DOI:10.31557/APJCP.2026.27.9.3315
Cancer Research Trends in East and Southeast Asia

RESEARCH ARTICLE

Editorial Process: Submission:30/12/2025 Acceptance:02/09/2026 Published:27/09/2026

Regional Differences in Cancer Research Trends in East and Southeast Asia Based on Analyses of the PubMed Registry

SongYi Han[1], Tae-Min Kim[2,3]*

Abstract

Objective: Cancer represents a major and growing public health burden across Asia, yet cancer research capacity, priorities, and methodologies vary substantially between East Asia (EA) and Southeast Asia (SEA). How scientific activity has evolved across these regions over time remains incompletely characterized. **Methods:** We conducted a large-scale bibliometric and text-mining analysis of PubMed-indexed cancer-related publications from 2000 to 2025, covering 18 countries and territories in EA and SEA. Publication metadata, Medical Subject Headings (MeSH), author affiliations, and country information were systematically extracted and harmonized. Temporal publication trends were assessed using raw counts and GDP-normalized metrics. Tumor-type emphasis and methodological focus were examined using MeSH-based analyses. Latent research themes were identified using non-negative matrix factorization (NMF). International collaboration patterns were reconstructed through country-level co-affiliation networks across three time periods. **Results:** Cancer research output increased markedly across both regions, with EA accounting for substantially higher absolute publication volumes, driven primarily by China, Japan, and Korea. After normalization by GDP, several smaller but research-intensive systems—including Korea, Taiwan, Singapore, and Vietnam demonstrated high productivity relative to economic scale. Tumor-type analyses revealed strong alignment between research focus and regional cancer epidemiology, including liver and cholangiocarcinoma in SEA and gastric and lung cancer in EA. Methodological comparisons showed a clear regional divergence: EA research emphasized molecular, mechanistic, and translational approaches, whereas SEA research was enriched for cohort-based, epidemiologic, and public–health-oriented studies. Topic modeling and collaboration networks further highlighted complementary regional strengths and increasing global integration over time. **Conclusions:** Cancer research in EA and SEA has expanded and diversified over the past 25 years, reflecting regional differences in disease burden, infrastructure, and policy priorities. Our findings underscore the importance of targeted public health strategies, sustained investment in regionally relevant cancer research capacity, and strengthened international collaboration frameworks to promote more equitable and effective cancer control across Asia.

Keywords: Cancer, Prevalence, East Asia, Southeast Asia, Bibliographic analysis

Asian Pac J Cancer Prev, **27 (9)**, 3315-3324

Introduction

Cancer has become one of the most pressing public health challenges worldwide, responsible for nearly 10 million deaths each year [1]. However, the epidemiology and burden of cancer vary substantially across regions, reflecting differences in lifestyle, environmental exposure, infection prevalence, and socioeconomic development [2]. East and Southeast Asia encompass countries at markedly different stages of development from high-income economies to emerging nations, but all face a rising cancer burden [3]. In this region, rapid economic growth, urbanization, and ecological changes over the past two decades have profoundly reshaped disease patterns [4, 5]. Understanding how scientific attention to cancer has evolved within this diverse context is critical to identify unmet research needs and promote balanced, equitable progress in cancer control.

Bibliometric analyses using PubMed data offer an objective means to quantify research activity and to trace how scientific priorities shift over time [6]. PubMed-based analyses are particularly advantageous for biomedical research due to structured MeSH (medical subject headings) indexing and standardized metadata fields [7]. Although global biomedical output has expanded exponentially, relatively few comparative studies have systematically examined the temporal and regional trajectories of cancer research in East and Southeast Asia. Earlier bibliometric investigations have primarily focused on global publication volumes or citation

*[1]East Asian Studies, Ewha Womans University, Seoul, Korea. [2]Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, Korea. [3]Cancer Research Institute, College of Medicine, The Catholic University of Korea, Seoul, Korea. *For Correspondence: tmkim@catholic.ac.kr*

impacts [8, 9] overlooking the substantial heterogeneity in national research capacity, funding mechanisms, and networks within Asia. Consequently, regional disparities in research productivity, thematic focus, and international collaboration remain largely unexplored [10].

To address these gaps, we constructed a comprehensive PubMed-based registry of cancer-related publications spanning 2000–2025, encompassing 18 countries and territories across East and Southeast Asia. Using this dataset, we applied quantitative bibliometric and text-mining approaches to investigate how the scale, thematic composition, and methodological orientation of cancer research have evolved over time. Temporal analyses were used to assess publication growth, while non-negative matrix factorization (NMF) was employed to uncover latent research domains [11]. MeSH terms were analyzed to delineate tumor-type focus and methodological trends, enabling cross-country comparisons of disease emphasis and study design. Finally, international co-authorship networks were reconstructed to explore the evolution of regional and global research partnerships [12, 13]. By integrating bibliometric, textual, and network analyses, this study provides a data-driven overview of 25 years of cancer research across East and Southeast Asian countries. Our findings reveal how research priorities, methodological strategies, and international collaborations have diversified in parallel with socioeconomic transitions, offering new insights into the evolving landscape of cancer research and informing strategies for more equitable, synergistic progress across the Asian region.

Materials and Methods

Data source and study population

We conducted a large-scale bibliometric and text-mining study of cancer-related publications from 18 East and Southeast Asian countries and territories from 2000, when a full-scale database of these was established, to 2025. PubMed is a database of academic papers in medicine and life science, helping researchers stay informed of the latest and previous research (https://pubmed.ncbi.nlm.nih.gov/about/). PubMed was selected as the primary data source due to its comprehensive biomedical coverage and controlled vocabulary indexing of MeSH (https://www.nlm.nih.gov/mesh/meshhome.html). All records were retrieved programmatically using R version 4.4.3 and the rentrez package (https://github.com/ropensci/rentrez/wiki), which provides structured API-based access to the NCBI Entrez system. Extracted metadata were PubMed ID, publication year, journal title, article title, MeSH descriptors, author affiliations, and country indicators. Countries were assigned based on affiliation strings, and standardized using the countrycode R package and ISO-3166 naming conventions. A comprehensive dictionary of geographical variants, historical names, multilingual forms, and abbreviated patterns was implemented to ensure accurate country recognition. The study included all cancer-related publications from the following countries: East Asia (seven countries): China, Japan, Republic of Korea, Taiwan, Hong Kong, Macao, Mongolia and Southeast Asia (11 countries): Brunei Darussalam, Cambodia, Indonesia, Lao PDR, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste, Viet Nam. Any publication containing ≥1 affiliation from these countries was included. After data cleaning, deduplication, and harmonization, the final dataset consisted of >900,000 unique PubMed records. The PubMed search was performed on October, 2025, and all records indexed up to that date were included. As a result, publication counts for 2025 reflect incomplete year-to-date data and should be interpreted with caution in temporal comparisons. To complement publication counts with impact-oriented indicators, article-level citation counts linked from OpenAlex were used to calculate total citations, mean and median citations per article, and H-index at the country level. Because citation accrual depends on publication age, citation summaries were presented both for publications up to 2023 only.

Identification and processing of MeSH terms

MeSH descriptors were decomposed into individual terms using the tidyr R package, harmonized by removing subheadings (e.g., "/genetics"), and normalized to title case. To identify tumor types, we selected terms ending in "neoplasms" or containing canonical oncologic entities (e.g., carcinoma, sarcoma, lymphoma, leukemia, melanoma, glioma, and adenocarcinoma) following established MeSH ontology conventions. Each unique PMID (PubMed ID)–tumor-type pair was counted once to avoid inflation from multiple subheadings. Tumor-type frequencies were summarized for each country. Within-country proportional representation was computed by normalizing counts to total tumor-related MeSH terms. The top 20 globally frequent tumor types were visualized, enabling cross-country comparison of dominant cancer sites. For methodological analyses, MeSH descriptors indicating study design, laboratory techniques, or analytic frameworks were extracted (e.g., retrospective studies, polymerase chain reaction, immunohistochemistry), guided by prior methodological bibliometric studies [14]. Regional differences were quantified using normalized proportional frequencies, and the 25 most discriminative terms are presented.

Topic modeling using deconvolution

To characterize latent thematic structures in cancer research, we constructed a binary MeSH document–term matrix in which each publication was encoded by presence/absence of standardized MeSH descriptors. Terms appearing in fewer than 50 publications were removed to minimize data sparsity. We applied non-negative matrix factorization (NMF), a widely used unsupervised method for biomedical topic discovery, using the NMF R package with rank K = 30, Brunet update, and 200 iterations. NMF decomposed the corpus into a W matrix (documents × topics) as topic weights per publication and an H matrix (topics × MeSH terms) as term loadings defining each topic. Each publication was assigned to East Asia or Southeast Asia based on its affiliation-derived country. Regional enrichment for each topic was computed as the normalized difference in mean topic weights between East Asia or Southeast Asia. The five most enriched topics per

region were visualized along with their top ten defining MeSH terms from the H matrix.

Measurement of international collaboration

Author affiliations were parsed to identify country names and resolve common abbreviations or historical variants. For each PubMed ID, all unique countries listed in the affiliation field were extracted. Undirected collaboration edges were formed between every pair of countries co-appearing in the same publication. Edge weights represented the number of shared PubMed ID. Countries were grouped into East Asia and Southeast Asia as defined above. Collaboration networks were aggregated for three time periods - 2000–2009, 2010–2019, and 2020–2025 - and only edges linking each Asian region to non-Asian international partners were retained. The top 15 partner countries for each region and each period were visualized using bar plots.

Robustness of the results

To improve validity and robustness, we harmonized records to a canonical schema, removed duplicate PMIDs, standardized and cleaned MeSH descriptors by removing major-topic markers and subheadings, excluded overly general tumor labels, and prevented duplicate PMID-country-term counting. For methodological and topic analyses, very rare and overly ubiquitous terms were filtered to reduce noise. In addition, publication trends were examined using both raw counts and GDP-normalized counts as a sensitivity analysis for cross-country comparability. For topic comparisons, NMF topic weights were normalized within region before East Asia–Southeast Asia comparisons.

Results

Temporal and regional trends in cancer research within East and Southeast Asia

We first examined temporal trends in cancer-related publications across East Asia and Southeast Asia over a 25-year period (Figure 1, top panels). Overall, East Asian countries had a substantially higher research output than Southeast Asian countries throughout the study window. China exhibited the most dramatic expansion, with the counts of cancer-related publications rising sharply after 2010 and surpassing 40,000 articles annually by the early 2020s. This rapid acceleration reflects large-scale national investments in biomedical science and research infrastructure during the past decade [15, 16]. Japan maintained consistently high output with moderate long-term growth, while South Korea and Taiwan showed steady and sustained increases. In Southeast Asia, research output was lower in absolute magnitude but demonstrated a clear upward trend. Singapore and Malaysia showed the largest and most sustained increases, with publication numbers rising steadily from the early 2000s to a peak around 2020. Thailand and Vietnam also exhibited marked expansion, particularly after 2010. Meanwhile, Indonesia, the Philippines, and Myanmar displayed more gradual growth. These trends highlight pronounced regional heterogeneity: East Asia is characterized by large-scale, high-volume research, whereas Southeast Asia shows emerging but uneven growth driven by a subset of rapidly

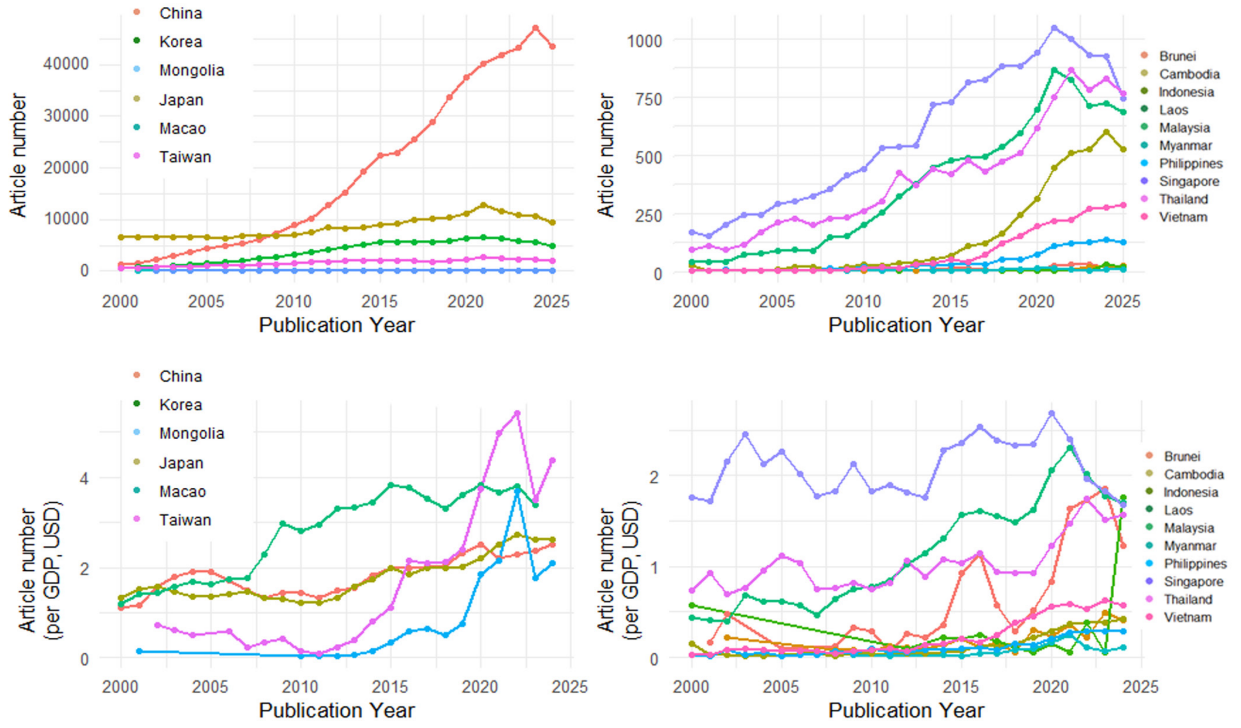

Figure 1. Temporal Trends in Cancer-Related Publications across East and Southeast Asia, 2000–2025. (Top panels) Annual counts of PubMed-indexed cancer-related publications are shown for 18 countries and territories in East Asia (left) and Southeast Asia (right). East Asian countries demonstrated markedly higher research output throughout the study period. (Bottom panels) Article annual counts are shown normalized by national gross domestic product (GDP; publications per billion USD). Data for East Asia are shown on the left and Southeast Asia on the right.

developing nations.

When adjusted for national economic scale, the comparative landscape of research productivity shifted considerably (Figure 1, bottom panels). Within East Asia, China's large absolute publication volume became less dominant when normalized by annual gross domestic production (GDP) plateauing around 2 to 3 articles per billion USD GDP. In contrast, South Korea and Taiwan produced approximately 4–5 articles per billion USD GDP by the mid-2020s. Japan demonstrated moderate but stable productivity, indicating that increases in research output in this country have not kept pace with long-term economic growth. Mongolia showed a marked rise in GDP-normalized output after 2015. In Southeast Asia, several countries demonstrated high efficiency despite modest absolute output. Singapore was consistently one of the most research-productive economies, maintaining more than 2 articles per billion USD GDP since the mid-2010s. Malaysia and Thailand exhibited gradual improvements in productivity, while Vietnam showed a notable increase in the 2010s. Meanwhile, Indonesia and the Philippines remained below 1 article per billion USD GDP. These patterns indicate that while economic size shapes total output, new knowledge and technologies created through research and development appear to have been shared and disseminated within Southeast Asia, leading to new achievements.

Citation-based indicators generally tracked publication-volume patterns, with China, Japan, and Korea exhibiting the largest total citation counts and H-indices, reflecting their leading roles in regional research output (Supplementary Table S1). At the same time, citation-based measures provided additional resolution beyond article counts alone. In East Asia, Hong Kong had the highest mean citations per article, while Korea and Taiwan showed relatively high median citation values, suggesting broader citation impact across publications rather than influence driven only by a small number of highly cited papers. In Southeast Asia, Singapore stood out with the highest regional values for total citations, mean citations per article, and H-index, and Malaysia also showed relatively strong citation performance. In contrast, countries such as Indonesia and Vietnam, although contributing increasing numbers of publications, had lower average citation-based impact metrics. Overall, these results indicate that publication volume and citation impact were aligned only partially, with several smaller but research-intensive systems showing disproportionately strong citation profiles relative to output size. To reduce bias from incomplete citation accrual in recent publications, these analyses were limited to articles published up to 2023.

Regional patterns of tumor types and methodologies in cancer research

Analysis of 25 years of PubMed-indexed cancer research revealed distinct cross-country patterns in tumor-type emphasis across 18 Asian countries examined (Figure 2A). The country-by-tumor-type heatmap demonstrated substantial heterogeneity in research focus that closely reflects known regional cancer epidemiology. Across the region, breast, lung, liver, stomach, and colorectal cancers emerged as the most frequently studied tumor types consistent with their high burden across Asia and their ranking among the leading causes of cancer morbidity and mortality globally [1].

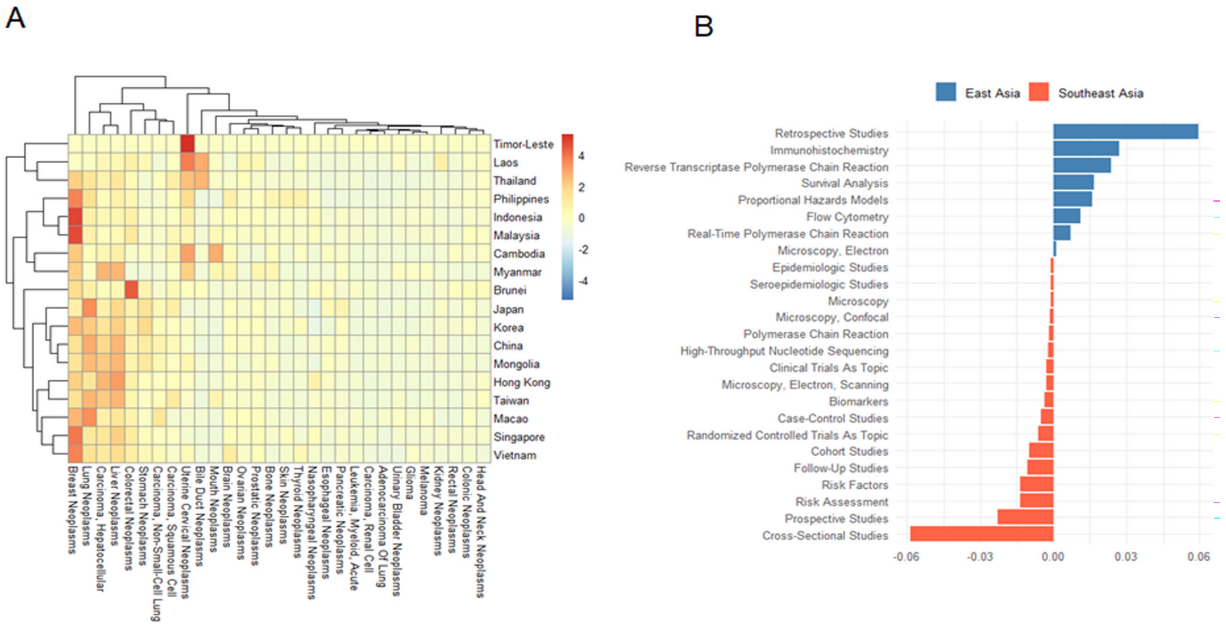

Figure 2. Cancer Research Trends In Terms of Tumor Types and Methodological Aspects. (A) Heatmap showing the relative distribution of the top 30 cancer types studied across 18 East and Southeast Asian countries. Values represent normalized publication frequencies per tumor type within each country. Countries and tumor types are clustered to highlight regional patterns, including a strong emphasis on breast, lung, liver, and colorectal cancers across most nations, with notable country-specific enrichment in several cancer types. (B) Differential analysis of methodological MeSH terms comparing cancer research practices between East Asia and Southeast Asia. Bars show the proportional differences in term usage.

East Asian countries, particularly China, Japan, and Korea, showed a strong research concentration in gastric cancers and hepatocellular carcinomas, mirroring the region's historically elevated incidence of these cancers attributable to Helicobacter pylori prevalence and dietary salt intake, alcohol consumption patterns, and hepatitis B virus (HBV)–related hepatocarcinogenesis [17]. Among the major tumor types, breast cancers were a strong focus of research in Southeast Asian countries. The heatmap also displayed research priorities shaped by infectious and environmental carcinogenic exposures in Southeast Asia. For instance, cervical cancers, driven largely by HPV infection, appeared more prominently in the research profiles of Cambodia, Myanmar, and Indonesia, consistent with persistent disparities in screening access and HPV vaccination uptake. Notably, cholangiocarcinoma, a hallmark cancer type of the Greater Mekong subregion, appeared as a distinct focus for Thailand and Laos, indicating successful reflection of local disease epidemiology in research priorities [18]. Countries with smaller research ecosystems such as Brunei, Laos, Cambodia, and Timor-Leste showed sparse distributions dominated by a few globally common cancers (breast, lung, liver). Overall, the tumor-type landscape captured by MeSH analysis demonstrates that cancer research in Asia closely follows the geographic distribution of cancer burden, probably reflecting population risk factors, infectious disease prevalence, screening policies, and national research capacity.

To compare how cancer research methodologies differ between East Asia and Southeast Asia, we analyzed the relative frequencies of MeSH descriptors related to study design, laboratory techniques, and analytic procedures after removing all tumor-type terms. A total of high-confidence methodological terms appearing at least 20 times across the dataset were included. For each term, proportional frequencies were computed for East Asian and Southeast Asian countries and shown as a differential bar plot (Figure 2B). The analysis revealed striking regional contrasts in methodological emphasis, suggesting divergent scientific priorities and research infrastructures across Asia. For example, East Asian cancer research was predominantly characterized by molecular, laboratory-based, and mechanistic approaches. Terms such as Retrospective Studies, Immunohistochemistry, Reverse Transcriptase Polymerase Chain Reaction, Proportional Hazards Models, Flow Cytometry, and Real-Time Polymerase Chain Reaction were strongly enriched in East Asia. In contrast, Southeast Asian cancer research demonstrated clear enrichment for population-level and observational epidemiologic designs. Terms such as Cross-Sectional Studies, Prospective Studies, Risk Assessment, Risk Factors, Case-Control Studies, Cohort Studies, and Follow-Up Studies were more frequent in Southeast Asia publications. This pattern indicates a strong focus on cancer surveillance, public-health epidemiology, environmental or infectious risk evaluation, and community-based cohort infrastructures. The high regional prevalence of infectious disease–associated malignancies (e.g., liver cancer linked to hepatitis, cervical cancer linked to HPV) may further contribute to this emphasis on epidemiologic methodology rather than experimental methods. Overall, the differential method usage highlights two complementary but distinct regional research profiles. East Asia's strong molecular and mechanistic orientation reflects its investment in laboratory infrastructure, translational oncology, and precision-medicine initiatives. Southeast Asia's emphasis on observational epidemiology indicates active national cancer registry development, environmental exposure research, and population-based cancer control efforts.

Thematic analysis by nonnegative matrix factorization

We applied non-negative matrix factorization (NMF) to word-frequency matrices derived from MeSH terms to decompose regional research activities into latent thematic components. NMF was selected because it generates additive, parts-based representations that are often easier to interpret in bibliometric applications than probabilistic alternatives such as latent Dirichlet allocation (LDA), while also providing topic weights that can be directly compared across countries and regions. The NMF-based topic decomposition identified 30 distinct research themes spanning molecular, clinical, epidemiologic, and translational aspects of cancer research (Figure 3). The topic modeling analysis revealed clear regional distinctions in the thematic composition of cancer research across East Asia and Southeast Asia, highlighting contrasting methodological emphases, disease foci, and experimental systems used in cancer research in the two regions.

East Asia enriched topics (Topic 13, 14, 20, 4, and 9) predominantly represent preclinical, mechanistic, and molecular oncology research. For example, topic 13 centers on cellular processes underlying tumor progression, such as cell movement, invasiveness, epithelial-mesenchymal transition, metastasis, and cadherin biology, suggesting a strong emphasis on molecular determinants of cancer aggressiveness. Topic 14 reflects extensive use of mouse xenograft models, including nude and inbred strains, tumor burden assays, and transplantation systems, indicating a major investment in in vivo experimental oncology platforms. Topic 20 is driven by terms such as lung neoplasms, adenocarcinoma, EGFR signaling, gefitinib, and quinazolines suggestive of research focus on EGFR-mutant lung cancer research, targeted therapies, and precision oncology. Topic 4 represents molecular assay technologies, including RT-PCR, real-time PCR, western blotting, transfection, and flow cytometry. This pattern reflects strong infrastructure for molecular diagnostics and gene-level functional assays. Topic 9 highlights gene expression regulation, neoplastic signaling, and lncRNA biology closely tied to transcriptomic and epigenetic cancer research. These topics reflect a consistent thematic profile: East Asian cancer research is oriented toward molecular, experimental, and preclinical discovery, supported by laboratory infrastructure and high publication volume in mechanistic studies.

The Southeast Asia–enriched topics (Topic 11, 15, 16, 3, and 8) reveal a distinct pattern focused on clinical, demographic, and therapeutic assessment, often rooted in population-level cancer burdens. For example,

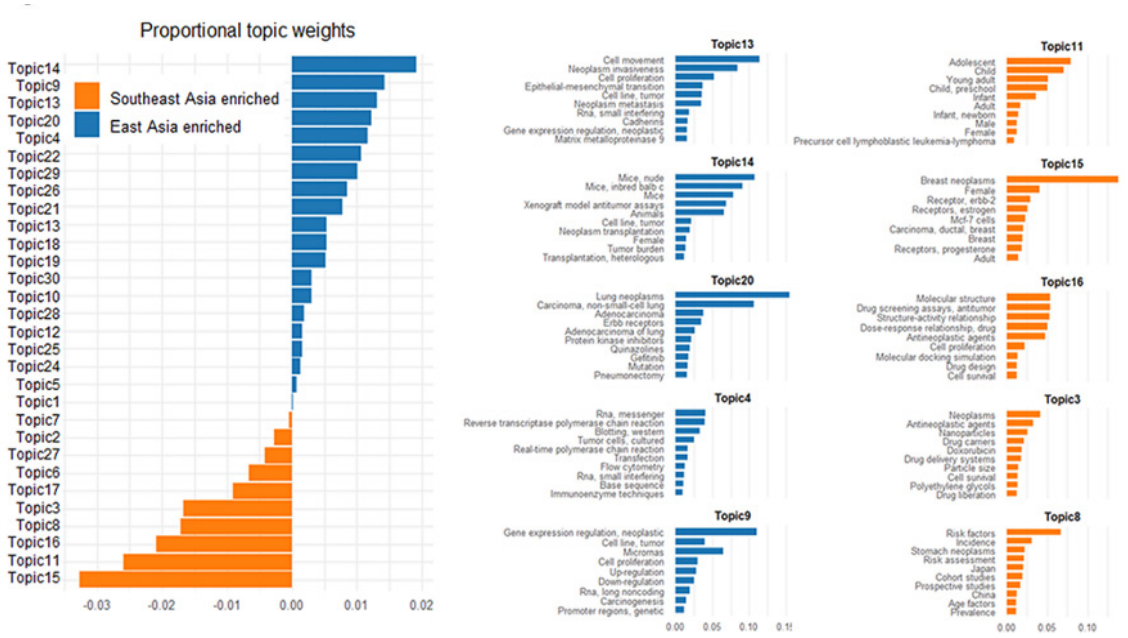

Figure 3. Regional Enrichment of Cancer Research Topics and Representative MeSH Terms in East Asia and Southeast Asia. (A) Differential topic enrichment between East Asia and Southeast Asia based on proportional topic weights derived from topic modeling of cancer-related publications. Positive values (blue) indicate topics significantly enriched in East Asia, whereas negative values (orange) reflect topics enriched in Southeast Asia. (B) Representative MeSH terms for the top region-enriched topics. Blue panels show the 10 highest-loading terms for the five most strongly East Asia–enriched topics, which primarily reflect molecular and mechanistic cancer biology. Orange panels show the 10 highest-loading terms for the five most strongly Southeast Asia–enriched topics, which emphasize clinical, epidemiologic, and translational themes.

topic 11 emphasizes pediatric, adolescent, and young adult populations, including leukemia and lymphoma, suggesting strong representation of clinical hematologic oncology in age-specific groups. Topic 15 is strongly associated with breast neoplasms, HER2/ER/PR receptor biology, MCF-7 cancer cell line models, and hormonal receptor characterization. This aligns with breast cancer being a major public health priority across Southeast Asia, where research often addresses population incidence, treatment response, and biomarker-guided therapy. Topic 16 reflects drug screening assays, structure-activity relationships, docking simulations, and antiproliferative agents, indicating active but more drug-development–oriented research. Topic 3 captures nanoparticles, drug carriers, particle size, and doxorubicin delivery reflecting Southeast Asian contributions to nanomedicine, drug delivery systems, and translational biomaterials. Topic 8 highlights risk factors, incidence, cohort studies, prevalence, and gastric or stomach neoplasms, reflecting epidemiological and clinical population research, including country-specific cancer burden analyses. Across these topics, Southeast Asia displays a research emphasis centered on clinical populations, cancer risk stratification, therapeutic evaluation, and translational drug development, rather than basic mechanistic work.

International collaboration patterns

To characterize how global research partnerships evolved over time, we analyzed country-level co-affiliations extracted from PubMed records and constructed regional international-collaboration profiles for East Asia and Southeast Asia across three periods: 2000–2009, 2010–2019, and 2020–2025 (Figure 4). The analysis highlights notable temporal shifts in collaboration intensity, partner diversity, and regional scientific integration. During the early 2000s, both East Asia and Southeast Asia exhibited modest but emerging global collaboration patterns. Notably, the United Kingdom appeared as the most frequent partner for both East Asia and Southeast Asia, surpassing even the United States during this period. This pattern was largely attributable to historically close affiliations involving Hong Kong and China. Secondary partners included United States, Canada, Germany, Australia, and several Western European countries, although East Asia displayed substantially larger collaboration volumes than Southeast Asia. The decade of 2010–2019 marked a major expansion of international collaboration for both regions. The United States emerged as the overwhelmingly dominant collaborator, far surpassing the United Kingdom and becoming the central hub for cancer research partnerships. East Asia demonstrated strong bilateral ties with major scientific nations including the United States, United Kingdom, Germany, Canada, Australia, Italy, and France. This shift reflects broad integration of China, South Korea, and Japan into global biomedical consortia and large-scale multicenter studies. Southeast Asia also showed substantial growth, maintaining strong ties with the United States, United Kingdom, but increasingly collaborating with countries such as Germany, France, India, the Netherlands, and Spain. In the most recent period (2020-2025), East Asia and Southeast Asia exhibited further diversification and intensification of international research connections. East Asia remained

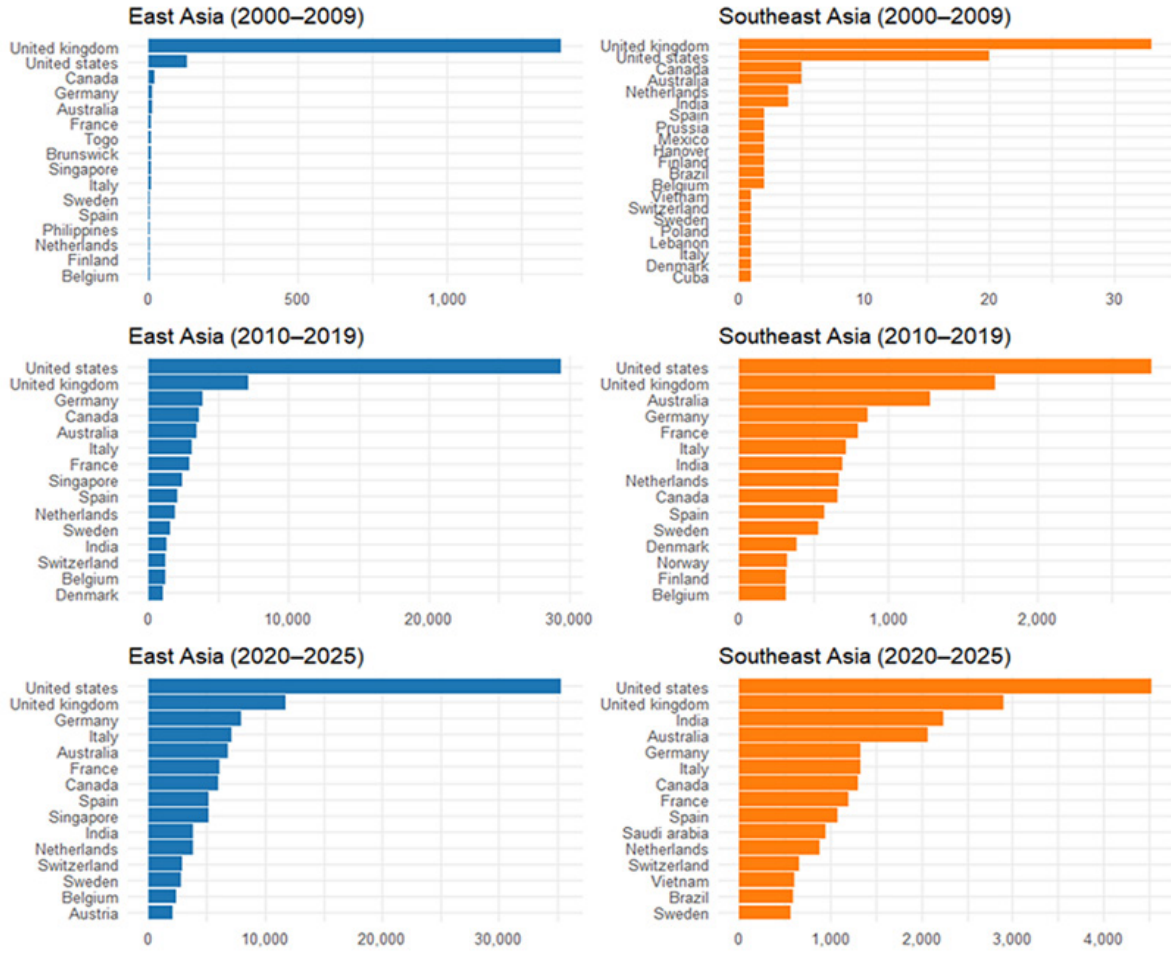

Figure 4. International collaboration networks of East and Southeast Asia. Bar plots show the top international partner countries collaborating with East Asia (left panels) and Southeast Asia (right panels), ranked by the number of co-authored PubMed-indexed cancer publications. Collaboration strengths were computed from country–country co-affiliation pairs extracted from institutional address fields. Three time periods of 2000–2009, 2010-2019 and 2020-2025 were investigated showing East Asia maintains larger absolute collaboration volumes, whereas Southeast Asia exhibits diversified and increasingly global connections.

globally dominant, with the United States as the primary partner followed by the United Kingdom, Germany, Italy, Australia, France, and Canada, reflecting continued growth of China and Korea in high-impact cancer research. Southeast Asia networks continued to broaden, with strong ties to the United States, United Kingdom and increasing collaboration with India, Germany, the Netherlands, Switzerland, and regional partners including Vietnam and Thailand. Across all periods, East Asia exhibited far higher collaboration volume and connectivity, consistent with its larger research output and infrastructure. However, Southeast Asia showed the fastest proportional growth, transitioning from sparse networks in the early 2000s to a highly interconnected position in recent years. These findings indicate that Asian cancer research has become increasingly globalized, with both East Asia and Southeast Asia deeply integrated into international research networks, albeit with differing magnitudes and partner compositions.

Discussion

This 25-year, PubMed-based analysis reveals pronounced structural and thematic contrasts in cancer research across East Asia and Southeast Asia. These differences reflect not only regional variation in cancer burden, but also long-standing disparities in research capacity, funding architecture, and potential health-system priorities in the two subregions. Asia already accounts for roughly half of global cancer deaths, underscoring the importance of understanding how research agendas have evolved in this part of the world.

Temporal trends show a steep and sustained expansion of cancer publications in East Asia, driven largely by China, Japan, and Korea, with more modest but accelerating growth in Southeast Asia. This pattern is consistent with global bibliometric work demonstrating that upper-middle and high-income countries now dominate oncology output and citation impact [19].

There is a substantial body of empirical research examining the relationship between research and development (R&D) activities and economic growth, with particular emphasis on R&D expenditure levels and patent outputs as key indicators of innovative capacity. Gross domestic product (GDP) is commonly adopted in this literature as a benchmark for national economic scale

and performance, allowing comparative assessment of how efficiently countries translate research investment into economic outcomes [20-22].

After normalization by national GDP, however, a different picture emerges: smaller but research-intensive systems such as Korea, Taiwan, Singapore, and, more recently, Mongolia and Vietnam, display disproportionately high productivity relative to economic size [10]. This suggests that targeted policy initiatives, academic–hospital integration, and focused investment in biomedical infrastructure can partly offset macro-economic constraints, although the relative contribution of each factor could not be directly assessed in the present study.

Our MeSH-based mapping of tumor types shows that country-level research emphases broadly mirror known epidemiologic profiles. In Southeast Asia, hepatocellular carcinoma and cholangiocarcinoma are relatively over-represented, aligning with the heavy burden of chronic hepatitis B and C infection [23] and the unique problem of liver-fluke–associated cholangiocarcinoma in the Mekong area [18, 24]. In East Asia, lung, gastric, and colorectal cancers dominate the research landscape, consistent with high smoking prevalence, dietary risk factors, and historically high incidence of upper-gastrointestinal malignancies in China, Japan, and Korea [25, 26]. The heat-map structure also highlights under-studied cancers in relatively lower-income Southeast Asian countries, suggesting gaps where disease burden may not yet be matched by research attention.

Methodological and thematic patterns captured through MeSH terms and NMF-derived topics further underscore a division of labor between the two subregions. East Asia shows clear enrichment of topics anchored in molecular oncology, experimental therapeutics, and translational research: terms related to signal transduction, transfection, xenograft models, and high-throughput assays cluster within East Asia-dominant topics. These findings are concordant with the rapid expansion of precision oncology and large-scale cancer genomics initiatives in Japan, Korea, and China, including national cancer genome projects and implementation of next-generation sequencing in routine clinical care. In contrast, Southeast Asia is enriched for cohort studies, case–control designs, risk-factor assessment, and cross-sectional surveys [27]. This methodological profile aligns with regional needs and constraints: many Southeast Asian countries are still consolidating basic cancer registration, screening programs, and treatment access, and policy-relevant evidence on risk factors, early detection, and health-system performance remains a priority [28]. Growing output in pharmaceutics, nanomedicine, and drug-delivery–related topics as detected in several NMF components also reflects the rise of engineering-driven and pharmaceutical sciences programs in countries such as Singapore, Thailand, and Malaysia. These differences should not be interpreted hierarchically. Rather, they represent complementary strengths along the cancer-control continuum. Deep investment of East Asian countries in molecular and translational research generates candidate biomarkers, drug targets, and mechanistic insights, while Southeast Asian population-based and health-services research provides the context needed to implement those advances in real-world, resource-constrained settings. The potential for synergy is evident in areas such as molecular epidemiology, implementation of precision screening, and context-specific cost-effectiveness analyses, where multi-country collaborations could integrate East Asia-generated "omics" discoveries with Southeast Asia's diverse cohorts and health-system experiments.

Our collaboration-network analysis shows that both East Asia and Southeast Asia have become progressively more integrated into global oncology research. In the early 2000s, the United Kingdom was the leading partner for East Asia; this likely reflects historical ties and editorial affiliations related to Hong Kong and mainland Chinese institutions during that period. By the 2010s and 2020s, however, the United States was the dominant primary collaborator for both East Asia and Southeast Asia, with the United Kingdom, Germany, Canada, Australia, and several European countries forming a secondary tier of partners. These patterns mirror broader global trends in biomedical collaboration, where trans-Pacific and Euro-Asian partnerships have intensified, but intra-regional collaboration within Asia has remained comparatively modest [29]. Strengthening East Asia and Southeast Asia research ties through shared biobanks, harmonized registries and clinical trial networks, interoperable data platforms, and regional training programs could help redistribute expertise and ensure that advances in precision oncology are equitably tested and implemented across the continent.

Several limitations warrant consideration. First, reliance on PubMed and MeSH indexing may under-represent locally published, non-English, or non-indexed journals, particularly in lower-income Southeast Asia countries. This may be especially relevant to the observed methodological divide, as some high-quality local epidemiological and public health research may be disseminated through regional journals not captured by PubMed. Accordingly, the observed differences in publication volume and methodological emphasis should be interpreted as patterns within the PubMed-indexed literature rather than exhaustive representations of all cancer research activity in the region; Supplementary Table S2 summarizes the possible implications of this indexing bias for Southeast Asian research patterns. Second, our heuristic filters for tumor types and methodological terms inevitably introduced some classification uncertainty, although multiple preprocessing steps were used to improve robustness, including descriptor cleaning, exclusion of overly general labels, duplicate-control rules, and low-frequency term filtering. In addition, country-level publication trends were examined using both raw and GDP-normalized counts, and the broad regional patterns were preserved. Third, publication counts and topic loadings are proxies for research activity, not direct measures of impact on clinical outcomes or policy change.

Despite these caveats, this study provides, to our knowledge, one of the most comprehensive, data-driven portraits of cancer research in East Asia and Southeast Asia over the last quarter century. By

combining bibliometrics, text mining, topic modeling, and network analysis, we highlight how regional research ecosystems have diversified in response to shifting disease burdens, technological innovation, and globalization of science. Importantly, the findings argue against a one-size-fits-all model for "ideal" cancer research: East Asia's molecular and translational strengths and Southeast Asia's epidemiologic and public-health orientation are both essential and more deliberate integration of these complementary strengths may position Asia as a global leader in context-sensitive precision oncology.

Our findings have several practical implications. First, from a public health perspective, the marked heterogeneity in research output and thematic emphasis across Asian countries suggests that cancer-control strategies should be more closely aligned with local disease burden, implementation needs, and health-system capacity. Second, from a research funding perspective, the observed concentration of output in a limited number of countries and topics supports more strategic investment in underrepresented settings, cancer types, and methodological domains, including epidemiologic surveillance, translational research, and data infrastructure. Third, from an international collaboration perspective, future frameworks should move beyond simple co-authorship expansion toward more equitable partnerships that emphasize shared data resources, researcher training, regional consortia, and sustained capacity building, particularly for lower-resourced Southeast Asian countries.

In conclusion, this study shows that cancer research across East Asia and Southeast Asia has expanded along distinct but potentially complementary trajectories. Reducing regional inequities will require better alignment of research priorities with national cancer burden, more strategic investment in underrepresented settings and domains, and more equitable collaboration frameworks centered on shared infrastructure, training, and sustained regional partnerships.

Author Contribution Statement

S.H. designed the study, conducted data analysis and wrote drafts of the paper. T.M.K. conceived the idea, supervised the project and wrote drafts of the paper.

Acknowledgements

The authors sincerely thank all contributors whose invaluable insights and efforts were essential to the successful completion of this manuscript.

Human ethics and consent to participate declarations

Not applicable. This study used only publicly available bibliographic metadata and did not involve human participants, patient data, or identifiable personal information. Therefore, ethics committee approval and informed consent were not required.

Conflict of interest

The Authors declare that they have no conflict of interest.

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. https://doi.org/10.3322/caac.21834.
2. Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global cancer transitions according to The human development index (2008-2030): a population-based study. Lancet Oncol. 2012;13(8):790-801. https://doi.org/10.1016/s1470-2045(12)70211-5.
3. Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, Abdel-Rahman O, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: A systematic analysis for the Global Burden of Disease Study. JAMA Oncol. 2019;5(12):1749-68. https://doi.org/10.1001/jamaoncol.2019.2996.
4. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990-3. https://doi.org/10.1038/nature06536.
5. Dhillon PK, Jeemon P, Arora NK, Mathur P, Maskey M, Sukirna RD, et al. Status of epidemiology in the WHO South-East Asia Region: burden of disease, determinants of health and epidemiological research, workforce and training capacity. Int J Epidemiol. 2012;41(3):847-60. https://doi.org/10.1093/ije/dys046.
6. Manoj Kumar L, George RJ, P SA. Bibliometric analysis for medical research. Indian J Psychol Med. 2023;45(3):277-82. https://doi.org/10.1177/02537176221103617.
7. Lipscomb CE. Medical Subject Headings (MeSH). Bull Med Libr Assoc. 2000;88(3):265-6.
8. Cabral BP, da Graça Derengowski Fonseca M, Mota FB. The recent landscape of cancer research worldwide: a bibliometric and network analysis. Oncotarget. 2018;9(55):30474-84. https://doi.org/10.18632/oncotarget.25730.
9. Glynn RW, Chin JZ, Kerin MJ, Sweeney KJ. Representation of cancer in the medical literature--a bibliometric analysis. PLoS One. 2010;5(11):e13902. https://doi.org/10.1371/journal.pone.0013902.
10. King DA. The scientific impact of nations. Nature. 2004;430(6997):311-6. https://doi.org/10.1038/430311a.
11. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. 1999;401(6755):788-91. https://doi.org/10.1038/44565.
12. Leydesdorff L, Wagner CS. International collaboration in science and the formation of a core group. J Informetr. 2008;2(4):317-25. https://doi.org/10.1016/j.joi.2008.07.003
13. Newman ME. The structure of scientific collaboration networks. Proc Natl Acad Sci U S A. 2001;98(2):404-9. https://doi.org/10.1073/pnas.98.2.404.
14. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J. 2009;26(2):91-108. https://doi.org/10.1111/j.1471-1842.2009.00848.x.
15. Bazavan A. Chinese government's shifting role in the national innovation system. Technol Forecast Soc Change. 2019;148. https://doi.org/10.1016/j.techfore.2019.119738.
16. U.S. Commission on International Religious Freedom. 2025 annual report: China chapter [Internet]. Washington (DC): U.S. Commission on International Religious Freedom; 2025 [cited 2026 Mar 26]. Available from: https://www.uscirf.gov/sites/default/files/2025-03/2025 USCIRF Annual Report.pdf

17. GBD 2019 Asia and All Cancers Collaborators. Temporal patterns of cancer burden in asia, 1990-2019: A systematic examination for the global burden of disease 2019 study. Lancet Reg Health Southeast Asia. 2024;21:100333. https://doi.org/10.1016/j.lansea.2023.100333.
18. Li X, Guan R, Zhang S. Factors contributing to the high malignancy level of cholangiocarcinoma and its epidemiology: literature review and data. Biology (Basel). 2025;14(4). https://doi.org/10.3390/biology14040351.
19. Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references: Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references. J Assoc Inf Sci Technol. 2014;66(11). https://doi.org/10.1002/asi.23329.
20. Huňady J, Orviská M. The impact of research and development expenditures on innovation performance and economic growth of the country the empirical evidence. CBU Int Conf Proc. 2014;2:119-125. https://doi.org/10.12955/cbup.v2.454.
21. Guerrero M, Cunningham J, Urbano D. Economic impact of entrepreneurial universities' activities: an exploratory study of the United Kingdom. Res Policy. 2015;44:748. https://doi.org/10.1016/j.respol.2014.10.008.
22. Sterlacchini A. R&D, higher education and regional growth: Uneven linkages among European regions. Res Policy. 2008;37(6):1096-107. https://doi.org/10.1016/j.respol.2008.04.009
23. Ott JJ, Stevens GA, Groeger J, Wiersma ST. Global epidemiology of hepatitis B virus infection: new estimates of age-specific HBsAg seroprevalence and endemicity. Vaccine. 2012;30(12):2212-9. https://doi.org/10.1016/j.vaccine.2011.12.116.
24. Kim DY. Changing etiology and epidemiology of hepatocellular carcinoma: Asia and worldwide. J Liver Cancer. 2024;24(1):62-70. https://doi.org/10.17998/jlc.2024.03.13.
25. Inoue M, Tsugane S. Epidemiology of gastric cancer in Japan. Postgrad Med J. 2005;81(957):419-24. https://doi.org/10.1136/pgmj.2004.029330.
26. Lam WK, White NW, Chan-Yeung MM. Lung cancer epidemiology and risk factors in Asia and Africa. Int J Tuberc Lung Dis. 2004;8(9):1045-57.
27. Torre LA, Siegel RL, Ward EM, Jemal A. Global cancer incidence and mortality rates and trends an update. Cancer Epidemiol Biomarkers Prev. 2016;25(1):16-27. https://doi.org/10.1158/1055-9965.Epi-15-0578.
28. Anchalee N, Thinkhamrop K, Suwannatrai AT, Titapun A, Loilome W, Kelly M. Spatio-temporal analysis of cholangiocarcinoma in a high prevalence area of northeastern Thailand: a 10-year large scale screening program. Asian Pac J Cancer Prev. 2024;25(2):537-46. https://doi.org/10.31557/apjcp.2024.25.2.537.
29. Gaillard J. Measuring research and development in developing countries: main characteristics and implications for the Frascati Manual. Sci Technol Soc. 2010;15:77-111. https://doi.org/10.1177/097172180901500104.