Discover Oncology

Research

Research trends and hotspots of single nucleotide polymorphisms in endometrial cancer: a bibliometric analysis

Renjie Wang[1] · Zhihong Jia[1] · Liang Peng[2] · Jinghui Xu[1] · Qiying Zhu[1] · Yinghong Wu[1]

Received: 5 December 2024 / Accepted: 5 May 2025
Published online: 12 May 2025

Abstract

Background Endometrial cancer (EC) is a common gynecological malignancy with increasing incidence, especially in developed nations. Understanding genetic variations, particularly single nucleotide polymorphisms (SNPs), is crucial for uncovering the disease's pathogenesis, progression, and treatment responses. This study explores the global research landscape of SNPs in EC, focusing on field evolution, key contributors, and emerging trends.

Methods A systematic search of the Web of Science Core Collection (WoSCC) retrieved 838 publications on SNPs in EC from 1991 to 2024. Bibliometric indicators, including publication volume, citation counts, and keyword occurrences, were analyzed using VOSviewer, CiteSpace, and the R package "bibliometrix" for visual mapping and trend analysis.

Results The United States (230 publications) and China (182 publications) were leaders in research output. Harvard University and the National Cancer Institute were prominent contributors. Key themes included "microsatellite instability" (a hallmark of DNA mismatch repair deficiency) and "genome-wide association studies" (GWAS), identifying susceptibility loci like HNF1B and CYP19A1. Recent trends, such as "Mendelian randomization," have enhanced causal inference in risk factor studies. SNP research has advanced risk prediction models and personalized therapeutic strategies, such as hormone therapy tailored to genetic profiles.

Conclusion SNP research has deepened our understanding of EC's genetic basis, with a growing emphasis on Mendelian randomization and GWAS. These advancements have refined risk prediction and opened new avenues for personalized medicine. Integrating SNP data with environmental and hormonal factors remains crucial for advancing prevention, diagnosis, and treatment strategies in EC.

Keywords Endometrial cancer · Single nucleotide polymorphisms · Bibliometric analysis · Personalized medicine

1 Background

Endometrial cancer (EC) is one of the most common gynecological malignancies globally, with a particularly high incidence in developed countries [1]. According to recent reports, EC is the most prevalent gynecological cancer in many Western nations, accounting for approximately 7% of all cancers diagnosed in women [2]. The incidence rates have been steadily increasing over recent decades, especially among older women, although a notable trend of earlier onset among younger women has also been observed [3].

Renjie Wang and Zhihong Jia have contributed equally to this work.

✉ Yinghong Wu, wuyinghong84@163.com | [1]Department of Pathology, The Second People's Hospital of Jingdezhen, Jingdezhen 333000, Jiangxi, China. [2]Department of Gynecology, The Second People's Hospital of Jingdezhen, Jingdezhen 333000, Jiangxi, China.

EC arises from the endometrium, the inner lining of the uterus, and is typically associated with hormonal imbalances, particularly prolonged exposure to unopposed estrogen without the balancing effects of progesterone. This hormonal imbalance can lead to endometrial hyperplasia, which may eventually progress to malignancy [4]. While most EC cases are diagnosed in postmenopausal women, the rising incidence in premenopausal individuals highlights the necessity for ongoing awareness and early detection efforts [3]. The progression of EC can significantly impact a patient's quality of life, leading to symptoms such as abnormal uterine bleeding and pelvic pain, along with posing substantial health risks [5].

Recent advancements in genetic research have emphasized the significance of single nucleotide polymorphisms (SNPs) in understanding various cancers, including EC [6, 7]. SNPs are the most common form of genetic variation among individuals, involving alterations in single nucleotides at specific genomic locations. These genetic variations can have a significant impact on gene expression and function, potentially resulting in cancerous changes. For example, SNPs in the HNF1B gene have been associated with increased EC risk, while SNPs in CYP19A1, which influence estrogen biosynthesis, have been linked to hormone-driven EC in postmenopausal women [8]. Furthermore, SNP genotyping has been utilized to identify high-risk individuals early, enabling preventive interventions such as lifestyle modifications or earlier screening protocols [9]. Research has shown that specific SNPs may influence susceptibility to EC, affect its progression, and even dictate patient responses to treatment [8]. For instance, genetic profiling has guided personalized hormonal therapies by identifying patients more likely to benefit from treatments targeting estrogen pathways [10]. This burgeoning field holds promise for the future of personalized medicine, where SNP analysis could inform tailored diagnostic and therapeutic strategies based on individual genetic profiles.

Bibliometric analysis serves as a powerful tool to explore extensive literature, which could evaluate scientific literature by analyzing publication trends, citation patterns, and collaborative networks [11]. This method is essential for understanding the development of specific research fields, such as EC and SNPs. Previous bibliometric studies have illuminated various aspects of EC research, such as genetic risk factors, treatment strategies, and disease progression. For instance, bibliometric analyses of EC have highlighted significant advancements in molecular biology, particularly regarding microsatellite instability and genome-wide association studies (GWAS) [1]. Similarly, research focused on SNPs has underscored their increasing significance in cancer susceptibility and personalized medicine, showcasing key authors and institutions leading research in this area [8]. These studies have been instrumental in mapping the global research landscape, identifying influential publications and emerging trends.

Despite the growing interest in both EC and SNPs, there remains a notable gap in comprehensive bibliometric analyses that specifically address SNP research within the context of EC. For example, while studies have demonstrated the role of SNPs in refining risk prediction models or guiding hormone therapy [12], few studies have summarized the most prominent research hotspots. This study aims to fill that gap by conducting an in-depth bibliometric analysis of SNP studies related to EC, offering insights into the field's evolution and identifying key research trends.

2 Methods

2.1 Search strategies and data collection

The literature search was conducted using the Web of Science Core Collection (WoSCC), a leading database for scientific research [13], with the following search formula: TS = (("endometrial cancer*" OR "endometrial carcinoma*" OR "endometrial malignancy*" OR "endometrial malignant neoplasm*" OR "endometrial malignant tumor*" OR "cancer* of endometrium" OR "carcinoma* of endometrium" OR "malignancy* of endometrium" OR "malignant neoplasm* of endometrium" OR "malignant tumor* of endometrium")) AND TS = (SNP OR "single nucleotide polymorphism" OR "Single Nucleotide Polymorphisms" OR "single nucleotide polymer" OR "Haplotype-tagging SNPs" OR "gene polymorphism" OR "genotype" OR "polymorphism genetics" OR "genetic polymorphism") [14, 15]. The literature retrieval was conducted on a single day, June 14, 2024, to ensure consistency and minimize the impact of database updates. During the data extraction phase, two researchers independently conducted the literature search using the same formula. The inclusion criteria for the publications were: (1) specifically addressing SNPs in relation to EC; (2) document type being "article;" and (3) language being "English." Publications of other document types or in languages other than English were excluded.

2.2 Statistical analysis

Relevant data was extracted from the retrieved literature records, and Microsoft Excel was utilized to identify and compute bibliometric indicators. These indicators encompassed key aspects of publications, including annual publication volume, citation frequency, average citation rate, journal names, journal impact factors (IF), publishing countries/regions, institutions, and authors.

Three advanced bibliometric tools were employed for visual analysis, including VOSviewer (version 1.6.20), CiteSpace (version 6.3.R1), and the R (version 4.3.3) package "bibliometrix". VOSviewer was crucial for mapping institutional cooperation, author collaborations, co-authorships, citations, and co-citations, allowing for the visualization and exploration of complex networks of collaboration within the academic field [16]. This analysis provides deeper insights into the interconnections among authors, institutions, and publications. The co-occurrence maps provided insights into collaborative networks and research trends, where the size of the nodes represented the frequency of occurrence, and the thickness of the connecting lines indicated the strength of co-occurrence relationships.

The keyword co-occurrence analysis was conducted using VOSviewer, where the size of each node represented the number of publications, the thickness of the lines indicated the strength of the links, and the color of the nodes denoted different clusters or time periods. The keyword burst detection was performed using CiteSpace to identify emerging trends and research hotspots in the field, with parameters set for time slicing from January 1994 to June 2024. Node types were specified as keywords, with the threshold (top N per slice) set to 5. Pruning was configured to pathfinder plus pruning merged network.

To quantify the academic impact of individuals and journals, the H-index was employed. The H-index is a critical measure to evaluate the academic contributions of researchers and can predict their future scientific achievements. In this study, each author's H-index was obtained from the WoSCC. Additionally, the G-index and M-index were incorporated into the analysis. Journal IF, sourced from the Journal Citation Reports (JCR), provide a quantitative measure reflecting the average number of citations to articles published in journals, books, or symposium series [17]. The JCR also categorizes journals into quartiles (Q1–Q4) based on their field and impact factor, aiding researchers in identifying high-impact publication venues [18]. Furthermore, TC (Total Citations) and TP (Total Publications) were calculated to provide additional metrics for evaluating the academic impact of individuals and journals. TC represents the total number of citations received by an individual or journal, while TP represents the total number of publications produced by an individual or journal. These metrics were calculated using the bibliometric data extracted from the literature records and were used to supplement the analysis of academic impact.

3 Results

3.1 An overview of publications

From the initial 1025 records identified, publications with other document types, including reviews (n = 97), editorial materials (n = 6), letters (n = 4), meeting abstracts (n = 34), and others (n = 24), were excluded. Non-English publications (n = 22) were also excluded. No duplicate records were identified in the dataset, resulting in a total of 838 eligible publications analyzed in this study. The flowchart of the data screening process is presented in Fig. 1. A total of 292 sources contributed to 838 documents published between 1991 and 2024. with an annual growth rate of 7.54%. There are 5,721 authors involved in the analysis, while each document averages 9.78 co-authors. The international co-authorship rate was 27.09%. The total number of references is 22,077, demonstrating the extensive literature base utilized. The average age of the documents is 12.9 years and each document averages 31.13 citations.

The publication trends in EC research with a focus on SNPs over the past 3 decades showed an overall fluctuating upward trend **(**Fig. 2**)**. Starting with a single publication in 1991, the number of papers gradually increased, reaching 12 by 2000. This growth continued steadily, peaking at 50 publications in 2011. However, after this peak, a notable decline was observed, with the number of publications dropping to 23 in 2017. Despite this decrease, the output remained stable in the following years, and a recovery was observed after 2020, resulted in a recent peak in 2022 with 43 publications.

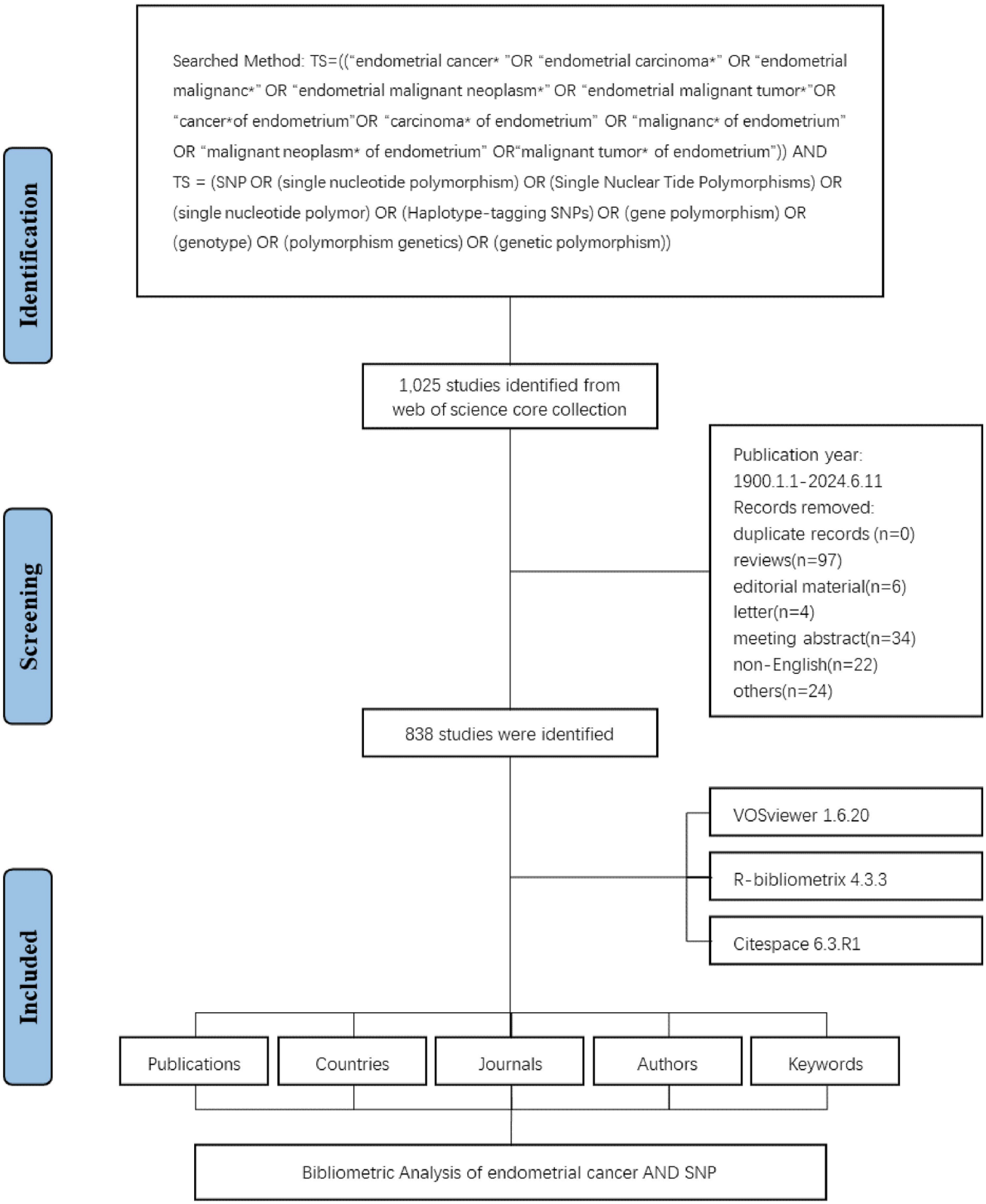


Fig. 1 Flowchart of the Literature Screening Process for Studies on SNPs in Endometrial Cancer

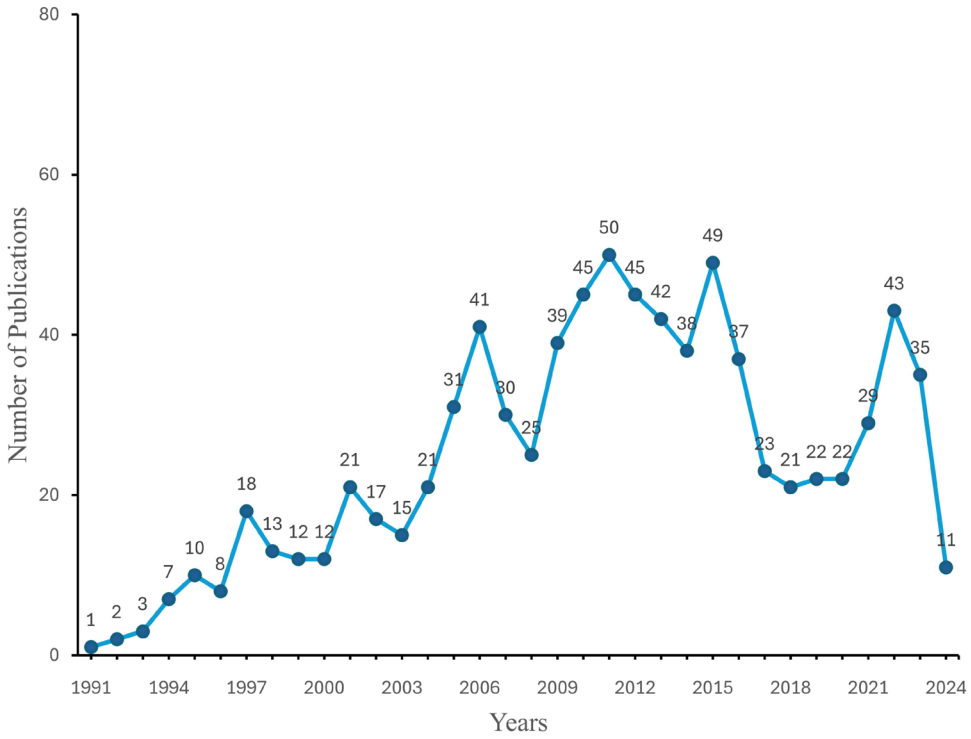


Fig. 2 Trends in Annual Publications from 1991 to 2024 on SNPs in Endometrial Cancer

3.2 Analysis of journals

A total of 292 journals published the 838 studies. In examining the impact of journals on SNP research in EC (Table 1), it is evident that titles such as *Gynecologic Oncology,* with an IF of 4.5, led in total publications (TP = 47) and maintained an H-index of 21, though it ranked fifth in total citations (TC = 738). *Cancer Epidemiology Biomarkers & Prevention,* with an IF of 3.7, came second in both total publications (TP = 32) and citations (TC = 984), and had the highest H-index (22) among the top journals, underscoring its significant influence. *Carcinogenesis*, ranked sixth in total publications (TP = 17) and seventh in total citations (TC = 576), with a solid H-index of 15, indicating a substantial impact despite its moderate publication volume.

The co-occurrence and coupling networks of journals revealed the interconnectedness and intellectual foundations of key publications in the field. In the co-occurrence network, 86 journals with at least 3 co-occurrences were analyzed **(**Fig. 3A**)**, *Gynecologic Oncology* and *Cancer Epidemiology Biomarkers & Prevention* emerged as central nodes, frequently co-cited alongside other influential journals like *Cancer Research* and *International Journal of Cancer*. Meanwhile, in the coupling network among 86 journals with at least 3 couples (Fig. 3B), journals are linked based on the extent to which they cite common references. *Gynecologic Oncology* remains a pivotal journal, frequently coupled with *International Journal of Cancer, Carcinogenesis*, and *Cancer Epidemiology Biomarkers & Prevention*. Journals such as *Cancer Research* and *Clinical Cancer Research* also exhibit strong coupling connections, suggesting shared intellectual foundations.

3.3 Analysis of the countries

The analysis of publications on SNPs in EC by country (Fig. 4A and Table 2), conducted among 346 countries or regions contributed to this filed, revealed significant disparities in research output and influence across various nations. While the United States and China lead in terms of publication volume, countries like Australia and the United Kingdom (UK) demonstrate high-impact research with substantial international collaborations.

The United States leads the field with 230 articles, ranking first 1,348 articles of TP (27.4%) and 9,215 TC, averaging 40.1 citations per article. The multiple-country publication (MCP) ratio for the United States is 0.378, indicating a moderate level of international collaboration. China follows with 182 articles (21.7%), ranking second in both TP

Table 1 Bibliometric indicators of Top 20 high-impact journals

Journal	H_index	IF	JCR_Quartile	PY_start	TP	TP_rank	TC	TC_rank
Cancer epidemiology biomarkers & prevention	22	3.7	Q1	2003	32	2	984	2
Gynecologic oncology	21	4.5	Q1	1995	47	1	738	5
Carcinogenesis	15	3.3	Q2	1997	17	6	576	7
International journal of cancer	15	5.7	Q1	1993	19	5	851	3
International journal of gynecological cancer	14	4.1	Q2	1998	22	4	185	32
Cancer research	13	12.5	Q1	1991	15	7	2157	1
Cancer	12	6.1	Q1	1995	14	8	262	25
Clinical cancer research	12	10	Q1	1998	12	9	3	936
Plos one	12	2.9	Q1	2009	24	3	288	20
Tumor biology	8	N/A	N/A	2011	12	10	121	54
BMC cancer	7	3.4	Q2	2008	10	13	136	45
British journal of cancer	7	6.4	Q1	1995	8	14	451	10
Cancer genetics and cytogenetics	7	N/A	N/A	1997	8	15	106	61
European journal of gynaecological oncology	7	0.5	Q4	1997	11	11	55	96
Familial cancer	7	1.8	Q3	2005	8	16	103	62
Human pathology	7	2.7	Q2	1996	7	21	135	47
Asian pacific journal of cancer prevention	6	N/A	N/A	2008	10	12	68	88
Cancer causes & control	6	2.2	Q2	2004	7	19	158	37
Cancer letters	6	9.1	Q1	2000	6	26	246	27
European journal of cancer	6	7.6	Q1	1994	6	27	143	40

H_index Measures both the productivity and citation impact of the publications in the journal. *IF (Impact Factor)* Indicates the average number of citations to recent articles published in the journal. *JCR_Quartile* Ranks the journal within its field according to the Journal Citation Reports, with quartiles indicating the position (Q1: top 25%, Q2: 25%-50%, Q3: 50%-75%, Q4: bottom 25%). *TP (Total Publications)* Total number of publications by the journal. *TP_rank* Rank of the journal based on the total number of publications. *TC (Total Citations)* Total number of citations received by the journal's articles. *TC_rank* Rank of the journal based on the total number of citations. *PY_start (Publication Year Start)* The year when the journal first started publishing

(636) and TC (2,201), with an average of 12.1 citations per article. However, its MCP ratio is relatively low at 0.071, suggesting fewer international collaborations. Japan contributes 59 articles (7%), ranking fifth in TP (244) but third in TC (1,794), with an average citation rate of 30.4 and an MCP ratio of 0.169. The UK, with 22 articles, ranks third in TP (299) and sixth in TC (1,176), with an average of 53.5 citations per article and an MCP ratio of 0.773.

Among the 40 countries involved in international collaborations with a minimum of three articles (Fig. 4B), the United States led with the highest number of documents (302) and citations (13,995), and it also displays the strongest collaboration network, with a total link strength of 386. The United States maintains close research collaborations with countries such as the UK, Germany, Canada, and China, which are prominent partners in its network. The UK follows with 63 documents, 3,048 citations, and a total link strength of 256. The UK's key collaborative partners include the United States, Germany, and Sweden, forming a strong international research network. Germany, which has contributed 49 documents and received 2,145 citations, has a total link strength of 232, reflecting robust collaborations with countries like the United States, the UK, and France.

3.4 Analysis of the authors

By analyzing the high-impact authors in SNP research in EC (Table 3), Zheng Wei emerges as the most prolific author with an H-index of 19, ranking first in TP (26) and second in TC (1,252). His significant research impact since 2006 is further reflected by his G-index of 26. Amanda B. Spurdle follows closely with an H-index of 18, ranking second in TP (24) and first in TC (1,466), highlighting her influential contributions since 2010. Rodney J. Scott, also with an H-index of 18, ranks fourth in TP (22) and sixth in TC (989), demonstrating his substantial impact in the field. Yong-Bing Xiang and Qiuyin Cai are also high-impact authors, both with an H-index of 17. Their significant contributions are further evidenced by their respective total link strengths of 95 and 94, indicating their substantial collaboration and impact in the field.

Among the 283 authors involved in collaborations with a minimum of three articles (Fig. 5), Zheng Wei leads with 18 documents and 448 citations, and has the strongest collaboration network, evidenced by a total link strength of 106. Zheng Wei maintains close collaborations with prominent researchers such as Xiao-Ou Shu, Hui Cai, and Wang-Hong Xu, forming a strong research cluster indicated by yellow nodes. Spurdle Amanda B (10 documents with total link strength of 77) and Dunning Alison M (7 documents with total link strength of 69) formed another cluster, as indicated by green nodes.

3.5 Analysis of the institutions

Harvard University emerges as the preeminent institution in SNP research on EC with 251 articles, underscoring its dominant role in the field. Both Harvard T.H. Chan School of Public Health and the University of Newcastle have made significant impacts, each publishing 85 articles (Fig. 6A). Among the 104 institutions involved in collaborations with a minimum of three articles (Fig. 6B), the National Cancer Institute (NCI) leads with 27 documents and 1,527 citations, demonstrating a total link strength of 140 that highlights robust collaborative ties. Vanderbilt University, with 26 documents and 737 citations, follows closely, showing substantial collaborative impact with a total link strength of 124. Harvard University, with its considerable document count of 48 and 2,928 citations, also shows a significant total link strength of 116. The University of Hawaii and the University of Cambridge further illustrate strong research output and collaborations, with total link strengths of 105 and 96, respectively, based on 13 and 17 documents.

3.6 Keyword co-occurrence network analysis

In the keyword co-occurrence network analysis, 248 keywords that appeared at least 10 times were identified (Fig. 7), revealed several key research areas in SNP-related studies on EC. The keywords can be classified into distinct thematic clusters based on their clinical or research relevance, as well as their co-occurrence relationships. One prominent cluster focuses on cancer types and their genetic associations. Keywords such as "Breast Cancer", which appears in 188 documents and has received 896 citations, and "endometrial cancer" are frequently co-cited. Other cancer-related terms, such as "prostate cancer," "colorectal cancer," and "lung cancer," also appear frequently, demonstrating the broad scope of SNP research across multiple cancer types. These keywords often co-occur with terms like "mutations" and "microsatellite instability," highlighting the genetic and molecular mechanisms underlying different cancers. Another major cluster revolves around "genetic and molecular mechanisms," with keywords such as "expression" (found in 173 documents, with 800 citations) and "gene" being central to this group. These terms are closely linked with "variants," "carcinoma," and "genome-wide association," emphasizing the research focus on understanding gene expression patterns. "P53" and "BRCA1," well-known cancer-related genes, also appear in this cluster. A third cluster focuses on "risk factors and epidemiology" in cancer research. Keywords like "risk" (159 documents, 784 citations), "obesity," and "susceptibility loci" are frequently co-cited, reflecting ongoing efforts to identify genetic and lifestyle factors that contribute to cancer susceptibility. This cluster also included terms such as "hormones" and "estrogen," linking hormonal pathways to cancer risk in women, especially in breast and endometrial cancers.

3.7 Analysis of burst keywords

The analysis of burst keywords illustrated the evolving research trends and focal points in SNP studies on EC over the past three decades (Fig. 8). The keyword "tumor suppressor gene" experienced the strongest citation burst (14.73) from 1994 to 2002, marking a critical period of research into genetic mechanisms that may inhibit or promote cancer development. Around the same time, "polymerase chain reaction" (1995–1997) and "DNA" (1996–2002) also saw substantial bursts, with strengths of 5.25 and 5.91, respectively. In the late 1990s and early 2000s, attention shifted toward "microsatellite instability", which had a burst strength of 13.3 from 1997 to 2003, reflecting its importance in identifying genetic markers and understanding cancer prognosis. "Nonpolyposis colorectal cancer", closely related to genetic predispositions like Lynch syndrome, had a prolonged burst from 1999 to 2010 (strength 4.29), showing sustained interest in hereditary cancer syndromes. The keyword "germline mutations" (2001–2004, strength 5.04) also highlights research in inherited genetic mutations that increase cancer risk.

In the mid-2000s, keywords such as "association" (2004–2007, strength 4.8) and "carcinogenesis" (2004–2009, strength 4.05) indicated a growing focus on linking genetic variants to cancer development. "Lung cancer" (2005–2013, strength 5.94) and "postmenopausal women" (2006–2010, strength 8) reflected an expansion of SNP research to other cancer types

Fig. 3 Network Analyses of Journals in Endometrial Cancer. **A.** The Co-occurrence Networks of Journals. Journal Link Strength in co-occurrence networks measures the frequency with which two journals are cited together within the same articles or references. **B.** The Coupling Networks of Journals. Journal Link Strength in coupling networks assesses the extent to which journals are linked based on the common references cited in their articles

and specific at-risk populations. The keyword "genome-wide association" (2015–2021, strength 9.6) reflects advancements in identifying genetic variants across populations. Simultaneously, "obesity" (2015–2021, strength 4.2) and "meta-analysis" (2016–2024, strength 4.57) suggest a growing interest in understanding lifestyle factors and synthesizing large datasets to identify genetic risk factors. Emerging methodologies are also indicated by the burst in "Mendelian randomization" (2022–2024, strength 5.52), signaling the increasing application of advanced statistical techniques to infer causal relationships in this disease.

4 Discussion

4.1 Key findings

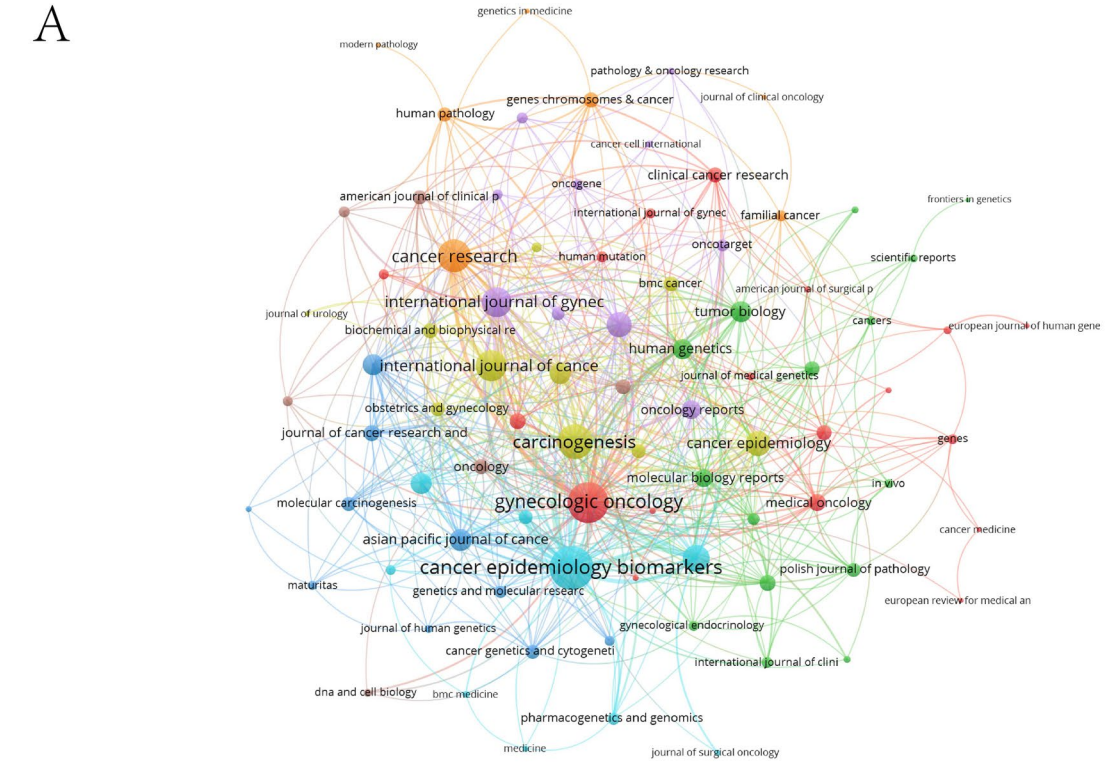
The bibliometric analysis provided a comprehensive overview of the evolution and key research foci in SNP research on EC over the past three decades. Our analysis identified the United States and China as leading contributors to the field, with notable institutions such as Harvard University and the NCI making significant impacts. Key research themes emerged, initially centering on 'endometrial carcinoma' and 'microsatellite instability', and more recently shifting towards 'mendelian randomization' and 'genome-wide association'. The analysis underscores the crucial role of SNPs in understanding EC's genetic basis and highlights the potential for targeted therapeutic strategies based on these genetic variations. While these findings provide a valuable summary of the research landscape, they also underscore critical gaps in the field, such as limited integration of SNP data into clinical practice and a need for more diverse population studies. These gaps represent significant opportunities for advancing the field by bridging the translational divide between molecular insights and their application to personalized medicine. For example, while GWAS have identified several susceptibility loci, such as HNF1B and CYP19A1, their clinical utility remains underexplored [19]. Addressing these gaps could lead to breakthroughs in risk prediction, early detection, and tailored treatment approaches.

Among institutions, Harvard University stands out as the most prolific, contributing 251 articles on SNPs in EC. Harvard's research encompasses both fundamental and translational studies, exploring the role of genetic variations in predicting treatment outcomes and advancing personalized medicine approaches [20, 21]. Its groundbreaking work in genetics has significantly contributed to our understanding of the genetic underpinnings of EC. In addition, the NCI plays a pivotal role in advancing cancer genetics research, supporting a broad spectrum of collaborative studies that integrate SNP data with other biomarkers to enhance diagnostic and prognostic models. [22, 23] Through these extensive collaborations, the NCI has helped shape the direction of genetic research in EC, fostering new insights and therapeutic strategies.

It is crucial to acknowledge the importance of global collaboration in advancing scientific knowledge. The analysis shows that countries like Australia and the United Kingdom, despite their relatively lower publication volumes, exhibit strong international collaborations, as evidenced by their high MCP ratios. This underscores the value of cross-border partnerships in enhancing research impact and facilitating the sharing of knowledge and resources.

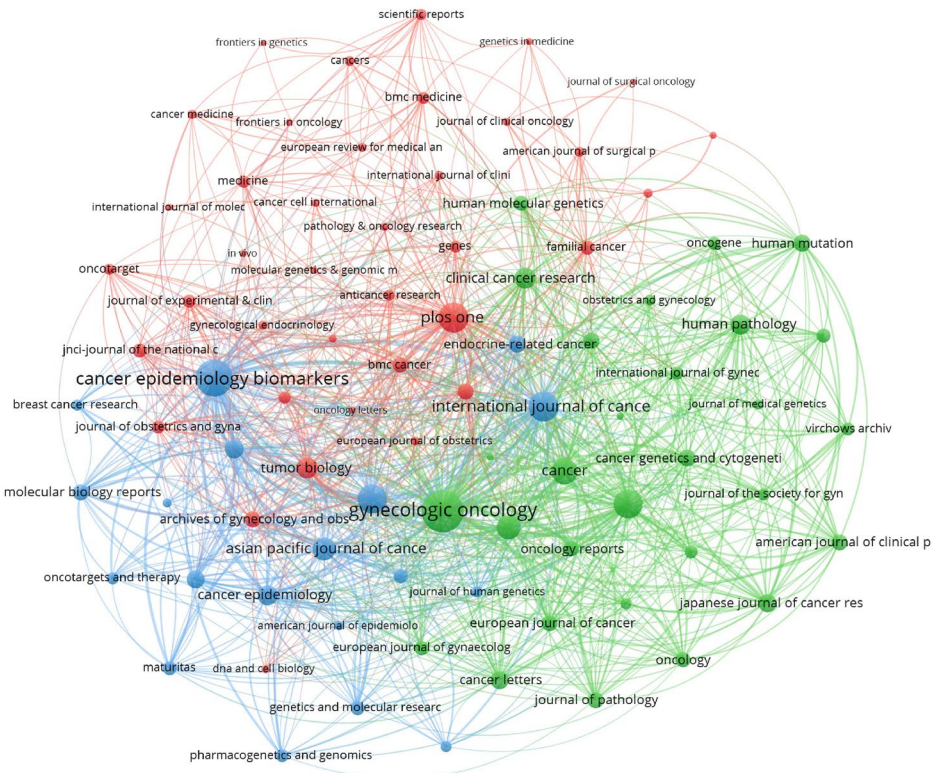
This bibliometric analysis highlights key contributors to SNP research in EC, Zheng Wei [24] and Amanda B. Spurdle [23], for their pivotal roles in advancing the clinical application of genetic findings. Zheng Wei, with an H-index of 19, has significantly contributed to understanding how SNPs impact treatment responses and survival outcomes in EC, reflected by his high citation count of 1,252. Similarly, Amanda B. Spurdle, with an H-index of 18, has utilized SNP data for risk stratification and personalized treatment plans, especially for patients with hereditary risks. Their work underscores the growing importance of integrating genetic data into clinical practice as the field moves toward precision medicine. Along with other researchers like Rodney J. Scott and Yong-Bing Xiang, their efforts reflect a shift towards personalized medicine in EC, where genetic heterogeneity plays a key role in disease progression. Additionally, their extensive collaborative networks emphasize the importance of international cooperation, as evidenced by the 15.38% co-authorship rate in SNP research on EC, facilitating the validation of genetic associations and their translation into clinical practice. Together, these authors exemplify the interconnectedness of global research efforts, driving forward innovations in EC management through both individual and collaborative contributions.

A



VOSviewer

B



VOSviewer

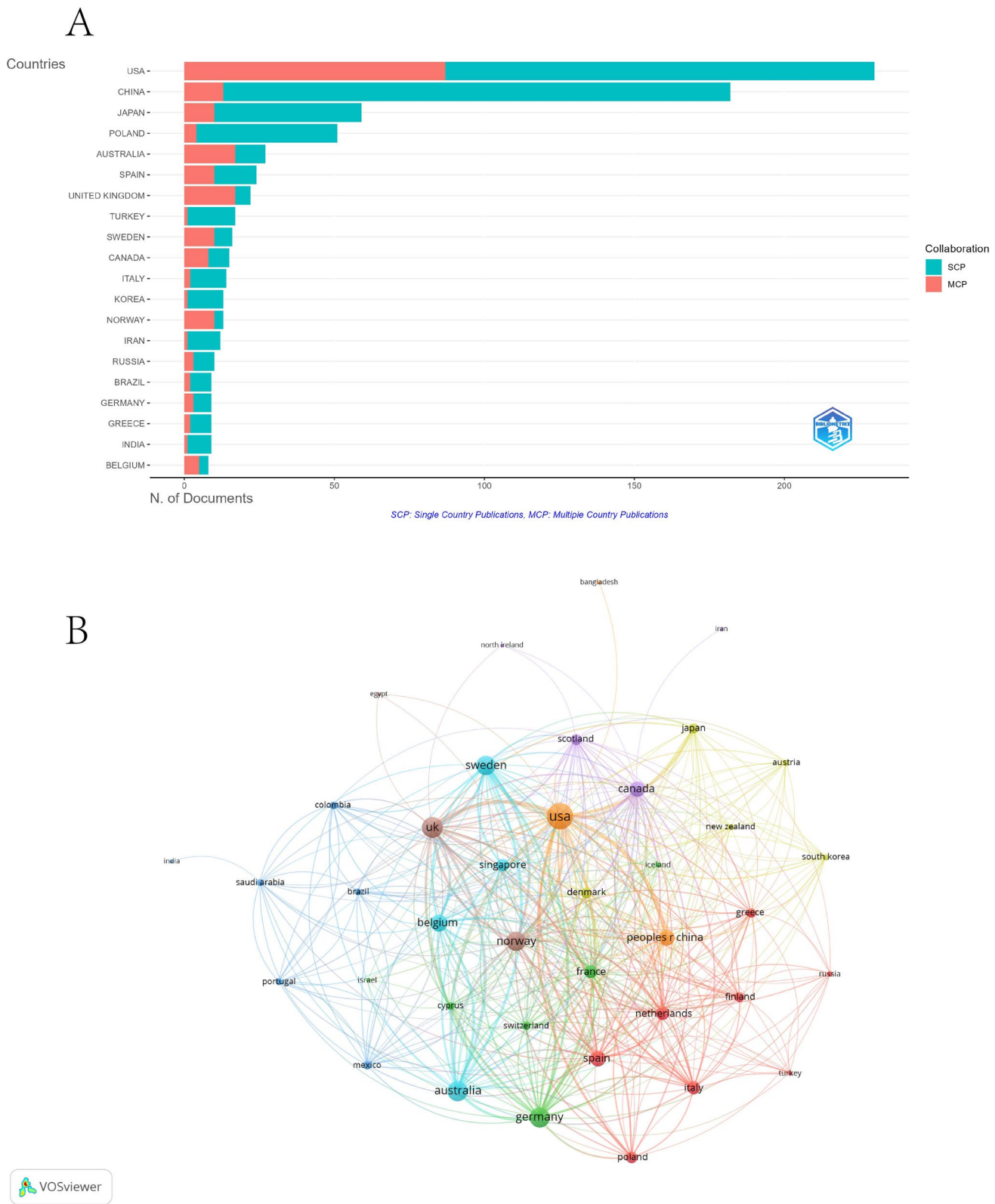


Fig. 4 Global Research Distribution and Collaboration in SNP Studies on Endometrial Cancer. **A**. Map displaying the geographic distribution of publications by corresponding authors' countries, categorized into Single Country Publications (SCP) and Multi-Country Publications (MCP). **B**. Diagram illustrating publication volumes and cooperative networks among countries. Color depth indicates the volume of publications, with the United States ranked first. Lines represent collaborative connections between authors from different countries

Table 2 Publication and Citation Profiles of Top 20 Leading Countries

Country	Articles	Freq	MCP_Ratio	TP	TP_rank	TC	TC_rank	Average Citations
USA	230	0.274	0.378	1348	1	9215	1	40.1
China	182	0.217	0.071	636	2	2201	2	12.1
Japan	59	0.070	0.169	244	5	1794	3	30.4
Poland	51	0.061	0.078	230	6	617	9	12.1
Australia	27	0.032	0.630	292	4	1165	7	43.1
Spain	24	0.029	0.417	116	10	1683	4	70.1
United Kingdom	22	0.026	0.773	299	3	1176	6	53.5
Turkey	17	0.020	0.059	60	16	228	15	13.4
Sweden	16	0.019	0.625	122	9	377	12	23.6
Canada	15	0.018	0.533	110	11	755	8	50.3
Italy	14	0.017	0.143	91	14	215	17	15.4
Korea	13	0.016	0.077	36	21	270	14	20.8
Norway	13	0.016	0.769	152	8	432	11	33.2
Iran	12	0.014	0.083	40	20	136	21	11.3
Russia	10	0.012	0.300	29	23	203	20	20.3
Brazil	9	0.011	0.222	46	18	90	25	10
Germany	9	0.011	0.333	228	7	96	24	10.7
Greece	9	0.011	0.222	32	22	211	18	23.4
India	9	0.011	0.111	26	25	83	26	9.2
Belgium	8	0.010	0.625	88	15	353	13	44.1

Articles: Publications where the corresponding authors are from a specific country. *Freq (Frequency of Total Publications)* Frequency of publications from the country relative to the total dataset. *MCP_Ratio (Multiple Country Publications Ratio)* Proportion of publications involving collaborations with authors from other countries. *TP (Total Publications)* Total number of publications from the country. *TP_rank* Rank of the country based on the total number of publications. *TC (Total Citations)* Total number of citations received by the country's publications. *TC_rank* Rank of the country based on the total number of citations. *Average Citations* The average number of citations per publication from the country

These findings highlight the potential for leveraging international networks to address emerging challenges in the field. For instance, collaborative efforts could prioritize underrepresented populations in SNP research, improving the generalizability of findings and ensuring that advancements benefit diverse patient groups. Additionally, expanding global partnerships could accelerate efforts to integrate SNP data with environmental and lifestyle factors, a critical step toward developing robust, clinically actionable risk prediction models.

4.2 Research hotspots

The keyword co-occurrence network revealed several distinct research hotspots in SNP-related studies on EC, reflecting the evolving focus of the field. One major cluster center around cancer types and genetic associations, with keywords like "endometrial carcinoma" and "microsatellite instability" frequently appearing. These terms underscore the foundational understanding of EC genetics, particularly the role of genomic instability in cancer development. As previous studies have demonstrated, microsatellite instability is a key marker for DNA mismatch repair deficiencies, which can drive tumorigenesis in EC, especially in cases associated with Lynch syndrome [25]. Similarly, the frequent appearance of "endometrial carcinoma" reflects ongoing efforts to map specific genetic variations associated with EC progression and prognosis [26], as SNPs have been linked to various aspects of disease susceptibility and progression.

Another prominent research area involves genetic mechanisms and molecular pathways, represented by keywords like "genome-wide association" (GWAS), "expression," and "gene". The increasing focus on GWAS highlights the field's shift towards more comprehensive approaches to identifying genetic risk factors. As outlined in recent studies, GWAS has become instrumental in uncovering susceptibility loci and understanding how SNPs contribute to disease risk across populations [27, 28]. For example, GWAS has identified several risk loci for EC, such as variants in the *HNF1B* gene, which have been linked to EC risk in multiple populations [29]. Keywords such as "expression" and "gene" emphasize

Table 3 Publication and Citation Profiles of Top 20 High-Impact Authors

Authors	H_index	g-index	m-index	PY_start	TP	TP_Frac	TP_rank	TC	TC_rank
Zheng Wei	19	26	1.00	2006	26	2.19	1	1252	2
Scott Rodney j	18	22	1.06	2008	22	1.40	4	989	6
Spurdle Amanda b	18	24	1.20	2010	24	0.94	2	1466	1
Salvesen Helga b	17	18	1.00	2008	18	0.86	12	808	14
Xiang Yong-Bing	17	23	0.94	2007	23	2.01	3	827	12
Lambrechts Diether	16	20	1.07	2010	20	0.67	6	935	9
Amant Frederic	15	18	1.00	2010	18	0.62	10	748	20
Cai Qiuyin	15	18	0.79	2006	18	1.95	11	442	46
De vivo Immaculata	15	21	0.79	2006	21	2.64	5	764	18
Dunning Alison m	15	19	1.00	2010	19	0.59	9	970	7
Shu Xiao-Ou	15	20	0.79	2006	20	1.32	8	1012	5
Hall Per	14	16	0.78	2007	16	0.51	16	832	11
O'mara Tracy a	14	20	1.00	2011	20	0.80	7	1066	4
Otton Geoffrey	14	16	0.82	2008	16	0.92	17	755	19
Trovik Jone	14	14	1.00	2011	14	0.38	23	785	15
Attia John	13	15	0.77	2008	15	0.80	18	720	23
Beckmann Matthias w	13	13	0.77	2008	13	0.28	24	783	16
Czene Kamila	13	14	0.77	2008	14	0.37	21	810	13
Easton Douglas f	13	17	0.93	2011	17	0.59	14	913	10
Mcevoy Mark	13	15	0.77	2008	15	0.80	19	720	23

H_index Measures both the productivity and citation impact of the author's publications. *g_index* Gives more weight to highly-cited articles, emphasizing the impact of top-performing papers. *m_index* The h-index divided by the number of years since the author's first published paper, measuring consistent impact over time. *TP (Total Publications)* Total number of publications by the author. *TP_rank* Rank of the author based on the total number of publications. *TP_FRAC (Fraction of Total Publications)* Proportion of the author's publications relative to the total number of publications in the dataset. *TC (Total Citations)* Total number of citations received by the author's publications. *TC_rank* Rank of the author based on the total number of citations. *Average Citations* The average number of citations per publication by the author. *PY_start (Publication Year Start)* The year when the author first started publishing

the exploration of gene expression patterns and molecular mechanisms that influence EC pathogenesis, pointing to the growing importance of integrating SNP data with gene regulatory networks to understand disease heterogeneity.

The third research hotspot revolves around risk factors and epidemiology, with keywords like "obesity," "hormones," and "risk" frequently co-occurring. These terms reflect the increasing recognition that both genetic and environmental factors, such as lifestyle and hormonal influences, play a critical role in EC risk. For instance, obesity has been widely studied as a risk factor for EC, particularly in relation to SNPs that affect metabolic and hormonal pathways [30, 31]. Studies have shown that obesity-related SNPs, such as those near the *FTO* gene, are associated with increased EC risk, likely due to their role in regulating adiposity and insulin resistance [32]. The co-occurrence of "hormones" further underscores the well-established link between hormonal regulation—especially estrogen exposure—and EC risk, with genetic variants potentially modulating these effects [33]. Hormone-related SNPs in genes like *CYP19A1*, which influences estrogen biosynthesis, have been implicated in EC susceptibility, particularly among postmenopausal women [34].

In addition to these established research areas, the analysis of keyword bursts highlights emerging trends, particularly the recent prominence of "mendelian randomization" (MR) and "genome-wide association." The emerging focus on MR as a research hotspot is particularly noteworthy. MR studies have provided key insights into the causal role of hormonal, metabolic, and lifestyle factors in EC risk. For instance, a study by Nead et al. (2015) employed MR analysis to demonstrate a causal association between insulin levels and EC risk, highlighting the role of metabolic factors in EC development [6]. Similarly, a recent study by Katagiri et al. (2023) used MR to identify protective effects of increased physical activity against EC risk, suggesting lifestyle interventions as a potential strategy for prevention [35]. These studies underscore the utility of MR in disentangling the causal relationships between genetic predispositions and modifiable risk factors, offering actionable insights for prevention and treatment. This shift towards MR signifies a broader methodological evolution in the field, enabling researchers to move beyond simple correlation studies and explore more complex interactions between genetic variants and disease. The application of MR in EC research holds significant promise, as it allows for the

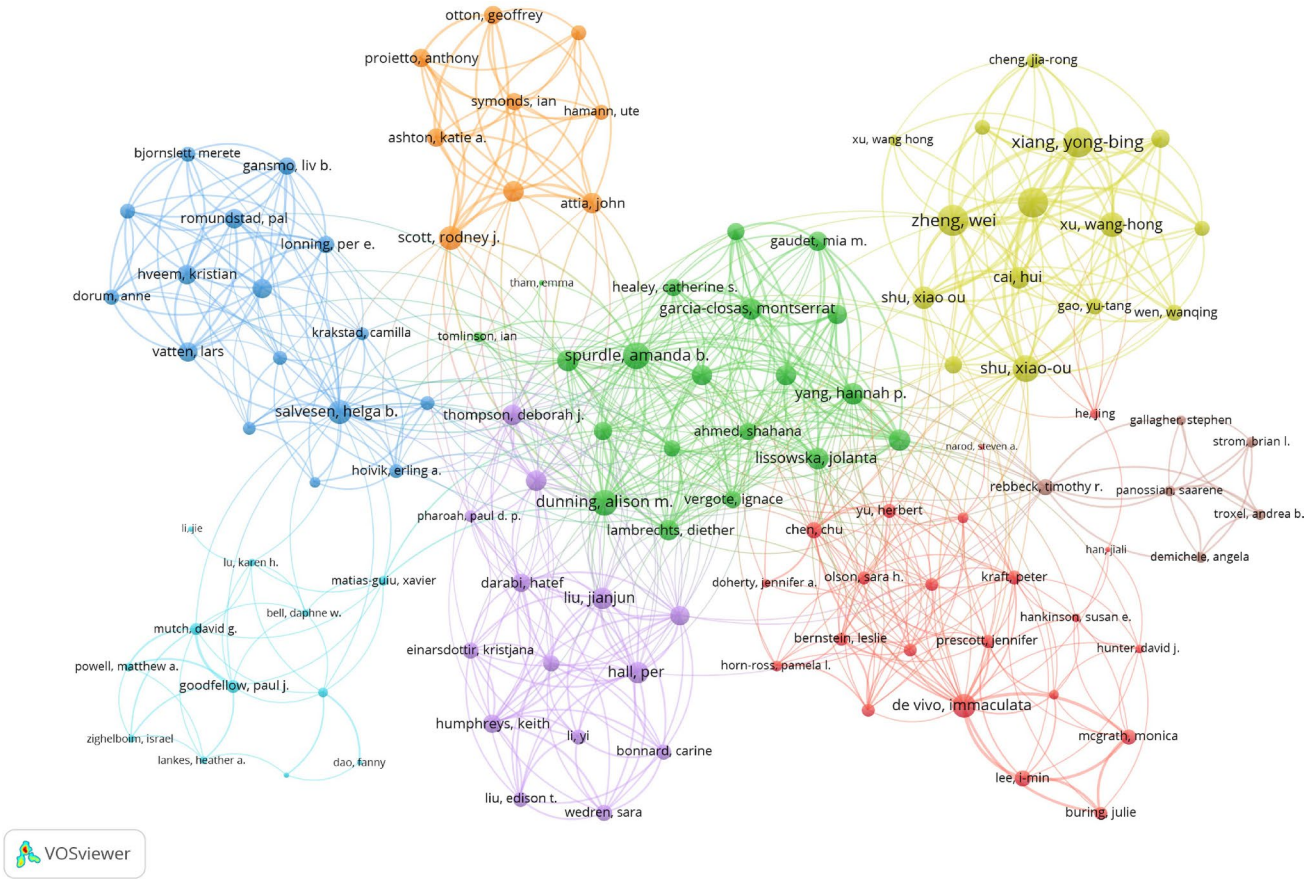


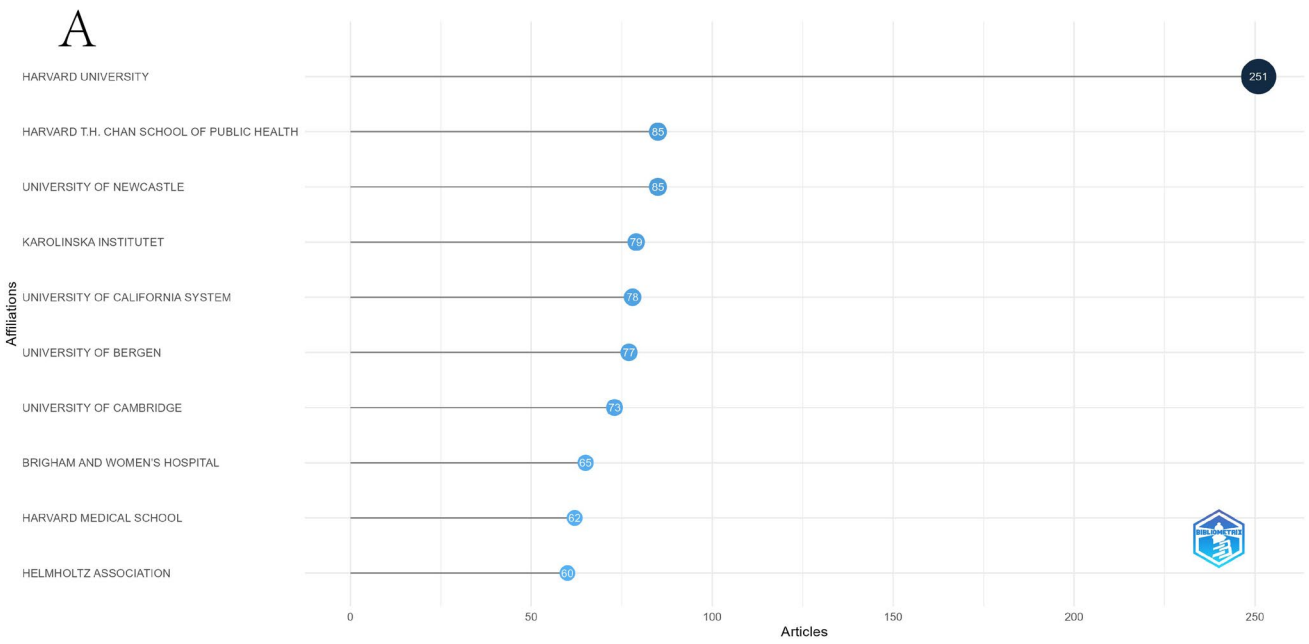
Fig. 5 Author Collaboration Network. Node size reflects the volume of publications, while link thickness indicates the strength of co-authorships. Different colors represent distinct research clusters, with total link strength showing the extent of collaborative interactions

dissection of causal pathways with greater precision, potentially informing more targeted prevention and treatment strategies [36].

4.3 Implications for further studies

Our bibliometric analysis, particularly the keyword burst results, reveals several emerging trends that will likely shape the future of SNP research in EC. One of the most recent and significant keywords to emerge is "Mendelian randomization", reflecting growing interest in this methodology as a powerful tool for identifying causal relationships between genetic variations and EC risk or prognosis. Mendelian randomization offers a novel approach to understanding the role of SNPs in EC by using genetic variants as instrumental variables to infer causality between risk factors—such as hormonal influences or metabolic conditions—and disease outcomes [37, 38]. The Mendelian randomization studies may also explain the recovery of numbers of publications after 2020. Another keyword that has recently gained prominence is "genome-wide association," underscoring the continued relevance of GWAS in uncovering novel genetic loci associated with EC [27]. GWAS has already identified several important SNPs linked to EC risk, such as variants in the *HNF1B* and *CYP19A1* genes [29], and future studies will likely expand on these findings by investigating underexplored populations or integrating multi-omics approaches. The combination of GWAS with other techniques, such as epigenome-wide association studies (EWAS) and transcriptome analyses, promises to provide a more comprehensive understanding of the molecular mechanisms driving EC, facilitating the development of more precise diagnostic and therapeutic tools.

The burst in the keyword "risk prediction" also points to a growing focus on translating genetic insights into clinical practice. With increasing knowledge of SNPs associated with EC, there is a clear opportunity to develop more robust risk prediction models that integrate genetic, environmental, and lifestyle factors. These models could enable earlier identification of at-risk individuals, guiding personalized screening and prevention strategies. For example, SNPs identified in GWAS could be incorporated into polygenic risk scores (PRS), which have already shown promise in assessing cancer risk in other



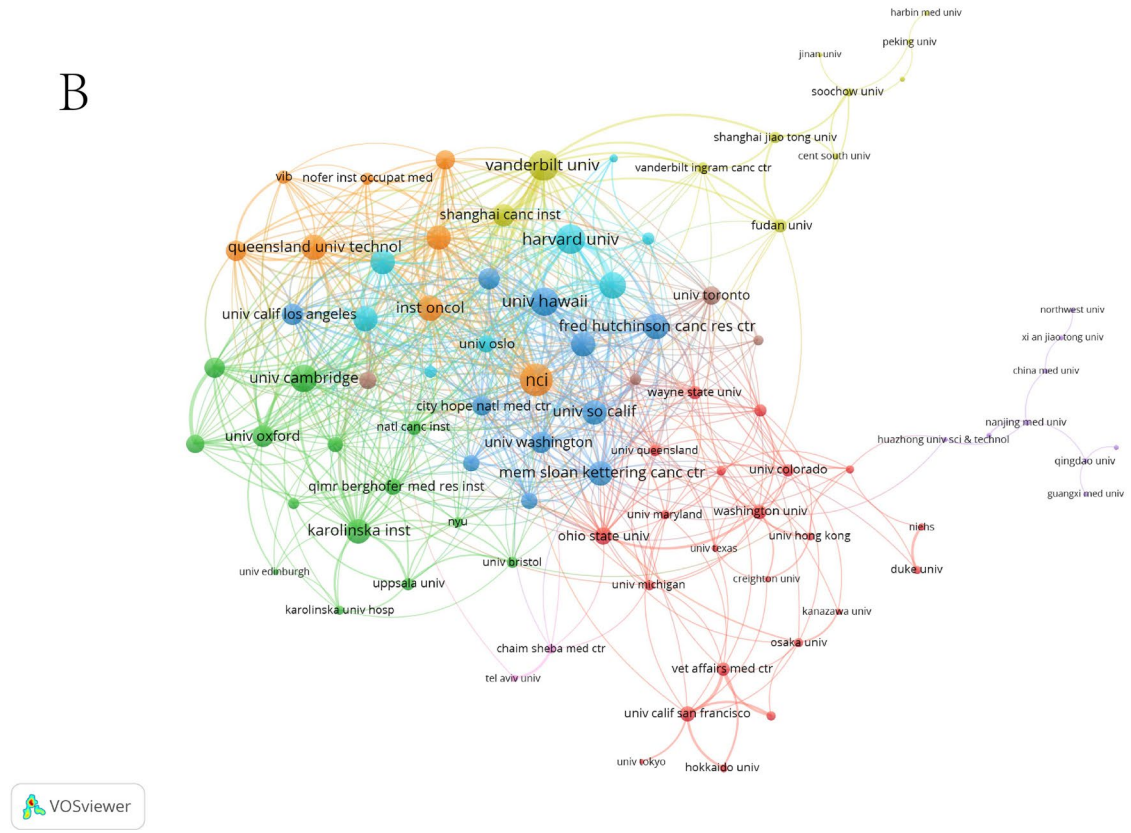


Fig. 6 Institutional Contributions and Collaborations in SNP Research on Endometrial Cancer. **A**. Top 10 institutions by article count and rank. Circle size indicates the number of articles published, with darker shades representing higher ranks in the field. **B**. Visualization map depicting collaboration networks among institutions. Nodes represent institutions, sized by publication count. Links illustrate co-authorships, with thickness indicating the strength of collaboration. Different colors denote various research clusters, highlighting the diversity and scope of institutional interactions

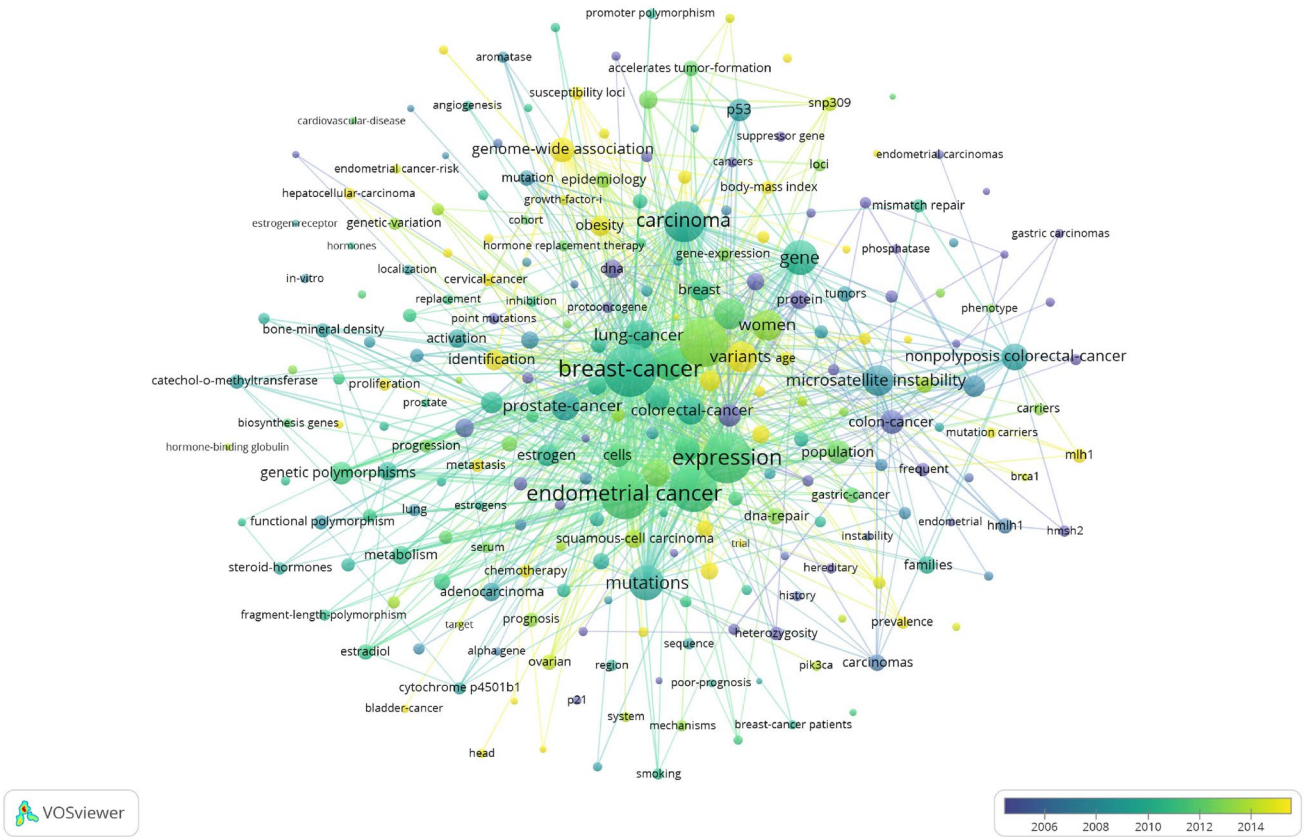


Fig. 7 Visual Analysis of Keyword Co-Occurrence Network in SNP Studies on Endometrial Cancer. Nodes represent keywords, sized according to their frequency of occurrence in the literature. Links between nodes indicate co-occurrences, with line thickness reflecting the strength of the association between keywords. Colors represent the average publication year of articles associated with each keyword, as detailed in the color gradient

populations [39]. As these tools become more refined, they may allow for more tailored risk assessments, particularly in populations with high EC incidence rates.

Furthermore, the recent emergence of "immune response" as a keyword reflects a growing interest in the role of SNPs in modulating immune-related pathways in EC. The connection between genetic variations and the efficacy of immunotherapies, such as immune checkpoint inhibitors, is an area of increasing research focus. Studies have indicated that certain SNPs may influence the immune microenvironment of tumors, potentially impacting the effectiveness of immunotherapy in EC patients. Future research in this area could lead to the identification of SNPs that predict response to immunotherapy, enabling more personalized treatment approaches based on a patient's genetic profile.

5 Limitations

While our study provides valuable insights, it is not without limitations. The reliance on English-language publications may overlook significant research conducted in other languages. Moreover, the literature search was conducted solely using the WoSCC. Although WoSCC is a commonly used database for bibliometric analyses and provides a standardized data source, studies published in other databases may have been overlooked. Additionally, the focus on bibliometric data limits the ability to assess the quality of the research methods used in individual studies.

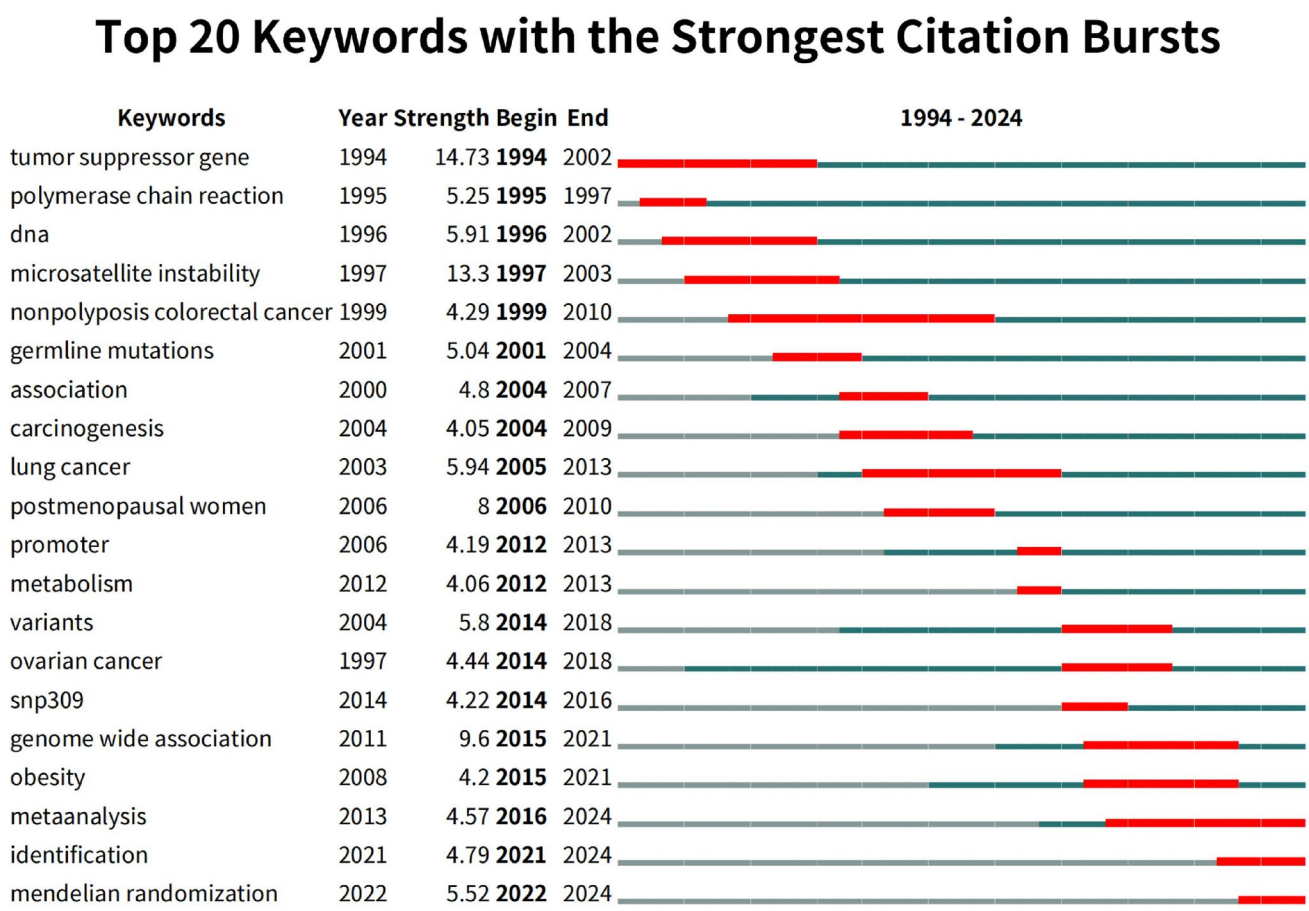


Fig. 8 Top 20 Keywords with the Strongest Citation Bursts in SNP Research on Endometrial Cancer. Chart of the top 20 keywords with significant citation bursts between 1994 and 2024, with blue lines marking the analysis period and red lines highlighting the burst phase

6 Conclusion

In conclusion, the findings from this bibliometric analysis provide a valuable overview of the research trajectory in SNP studies related to EC, highlighting both the achievements and the areas ripe for further exploration. By continuing to focus on the genetic basis of EC, researchers can contribute to the development of more effective, tailored therapeutic strategies, ultimately improving outcomes for patients affected by this disease.

Acknowledgements Not applicable.

Author contributions Renjie Wang and Zhihong Jia carried out the studies, participated in collecting data, and drafted the manuscript. Liang Peng, Qiying Zhu and Jinghui Xu performed the statistical analysis and participated in its design. Renjie Wang and Yinghong Wu helped to draft the manuscript. All authors read and approved the final manuscript.

Funding The study was funded by Planned Project of Jingdezhen Science and Technology Bureau (20224PTJS003).

Availability of data and materials All data generated or analysed during this study are included in this published article [and its supplementary information files].

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Oaknin A, Bosse TJ, Creutzberg CL, Giornelli G, Harter P, Joly F, et al. Endometrial cancer: ESMO clinical practice guideline for diagnosis, treatment and follow-up. Ann Oncol. 2022;33:860–77.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68:7–30.
3. Crosbie EJ, Kitson SJ, McAlpine JN, Mukhopadhyay A, Powell ME, Singh N. Endometrial cancer. Lancet. 2022;399:1412–28.
4. Bogani G, Ray-Coquard I, Concin N, Ngoi NYL, Morice P, Caruso G, et al. Endometrial carcinosarcoma. Int J Gynecol Cancer. 2023;33:147–74.
5. Makker V, MacKay H, Ray-Coquard I, Levine DA, Westin SN, Aoki D, et al. Endometrial cancer. Nat Rev Dis Primers. 2021;7:88.
6. Nead KT, Sharp SJ, Thompson DJ, Painter JN, Savage DB, Semple RK, et al. Evidence of a causal association between insulinemia and endometrial cancer: a Mendelian randomization analysis. J Natl Cancer Inst. 2015. https://doi.org/10.1093/jnci/djv178.
7. Pavličev M, McDonough-Goldstein CE, Zupan AM, Muglia L, Hu YC, Kong F, et al. A common allele increases endometrial Wnt4 expression, with antagonistic implications for pregnancy, reproductive cancers, and endometriosis. Nat Commun. 2024;15:1152.
8. Bafligil C, Thompson DJ, Lophatananon A, Smith MJ, Ryan NA, Naqvi A, et al. Association between genetic polymorphisms and endometrial cancer risk: a systematic review. J Med Genet. 2020;57:591–600.
9. Kim S, Misra A. SNP genotyping: technologies and biomedical applications. Annu Rev Biomed Eng. 2007;9:289–320.
10. Mukamurera N, Extension KP. Personalized Medicine: Tailoring Treatments to Individual Genetic Profiles. 2024; 3: 42-5
11 Öztürk O, Kocaman R, Kanbach DK. How to design bibliometric research: an overview and a framework proposal. Rev Manag Sci. 2024. https://doi.org/10.1007/s11846-024-00738-0.
12. Xia X, Zhang Y, Wei Y, Wang MH. Statistical methods for disease risk prediction with genotype data. Methods Mol Biol. 2023;2629:331–47.
13. Yeung AWK. A revisit to the specification of sub-datasets and corresponding coverage timespans when using Web of Science Core Collection. Heliyon. 2023;9: e21527.
14. Zheng W, Wu Y, Wang Y, Cheng J, Shen W. Bibliometric and visualized analysis of the top-100 highly cited articles on immunotherapy for endometrial cancer. Medicine (Baltimore). 2023;102: e34228.
15. Zhang Z, Yang Y, Kong W, Huang S, Tan Y, Huang S, et al. A bibliometric and visual analysis of single nucleotide polymorphism studies in depression. Curr Neuropharmacol. 2024;22:302–22.
16. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84:523–38.
17. Górski A, Zimecki M, Krotkiewski H. Journal impact factor and self-citations. Arch Immunol Ther Exp (Warsz). 2021;69:21.
18. Moussa S. A bibliometric investigation of the journals that were repeatedly suppressed from Clarivate's Journal Citation Reports. Account Res. 2023;30:592–612.
19 Wang X, Glubb DM, O'Mara TA. 10 Years of GWAS discovery in endometrial cancer: Aetiology, function and translation. EBioMedicine. 2022;77: 103895.
20. Shi J, Kraft P, Rosner BA, Benavente Y, Black A, Brinton LA, et al. Risk prediction models for endometrial cancer: development and validation in an international consortium. J Natl Cancer Inst. 2023;115:552–9.
21. O'Mara TA, Glubb DM, Amant F, Annibali D, Ashton K, Attia J, et al. Identification of nine new susceptibility loci for endometrial cancer. Nat Commun. 2018;9:3166.
22. Oaknin A, Gilbert L, Tinker AV, Brown J, Mathews C, Press J, et al. Safety and antitumor activity of dostarlimab in patients with advanced or recurrent DNA mismatch repair deficient/microsatellite instability-high (dMMR/MSI-H) or proficient/stable (MMRp/MSS) endometrial cancer: interim results from GARNET—a phase I, single-arm study. J Immunother Cancer. 2022;10:2.
23. Lindström S, Wang L, Feng H, Majumdar A, Huo S, Macdonald J, et al. Genome-wide analyses characterize shared heritability among cancers and identify novel cancer susceptibility regions. J Natl Cancer Inst. 2023;115:712–32.
24. Katagiri R, Iwasaki M, Abe SK, Islam MR, Rahman MS, Saito E, et al. Reproductive factors and endometrial cancer risk among women. JAMA Netw Open. 2023;6: e2332296.
25. De' Angelis GL, Bottarelli L, Azzoni C, De' Angelis N, Leandro G, Di Mario F, et al. Microsatellite instability in colorectal cancer. Acta Biomed. 2018;89:97–101.
26. Lax SF. Molecular genetic pathways in various types of endometrial carcinoma: from a phenotypical to a molecular-based classification. Virchows Arch. 2004;444:213–23.
27. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12: e1001779.
28. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9:356–69.
29. Spurdle AB, Thompson DJ, Ahmed S, Ferguson K, Healey CS, O'Mara T, et al. Genome-wide association study identifies a common variant associated with risk of endometrial cancer. Nat Genet. 2011;43:451–4.
30 Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med. 2003;348:1625–38.

31. Aune D, Navarro Rosenblatt DA, Chan DS, Vingeliene S, Abar L, Vieira AR, et al. Anthropometric factors and endometrial cancer risk: a systematic review and dose-response meta-analysis of prospective studies. Ann Oncol. 2015;26:1635–48.
32. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316:889–94.
33. Felix AS, Yang HP, Bell DW, Sherman ME. Epidemiology of endometrial carcinoma: etiologic importance of hormonal and metabolic influences. Adv Exp Med Biol. 2017;943:3–46.
34. Brinton LA, Felix AS, McMeekin DS, Creasman WT, Sherman ME, Mutch D, et al. Etiologic heterogeneity in endometrial cancer: evidence from a Gynecologic oncology group trial. Gynecol Oncol. 2013;129:277–84.
35. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey SG. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27:1133–63.
36. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37:658–65.
37 Mondal SK, Sen MK. Loss of phosphatase activity in PTEN (phosphatase and tensin homolog deleted on chromosome ten) results in endometrial carcinoma in humans: an in-silico study. Heliyon. 2020;6:1.
38 McElroy J, Kruse RL, Taylor JF. Candidate genetic polymorphisms and haplotypes associated with endometrial cancer risk. ISEE Conf Abstr. 2016. https://doi.org/10.1289/isee.2016.3530.
39. Dai J, Lv J, Zhu M, Wang Y, Qin N, Ma H, et al. Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. Lancet Respir Med. 2019;7:881–91.