VNUHCM Journal of Health Sciences 2026, 7(1): 920-927

Open Access Full Text Article

Research article

Immunoinformatics-driven identification and evaluation of specific epitopes from LCINS-Associated oncogenes for therapeutic vaccine design

Duy-Khang Nguyen [1,2], Xuan-Huy Le [1,2], Vi-Anh Nguyen-Hoang [1,2], Bao-Long Nguyen-Tran [1,2], Ngan-Giang Truong [1,2], Vinh-Hy Truong [1,2], Chi-Bao Bui [1,2], Truc Ly Nguyen [1,2,*]

ABSTRACT

Background: Lung cancer in never-smokers (LCINS) is an increasingly prevalent and clinically distinct subtype of lung cancer, particularly in East Asian populations. Despite its rising incidence, there are currently no antigen-specific immunotherapeutic strategies or vaccines targeting lung cancer in never-smokers. This study aimed to identify and evaluate immunogenic epitopes derived from key oncogenic drivers in lung cancer in never-smokers as preliminary steps toward multi-epitope vaccine design.

Methods: Protein sequences of five lung cancer in never-smokers associated oncogenes (EGFR, KRAS, HER2, BRAF, and MET) were retrieved from UniProtKB and analyzed using a multi-tiered immunoinformatics workflow. B-cell, major histocompatibility complex class I, and major histocompatibility complex class II epitopes were predicted using tools such as NetMHCpan, NetMHCIIpan, and BepiPred. Candidate epitopes were identified based on HLA binding affinity, antigenicity (VaxiJen), toxicity, allergenicity while the ability of MHC-II epitopes to trigger the production of cytokines (IFN-γ, IL-4, IL-10) will be further assessed. Population coverage analysis was performed using the IEDB Population Coverage tool.

Results: Out of 1027 initial epitope predictions, only 16 epitopes (8 MHC-I, 4 MHC-II, and 4 B-cell epitopes) met all selection criteria. The final epitope exhibited strong antigenicity, HLA binding potential, and favorable global population coverage (81.23%). Several epitopes were highly specific to LCINS-related driver mutations, suggesting relevance for precision immunotherapy.

Conclusion: This study presents a rational, immunoinformatics-guided approach to identify and prioritize epitope candidates from lung cancer in never-smokers oncogenes. These findings provide a foundation for future design and preclinical evaluation of personalized multi-epitope cancer vaccines tailored to lung cancer in never-smokers.

Key words: Lung cancer in never-smokers (LCINS), Immunoinformatics, Epitope prediction, Tumor-associated antigens, HLA binding, Population coverage

[1]*University of Health Sciences, Vietnam National University, Ho Chi Minh City, Vietnam*

[2]*Vietnam National University, Ho Chi Minh City, Vietnam*

Correspondence

Truc Ly Nguyen , University of Health Sciences, Vietnam National University, Ho Chi Minh City, Vietnam

Vietnam National University, Ho Chi Minh City, Vietnam

Email: nttly@uhsvnu.edu.vn

History

- Received: 03-07-2025
- Revised: 19-12-2025
- Accepted: 13-04-2026
- Published Online: 12-06-2026

DOI : 10.32508/vnuhcmj-hs.v7i1.689

INTRODUCTION:

Lung cancer is the leading cause of cancer-related deaths worldwide, with nearly 2.5 million new cases and over 1.8 million deaths reported globally in 2022 alone. In Vietnam, according to GLOBOCAN 2022, there were approximately 24,426 new cases and 22,597 deaths due to lung cancer annually [1] .Lung cancer ranks third in incidence and second in mortality among all cancers, only after liver cancer. Furthermore, survival analyses in lung cancer patients have shown that clinical performance status and demographic characteristics have a substantial impact on overall prognosis, highlighting the variety of disease outcomes [2] While tobacco smoking is the most well-known risk factor, the incidence of lung cancer in never-smokers (LCINS) is rising significantly, particularly among women, younger individuals, and Asian populations [3,4] .In the United States alone, over 20,000 LCINS-related deaths were recorded in 2023, making it the eighth leading cause of cancer death in the country [5] .Globally, LCINS now ranks fifth among the causes of cancer-related deaths [4] .

Recent studies suggest that LCINS is a distinct pathological entity with unique biological and genetic characteristics compared to smoking-associated lung cancer [6] .Unlike the latter, which typically presents with a high tumor mutational burden due to carcinogens in tobacco smoke, LCINS is primarily driven by specific activating oncogenic mutations [6] .Key oncogenes implicated in LCINS include EGFR, KRAS, HER2, BRAF, and MET [4,7–11] .Among these, EGFR mutations, particularly exon 19 deletions and L858R sub-

Cite this article : D N, X L, V N, B N, N T, V T, C B, T L N. **Immunoinformatics-driven identification and evaluation of specific epitopes from LCINS-Associated oncogenes for therapeutic vaccine design**. *VNUHCM J. Health Sci. 2026; 7(1): 920-927.*

stitutions, are the most prevalent, occurring in 40-60% of LCINS cases, especially in East Asian populations[8] .KRAS mutations, notably G12C, are observed in 10-15% of lung adenocarcinoma cases and are emerging as promising therapeutic targets[11,12] .HER2 exon 20 insertions, present in approximately 1-3% of LCINS, are associated with increased tumor aggressiveness[10] , whereas BRAF V600E mutations, found in about 1-2% of cases, activate the MAPK signaling pathway and promote tumor progression[9] .Additionally, MET exon 14 skipping mutations contribute to receptor dysregulation, tumor invasiveness, and resistance to targeted therapies[7] .

These mutations not only drive uncontrolled cell proliferation but also contribute to treatment resistance and immune evasion, limiting the effectiveness of current therapies such as tyrosine kinase inhibitors (TKIs). For instance, EGFR-mutant tumors often acquire secondary mutations like T790M or C797S upon osimertinib treatment, leading to resistance[13,14] .Similarly, KRAS G12C-targeting agents (e.g., sotorasib)[12] and capmatinib for MET-altered lung cancer[15] face comparable challenges. Furthermore, immune checkpoint inhibitors (ICIs) exhibit limited efficacy in LCINS, possibly due to low tumor mutational burden, poor antigen presentation, and an immunosuppressive tumor microenvironment[16].

Cancer vaccines offer a promising immunotherapeutic alternative by stimulating durable, tumor-specific immune responses. However, existing vaccine efforts in lung cancer, such as TG4010 (targeting MUC1) or GVAX (based on whole tumor cells), have largely focused on antigens prevalent in smokers. These may be less relevant for LCINS, where the mutational and antigenic landscape is distinct[17,18] .Currently, no vaccine platforms specifically designed for LCINS-associated driver mutations have been reported in the literature. Therefore, this study applied a comprehensive immunoinformatics-based approach to identify and prioritize potential B-cell, MHC class I, and class II T-cell epitopes derived from LCINS-specific oncogenes. These epitopes were evaluated for antigenicity, binding affinity, cytokine induction, and global population coverage. These findings provide initial immunological insights that may inform the development of personalized multi-epitope vaccines for lung cancer in never-smokers.

MATERIALS AND METHODS:

2.1. Retrieval of target proteins:

Five tumor-associated antigens commonly implicated in lung cancer in never-smokers (LCINS) /vcnwere selected based on literature evidence of their mutation frequency and oncogenic potential. These included: epidermal growth factor receptor (EGFR), KRAS proto-oncogene (KRAS), human epidermal growth factor receptor 2 (HER2/ERBB2), B-Raf proto-oncogene (BRAF), and MET proto-oncogene (MET). Protein sequences were retrieved from the UniProtKB database[19,20] in FASTA format using the following accession IDs: EGFR (P00533), KRAS (P01116), HER2 (P04626), BRAF (P15056), and MET (P08581). All sequences were stored and processed for subsequent in silico epitope prediction.

2.2. Identification of B-cell epitopes

Linear B-cell epitopes were predicted using the BepiPred tool available through the IEDB (Immune Epitope Database) analysis resource[21] .The amino acid sequences of the five selected proteins were submitted with default parameters, and regions with prediction scores above the 0.35 threshold, which is the default threshold in the BepiPred system we used, were considered putative B-cell epitopes. Only sequences $\geq$ 9 amino acids in length were retained for further analysis[21] .

2.3. Identification of MHC class I epitopes

Cytotoxic T lymphocyte (CTL) epitopes were identified using NetMHCpan-4.1, which predicts the binding affinity of 9-mer peptides to HLA class I molecules[22] .Common alleles representing broad ethnic distribution were selected, including HLA-A*01:01*, HLA-A*02:01*, HLA-A*26:01*, HLA-B*08:01*, HLA-B*15:01* and HLA-B*58:01*.

Binding predictions were filtered using the percentile rank (%Rank EL), which normalizes affinity scores across alleles by comparing each peptide's predicted elution ligand likelihood (Score EL) against a reference set of random natural peptides. Peptides with %Rank < 0.5% were classified as strong binders (SB) and those with 0.5% $\leq$ %Rank < 2% as weak binders (WB), per default NetMHCpan thresholds[20,22] .Only SB and WB peptides proceeding to downstream antigenicity, allergenicity, and toxicity screening were retained for multi-epitope vaccine construction.

2.4. Identification of MHC class II epitopes

Helper T lymphocyte (HTL) epitopes were predicted using the IEDB recommended 2.22 method, incorporating the NetMHCIIpan-4.3 prediction algorithm for 15-mer peptides[20,23] .The following HLA-DRB1 alleles were considered: DRB1*01:01*, DRB1*04:01*, DRB1*13:01* and DRB1*15:01*.

Binding predictions were filtered using the percentile rank (%Rank), which normalizes affinity scores across

alleles by comparing each peptide's predicted binding affinity against a reference set of random natural peptides. Peptides with %Rank < 1% were classified as and those with 1% ≤ %Rank < 5% as WB, per default NetMHCIIpan-4.3 thresholds[23] .Peptides with %Rank > 20 were excluded as potential HLA-irrelevant contaminants, while those with %Rank ≤ 5 were retained as high-confidence binders for downstream analyses[23] .Only SB and WB peptides proceeding to downstream antigenicity, allergenicity, and toxicity screening were retained for multi-epitope vaccine construction.

2.5. Antigenicity and safety evaluation:

Antigenicity was assessed using VaxiJen v2.0, applying the "tumor" model with a minimum threshold of 0.5. MHC-I epitopes with scores ≥ 1.5, and MHC-II epitopes with scores ≥ 0.7 and B-cell epitopes with scores ≥ 1.0, were considered highly antigenic[24] . Toxicity and allergenicity were evaluated using ToxinPred and AllergenFP respectively, and only non-toxic, non-allergenic epitopes were retained[20,25,26] .Cytokine-inducing potential was further evaluated using IFNepitope, IL4pred, and IL10pred, with predictions applied exclusively to MHC class II epitopes, as cytokine induction is primarily mediated by $CD4^+$ T helper cells upon antigen recognition. Only epitopes predicted to induce all three immunostimulatory cytokines (IFN-γ, IL-4, and IL-10) were retained for further analysis[27–29] .

2.6. Population coverage analysis

To assess the immunological relevance of the selected epitope set at a population level, IEDB's Population Coverage tool was used[20,30] .The combined HLA alleles associated with both MHC class I and II epitopes were inputted to calculate the projected percentage of individuals in different global regions, including Vietnam, expected to present at least one epitope-HLA combination. The results were expressed as total projected population coverage (%), average number of epitope-HLA hits per individual, and pc90 - the minimum number of combinations recognized by 90% of the population.

RESULTS:

3.1. Antigen retrieval and epitope identification:

Protein sequences of five LCINS-associated oncogenic targets (EGFR, KRAS, HER2, BRAF, and MET) were retrieved from the UniProtKB database using their respective accession IDs (Table 1). These sequences were processed in FASTA format and used as input for downstream epitope prediction analyses.

Initial epitope predictions yielded a total of 1027 candidates, including 644 MHC-I, 305 MHC-II, and 78 B-cell epitopes (*Data sheets supporting information*) . These were filtered through a multi-parameter in silico evaluation pipeline incorporating HLA binding affinity, antigenicity (VaxiJen scores), allergenicity, toxicity, and cytokine-inducing potential (IFN-γ, IL-4, IL-10). Only 16 epitopes met all stringent criteria, including 8 MHC-I, 4 MHC-II, and 4 B-cell epitope, representing a retention rate of approximately 1.56%, indicative of high selectivity.

Table 1: AccessionIDs of selected LCINS-related proteins from UniProtKB.

Protein	Gene symbol	Accession ID (UniProtKB)	Species
EGFR	EGFR	P00533	Homo sapiens
KRAS	KRAS	P01116	Homo sapiens
HER2	ERBB2	P04626	Homo sapiens
BRAF	BRAF	P15056	Homo sapiens
MET	MET	P08581	Homo sapiens

3.2. Evaluation of final epitopes

The final list of epitopes is summarized in Table 2. All selected epitopes exhibited strong to moderate binding affinity to commonly expressed HLA alleles. Antigenicity scores assessed by VaxiJen v2.0 were above the tumor-specific threshold (≥ 1.5 for MHC-I; ≥ 0.7 for MHC-II and ≥ 1.0 for B-cell). These epitopes were derived from multiple tumor driver antigens, providing multi-target coverage. Additionally, all epitopes were tested and evaluated to ensure they are non-allergenic and non-toxic. Specifically, for MHC-II epitopes, we further assessed their immunogenicity by evaluating their ability to induce the production of three immunostimulatory cytokines (IFN-γ, IL-4, and IL-10).

These epitopes were selected to ensure a comprehensive immune response, targeting both cytotoxic T cells (via MHC-I), helper T cells (via MHC-II), and B-cell activation.

3.3. Population coverage analysis

The twelve MHC-restricted epitopes were analyzed for projected population coverage using the IEDB Population Coverage tool. The cumulative global coverage was estimated at 81.23%, with an average of 2.25 epitope-HLA hits per person and a pc90 value of 0.53. Particularly high coverage was observed in Europe (89.53%), North Africa (60.27%), and South Asia

Table 2: Finalselected epitopes and their characteristics.

Type of epitope	Protein	Epitopes	Binding MHC alleles	Bindin affinity	VaxiJen scorea	Allergeni	Toxicity	IFN-γ induction	Interleuk 4	Interleukin-10
B-cell	EGFR	EEDMDD	NA	NA	1.2427	non-allergen	non-toxin	NA	NA	NA
	KRAS	TAGQEEY	NA	NA	1.0366	non-allergen	non-toxin	NA	NA	NA
		TSAKTRQ	NA	NA	1.1849	non-allergen	non-toxin	NA	NA	NA
	HER2	RSSSTRSG	NA	NA	1.1662	non-allergen	non-toxin	NA	NA	NA
MHC-II	EGFR	LEIIRGRT	DRB1_13	17.36	0.9818	non-allergen	non-toxin	positive	IL4-inducer	IL10-inducer
	HER2	RPRFRELV	DRB1_01	6.5	1.2602	non-allergen	non-toxin	positive	IL4-inducer	IL10-inducer
			DRB1_04	22.95		non-allergen	non-toxin	positive	IL4-inducer	IL10-inducer
	BRAF	QQLQAFK	DRB1_15	41.84	0.7115	non-allergen	non-toxin	positive	IL4-inducer	IL10-inducer
	MET	PSSLIVHF	DRB1_15	43.29	0.8773	non-allergen	non-toxin	positive	IL4-inducer	IL10-inducer
MHC-I	EGFR	GSTAENA	HLA-A*01:01	999.82	1.5050	non-allergen	non-toxin	NA	NA	NA
		DSRPKFRI	HLA-B*08:01	136.33	1.7348	non-allergen	non-toxin	NA	NA	NA
		LCNVESIC	HLA-B*58:01	44.07	1.5554	non-allergen	non-toxin	NA	NA	NA
	HER2	PAFDNLY	HLA-B*58:01	28.11	1.6379	non-allergen	non-toxin	NA	NA	NA
	BRAF	SLYHHLH	HLA-A*02:01	14.41	1.6985	non-allergen	non-toxin	NA	NA	NA
			HLA-B*08:01	280.41		non-allergen	non-toxin	NA	NA	NA
	MET	YVNDFFN	HLA-A*02:01	75.04	1.5357	non-allergen	non-toxin	NA	NA	NA
		SVKDRFI	HLA-A*26:01	389.36	1.6717	non-allergen	non-toxin	NA	NA	NA
			HLA-B*08:01	209.30		non-allergen	non-toxin	NA	NA	NA
			HLA-B*15:01	225.32		non-allergen	non-toxin	NA	NA	NA
		RLKETKD	HLA-B*15:01	165.53	1.7277	non-allergen	non-toxin	NA	NA	NA

[a]VaxiJenscore above 1.5 for MHC-I epitopes, above 0.7 for MHC-II epitopes and 1.0 forB-cell epitopes.

Table 3: Population coverage analysis of theselected MHC epitopes.

Population/area	Class combined		
	coverage	average_hit	pc90c
Central Africa	42.14%	0.81	0.17
Central America	10.59%	0.17	0.11
East Africa	55.85%	1.2	0.23
East Asia	62.83%	1.45	0.27
Europe	89.53%	2.83	0.96
North Africa	60.27%	1.39	0.25
North America	83.72%	2.37	0.61
Northeast Asia	50.80%	1.15	0.2
Oceania	46.12%	1.05	0.19
South Africa	49.07%	0.94	0.2
South America	43.48%	0.88	0.18
South Asia	62.65%	1.35	0.27
Southeast Asia	43.78%	0.97	0.18
Southwest Asia	59.73%	1.29	0.25
Vietnam	39.36%	0.8	0.16
West Africa	56.12%	1.27	0.23
West Indies	70.44%	1.74	0.34
World	81.23%	2.25	0.53
Average	55.98	1.33	0.3
Standard deviation	18.06	0.62	0.2

[a] projected population coverage
[b] average number of epitope hits/HLAcombinations recognized by the population
[c] minimum number of epitope hits/HLAcombinations recognized by 90% of the population

(62.65%) (Table 3).

However, regional coverage in Vietnam (39.36%), East Asia (62.83%), and Southeast Asia (43.78%) was moderate, highlighting the need for future expansion of allele selection to enhance regional specificity.

DISCUSSION:

This study presents a computational approach to the identification and evaluation of potential B-cell and T-cell epitopes derived from major oncogenic drivers in lung cancer in never-smokers (LCINS). Given the increasing prevalence of LCINS in East Asian populations and the absence of targeted immunotherapies, early-stage epitope discovery represents an important foundational step toward the rational design of therapeutic vaccines [3,4,6] .

Through the use of immunoinformatics tools, a total of 1027 epitopes were initially predicted from five LCINS-associated proteins: EGFR, KRAS, HER2, BRAF, and MET. Following rigorous screening based on HLA-binding affinity, antigenicity (via VaxiJen), predicted safety (non-toxic and non-allergenic), and cytokine-inducing potential, only 16 epitopes (8 MHC-I, 4 MHC-II, and 4 B-cell) were retained. This low selection ratio (~1.56%) underscores the high stringency of the evaluation pipeline, and reinforces the feasibility of using computational methods to preselect the most promising candidates for downstream validation.

Notably, several of the selected epitopes demonstrated strong binding to widely prevalent HLA alleles and high predicted antigenicity scores, including *DSRPKFREL* (EGFR) and *RLKETKDGF* (MET) for MHC class I, and *RPRFRELVSEFSRMA* (EGFR) for MHC class II. The ability of MHC-II epitopes to induce cytokines, such as IFN-γ and IL-4, predicted using sequence-based machine learning models, suggests potential for robust T-cell activation, pending future functional validation.

Population coverage analysis, performed using the IEDB Population Coverage Tool, indicated a global coverage of 81.23% with particularly high representation in Europe, and North America, West Indies. Coverage in Vietnam (39.36%) and Southeast Asia (43.78%) was moderate, suggesting that inclusion of additional alleles common in these regions may enhance local relevance in future studies.

Compared to previous vaccine efforts, such as TG4010 (targeting MUC1) or GVAX (based on whole-cell tumor platforms), this study adopts a mutation-guided reverse-design strategy that prioritizes molecular specificity and potential for personalization [17,18] . While these findings are preliminary and limited to in silico predictions, they offer valuable insights into the epitope landscape of LCINS and form the basis for future investigations into structure-based vaccine modeling, molecular docking, and laboratory validation.

The discrepancy in population coverage stems from ethnic variations in HLA allele frequencies. Our selection prioritized globally prevalent alleles for broad representativeness, inadvertently excluding East Asian-specific variants like HLA-A11:01 and HLA-B46:01. This resulted in lower coverage for Vietnam and Southeast Asia. We acknowledge this limitation and plan to incorporate region-specific alleles

in future analyses to improve accuracy and immunological relevance.

This in silico study identified promising multi-epitope vaccine candidates for lung cancer in never-smokers but has key limitations: (1) Limited HLA allele diversity: The selected alleles primarily represent Western populations, contributing to reduced predicted population coverage in Vietnam and Southeast Asia, and (2) Lack of experimental validation: Assessments of antigenicity (VaxiJen), allergenicity (AllerTOP), and toxicity (ToxinPred) remain computational predictions without wet-lab confirmation. To enhance the translational relevance and scientific robustness of this work, future efforts will focus on expanding the HLA dataset to include alleles prevalent in Vietnamese and Southeast Asian populations, followed by in vitro and ex vivo assays to confirm immunogenicity, safety, and overall vaccine potential.

CONCLUSION:

This study employed a comprehensive immunoinformatics framework to identify and evaluate potential B-cell and T-cell epitopes derived from five oncogenic driver proteins commonly mutated in lung cancer in never-smokers (LCINS). Out of 1,027 predicted candidates, 16 high-priority epitopes were selected based on stringent criteria, including strong HLA-binding affinity, high antigenicity, predicted safety (non-toxicity and non-allergenicity), and cytokine-inducing potential for MHC-II alleles. These epitopes represent promising immunological targets that may serve as a foundation for developing multi-epitope therapeutic vaccines specifically tailored to the LCINS population. Notably, several epitopes exhibited favorable binding to globally prevalent HLA alleles and demonstrated high predicted population coverage across multiple geographic regions. While the findings remain entirely in the in silico phase, they provide a solid basis for subsequent stages of experimental validation. Future work should include docking and molecular dynamics simulations to assess epitope–HLA binding affinity and structural stability, followed by constructing a multi-epitope vaccine and rigorously evaluating its immunogenicity and safety through in vitro assays, in vivo studies in relevant animal models, and organoid-based testing. These targeted steps will be essential for bridging the gap between computational predictions and meaningful translational clinical application.

LIST OF ABBREVIATIONS:

LCINS: Lung cancer in never-smokers
EGFR: Epidermal growth factor receptor
KRAS: KRAS proto-oncogene
HER2/ERBB2: human epidermal growth factor receptor 2
BRAF: B-Raf proto-oncogene
MET: MET proto-oncogene

SUPPLEMENTARY DATA:

Data sheets supporting information

AUTHORS' CONTRIBUTIONS

D.K.N.: Conceptualization, Data curation, Methodology, Formal analysis, Validation, Writing – review and editing. **X.H.L.**: Conceptualization, Data curation, Methodology, Formal analysis. **V.A.N.H.**: Data curation, Methodology, Formal analysis, Validation. **B.L.N.T.**: Data curation, Methodology, Formal analysis. **N.G.T.**: Data curation, Methodology, Visualization. **V.H.T.**: Data curation, Methodology, Formal analysis. **C.B.B.**: Data curation, Methodology, Formal analysis. **T.L.N.**: Project administration, Resources, Supervision, Funding acquisition, Investigation, Validation, Writing – original draft, Writing – review and editing.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known financial interests or personal relationships that could have appeared to influence the work reported in this paper.

RESEARCH ETHICS

This study did not involve human participants, animals, or identifiable personal data; therefore, ethical approval was not required.

ACKNOWLEDGEMENTS

This research is funded by Vietnam National University, Ho Chi Minh City, Vietnam (VNUHCM) under grant number C2026-44-12.

REFERENCES

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–63. Available from: https://10.3322/caac.21834.
2. Nguyen VT, Ho TD. Risk Factors of Lung Cancer Patients: A Survival Analysis with R. Journal of Science and Technology Development. 2020;23(3):660–9. Available from: https://10.32508/stdj.v23i3.1794.
3. Pelosof L, Ahn C, Gao A, Horn L, Madrigales A, Cox J, et al. Proportion of never-smoker noncell lung cancer patients at three diverse institutions. J Natl Cancer Inst. 2017;109(7). Available from: https://10.1093/jnci/djw295.
4. LoPiccolo J, Gusev A, Christiani DC, Jänne PA. Lung cancer in patients who have never smoked - an emerging disease. Nat Rev Clin Oncol. 2024;21(2):121–46. Available from: https://10.1038/s41571-023-00844-0.

5. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023;73(1):17–48. Available from: https://10.3322/caac.21763.
6. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012;150(6):1121–34. Available from: https://10.1016/j.cell.2012.08.024.
7. Awad MM, Oxnard GR, Jackman DM, Savukoski DO, Hall D, Shivdasani P, et al. MET exon 14 mutations in non-cell lung cancer are associated with advanced age and stage-dependent MET genomic amplification and c-Met overexpression. J Clin Oncol. 2016;34(7):721–30. Available from: https://10.1200/JCO.2015.63.4600.
8. Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, et al. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med. 2009;361(10):958–67. Available from: https://10.1056/NEJMoa0904554.
9. Paik PK, Arcila ME, Fara M, Sima CS, Miller VA, Kris MG, et al. Clinical characteristics of patients with lung adenocarcinomas harboring BRAF mutations. J Clin Oncol. 2011;29(15):2046–51. Available from: https://10.1200/JCO.2010.33.1280.
10. Mazières J, Peters S, Lepage B, Cortot AB, Barlesi F, Beau-Faller M, et al. Lung cancer that harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives. J Clin Oncol. 2013;31(16):1997–2003. Available from: https://10.1200/JCO.2012.45.6095.
11. Canon J, Rex K, Saiki AY, Mohr C, Cooke K, Bagal D, et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. Nature. 2019;575(7781):217–23. Available from: https://10.1038/s41586-019-1694-1.
12. Skoulidis F, Li BT, Dy GK, Price TJ, Falchook GS, Wolf J, et al. Sotorasib for lung cancers with KRAS p. G12C mutation. N Engl J Med. 2021;384(25):2371–81. Available from: https://10.1056/NEJMoa2103695.
13. Yang JC, Ahn MJ, Kim DW, Ramalingam SS, Sequist LV, Su WC, et al. Osimertinib in pretreated T790M-positive advanced non-cell lung cancer: AURA study phase II extension component. J Clin Oncol. 2017;35(12):1288–96. Available from: https://10.1200/JCO.2016.70.3223.
14. Thress KS, Paweletz CP, Felip E, Cho BC, Stetson D, Dougherty B, et al. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. Nat Med. 2015;21(6):560–2. Available from: https://10.1038/nm.3854.
15. Wolf J, Seto T, Han JY, Reguart N, Garon EB, Groen HJ, et al. Capmatinib in MET exon 14or MET-amplified non-cell lung cancer. N Engl J Med. 2020;383(10):944–57. Available from: https://10.1056/NEJMoa2002787.
16. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science. 2015;348(6230):124–8. Available from: https://10.1126/science.aaa1348.
17. Nemunaitis J, Jahan T, Ross H, Sterman D, Richards D, Fox B, et al. Phase 1/2 trial of autologous tumor mixed with an allogeneic GVAX vaccine in advanced-stage non-small-cell lung cancer. Cancer Gene Ther. 2006;13(6):555–62. Available from: https://10.1038/sj.cgt.7700922.
18. Quoix E, Lena H, Losonczy G, Forget F, Chouaid C, Papai Z, et al. TG4010 immunotherapy and first-line chemotherapy for advanced non-small-cell lung cancer (TIME): results from the phase 2b part of a randomised, double-blind, placebo-controlled, phase 2b/3 trial. Lancet Oncol. 2016;17(2):212–23. Available from: https://10.1016/S1470-2045(15)00483-0.
19. Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, et al. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2023;51:523–31. Available from: https://10.1093/nar/gkac1052.
20. Nguyen TL, Kim H. Identification and evaluation of tumor pyroptosis-associated antigens for design a vaccine candidate against lung cancer. 2024;Available from: https://doi.org/10.21203/rs.3.rs-5329441/v1.
21. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. Immunome Res. 2006;2(1):2. Available from: https://10.1186/1745-7580-2-2.
22. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. 2020;48:449–54. Available from: https://10.1093/nar/gkaa379.
23. Nilsson JB, Kaabinejadian S, Yari H, Kester MG, van Balen P, Hildebrand WH, et al. Accurate prediction of HLA class II antigen presentation across all loci using tailored data acquisition and refined machine learning. Sci Adv. 2023;9(47). Available from: https://10.1126/sciadv.adj6367.
24. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics. 2007;8(1):4. Available from: https://10.1186/1471-2105-8-4.
25. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GP, et al. In silico approach for predicting toxicity of peptides and proteins. PLoS One. 2013;8(9). Available from: https://10.1371/journal.pone.0073957.
26. Dimitrov I, Flower DR, Doytchinova I. AllerTOP—a server for in silico prediction of allergens. BMC Bioinformatics. 2013;14(Suppl 6):4. Available from: https://10.1186/1471-2105-14-S6-S4.
27. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. Biol Direct. 2013;8(1):30. Available from: https://10.1186/1745-6150-8-30.
28. Nagpal G, Usmani SS, Dhanda SK, Kaur H, Singh S, Sharma M, et al. Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. Sci Rep. 2017;7(1):42851. Available from: https://10.1038/srep42851.
29. Dhanda SK, Gupta S, Vir P, Raghava GP. Prediction of IL4 inducing peptides. Clin Dev Immunol. 2013;2013(1).
30. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics. 2006;7(1):153. Available from: https://10.1186/1471-2105-7-153.

VNUHCM Journal of Health Sciences 2026, 7(1):920-927

Open Access Full Text Article

Bài báo nghiên cứu

Xác định và đánh giá các epitope đặc hiệu từ các oncogene liên quan LCINS dựa trên tin sinh miễn dịch nhằm định hướng thiết kế vắc-xin điều trị

Duy-Khang Nguyen [1,2], Xuan-Huy Le [1,2], Vi-Anh Nguyen-Hoang [1,2], Bao-Long Nguyen-Tran [1,2], Ngan-Giang Truong [1,2], Vinh-Hy Truong [1,2], Chi-Bao Bui [1,2], Truc Ly Nguyen [1,2,*]

[1]*University of Health Sciences, Vietnam National University, Ho Chi Minh City, Vietnam*

[2]*Vietnam National University, Ho Chi Minh City, Vietnam*

Liên hệ

Truc Ly Nguyen , University of Health Sciences, Vietnam National University, Ho Chi Minh City, Vietnam

Vietnam National University, Ho Chi Minh City, Vietnam

Email: nttly@uhsvnu.edu.vn

Lịch sử

- Ngày nhận: 03-07-2025
- Ngày sửa đổi: 19-12-2025
- Ngày chấp nhận: 13-04-2026
- Ngày đăng: 12-06-2026

DOI : 10.32508/vnuhcmj-hs.v7i1.689

TÓM TẮT

Tổng quan: Ung thư phổi ở người không hút thuốc (Lung Cancer in Never-Smokers - LCINS) là một phân nhóm bệnh đang gia tăng với các đặc điểm lâm sàng và phân tử riêng biệt, đặc biệt phổ biến ở các quần thể Đông Á. Mặc dù tỷ lệ mắc ngày càng tăng, hiện vẫn chưa có các chiến lược miễn dịch trị liệu đặc hiệu kháng nguyên hoặc vắc-xin nhắm trúng đích dành riêng cho LCINS. Nghiên cứu này nhằm xác định và đánh giá các epitope có khả năng sinh miễn dịch bắt nguồn từ các oncogene chủ đạo trong LCINS, làm cơ sở bước đầu cho thiết kế vắc-xin ung thư đa epitope.

Phương pháp: Trình tự protein của năm oncogene liên quan đến LCINS (EGFR, KRAS, HER2, BRAF và MET) được thu thập từ cơ sở dữ liệu UniProtKB và phân tích thông qua quy trình tin sinh miễn dịch đa tầng tích hợp. Các epitope tế bào B, epitope liên kết với phân tử phức hợp hòa hợp mô chính lớp I (MHC-I) và lớp II (MHC-II) được dự đoán bằng các công cụ chuyên dụng như NetMHCpan, NetMHCIIpan và BepiPred. Các epitope tiềm năng được sàng lọc dựa trên ái lực liên kết với HLA, tính sinh kháng nguyên (đánh giá bằng VaxiJen), độc tính và khả năng gây dị ứng. Ngoài ra, khả năng của các epitope MHC-II trong việc kích thích sản xuất cytokine (IFN-γ, IL-4, IL-10) cũng được đánh giá nhằm dự đoán tiềm năng hoạt hóa đáp ứng miễn dịch qua trung gian tế bào T hỗ trợ. Phân tích độ bao phủ quần thể được thực hiện bằng công cụ Population Coverage của IEDB.

Kết quả: Từ tổng số 1027 epitope dự đoán ban đầu, có 16 epitope (gồm 8 epitope MHC-I, 4 epitope MHC-II và 4 epitope tế bào B) đáp ứng đầy đủ các tiêu chí lựa chọn. Các epitope này thể hiện tính sinh kháng nguyên cao, khả năng liên kết mạnh với HLA và độ bao phủ quần thể toàn cầu đạt 81,23%. Đáng chú ý, một số epitope cho thấy tính đặc hiệu cao đối với các đột biến sinh ung đặc trưng của LCINS, gợi ý tiềm năng ứng dụng trong miễn dịch trị liệu chính xác.

Kết luận: Nghiên cứu này đề xuất một cách tiếp cận hợp lý dựa trên tin sinh miễn dịch nhằm xác định và ưu tiên các epitope tiềm năng từ các oncogene liên quan LCINS. Các kết quả thu được cung cấp nền tảng khoa học quan trọng cho việc thiết kế và đánh giá tiền lâm sàng các vắc-xin ung thư đa epitope mang tính cá thể hóa, hướng tới các chiến lược miễn dịch trị liệu chính xác cho LCINS.

Từ khoá: Ung thư phổi ở người không hút thuốc (LCINS), tin sinh miễn dịch, dự đoán epitope, kháng nguyên liên quan khối u, liên kết HLA, độ bao phủ quần thể.

Trích dẫn bài báo này: D N, X L, V N, B N, N T, V T, C B, T L N. **Xác định và đánh giá các epitope đặc hiệu từ các oncogene liên quan LCINS dựa trên tin sinh miễn dịch nhằm định hướng thiết kế vắc-xin điều trị**.*VNUHCM J. Health Sci. 2026; 7(1):920-927.*