Dataset
42	KLF2	0.045328	NPC	*	GSE227541	Validation Dataset
43	NAPSB	0.081819	NPC	ns	GSE227541	Validation Dataset
44	VILL	3.08E-04	NPC	***	GSE227541	Validation Dataset
45	BUB1B	0.020611	NPC	*	GSE34573	Validation Dataset
46	GAS2L3	0.369364	NPC	ns	GSE34573	Validation Dataset
47	NFE2L3	0.019392	NPC	*	GSE34573	Validation Dataset
48	OIP5	0.321121	NPC	ns	GSE34573	Validation Dataset
49	PDGFRL	0.421874	NPC	ns	GSE34573	Validation Dataset
50	CD1D	0.339642	N	ns	GSE34573	Validation Dataset
51	CYP4B1	0.163918	N	ns	GSE34573	Validation Dataset
52	IL33	0.088838	N	ns	GSE34573	Validation Dataset
53	KLF2	0.020611	N	*	GSE34573	Validation Dataset
54	NAPSB	0.344879	N	ns	GSE34573	Validation Dataset

No	Gene	p_value	Group_Higher	Significance	GSE	Group
55	VILL	0.13598	N	ns	GSE34573	Validation Dataset
56	BUB1B	0.037579	NPC	*	GSE40290	Validation Dataset
57	NFE2L3	8.51E-07	NPC	***	GSE40290	Validation Dataset
58	OIP5	0.00742	NPC	**	GSE40290	Validation Dataset
59	PDGFRL	0.008652	NPC	**	GSE40290	Validation Dataset
60	CD1D	3.27E-04	N	***	GSE40290	Validation Dataset
61	CYP4B1	8.39E-04	N	***	GSE40290	Validation Dataset
62	KLF2	4.39E-09	N	***	GSE40290	Validation Dataset
63	VILL	0.729818	NPC	ns	GSE40290	Validation Dataset
64	BUB1B	7.38E-05	NPC	***	GSE61218	Validation Dataset
65	GAS2L3	0.143117	NPC	ns	GSE61218	Validation Dataset
66	NFE2L3	4.43E-07	NPC	***	GSE61218	Validation Dataset
67	OIP5	1.37E-05	NPC	***	GSE61218	Validation Dataset
68	PDGFRL	0.115778	NPC	ns	GSE61218	Validation Dataset
69	CD1D	0.205666	N	ns	GSE61218	Validation Dataset
70	CYP4B1	0.060156	N	ns	GSE61218	Validation Dataset
71	IL33	0.00112	N	**	GSE61218	Validation Dataset
72	KLF2	1.58E-04	N	***	GSE61218	Validation Dataset
73	VILL	0.005309	N	**	GSE61218	Validation Dataset
74	BUB1B	0.060156	NPC	ns	GSE64634	Validation Dataset
75	GAS2L3	0.622793	NPC	ns	GSE64634	Validation Dataset
76	NFE2L3	0.088758	NPC	ns	GSE64634	Validation Dataset
77	OIP5	0.034949	NPC	*	GSE64634	Validation Dataset
78	PDGFRL	0.302646	NPC	ns	GSE64634	Validation Dataset
79	CD1D	0.101606	N	ns	GSE64634	Validation Dataset
80	CYP4B1	0.068777	N	ns	GSE64634	Validation Dataset
81	IL33	0.033822	N	*	GSE64634	Validation Dataset
82	KLF2	0.013325	N	*	GSE64634	Validation Dataset
83	NAPSB	0.097008	N	ns	GSE64634	Validation Dataset
84	VILL	0.002568	N	**	GSE64634	Validation Dataset

DISCLAIMER: The above article has been published, as is, ahead-of-print, to provide early visibility but is not the final version. Major publication processes like copyediting, proofing, typesetting and further review are still to be done and may lead to changes in the final published version, if it is eventually published. All legal disclaimers that apply to the final published article also apply to this ahead-of-print version.