Check for updates

RESEARCH ARTICLE

REVISED Mapping the evidence base of kidney, prostate and urinary tract tumours to inform the WHO classification of tumours

[version 2; peer review: 2 approved with reservations]

Sophie Gruber[1], Karolina Worf[1], Inga Trulson[1], Laura Knoblauch[1], Tobias Gruber[1], Kirstina Schwamborn[2], Fiona Campbell[3], Richard Colling[4], Alex Inskip[5], Elena Plans-Beriso[6], Oana M. Craciun[6], Magdalena Chechlińska[7], Magdalena Kowalewska[7], Kateryna Maslova[7], Puay Hoon Tan[8], Harshima Wijesinghe[9], Ian A. Cree[9], Dilani Lokuhetty[9], Blanca Iciar Indave Ruiz[9], Stefan Holdenrieder[1], WCT EVI MAP group

[1]TUM University Hospital German Heart Center, Institute of Laboratory Medicine, Munich, 80636, Germany
[2]School of Medicine, Technical University of Munich, Institute of Pathology, Munich, 81675, Germany
[3]University of Birmingham, Centre for Evidence and Implementation Science, Birmingham, B15 2RT, UK
[4]University of Plymouth, Plymouth University Peninsula Medical School, Plymouth, England, PL4 8AA, UK
[5]Population Health Sciences Institute, Newcastle University, National Institute for Health and Care Research (NIHR) Innovation Observatory, Newcastle, NE1 7RU, UK
[6]National Center for Epidemiology, Instituto de Salud Carlos III, Department of Epidemiology of Chronic Diseases, Madrid, 28029, Spain
[7]Maria Sklodowska-Curie National Research Institute of Oncology, Department of Experimental Oncology, Warsaw, 02-781, Poland
[8]Luma Medical Centre, Department of Pathology, Singapore, 228510, Singapore
[9]World Health Organization, International Agency for Research on Cancer, Lyon, 69366, France

v2 **First published:** 30 Jun 2026, **6**:224
https://doi.org/10.12688/openreseurope.23926.1
Latest published: 29 Sep 2026, **6**:224
https://doi.org/10.12688/openreseurope.23926.2

Abstract

Background

The project 'World Health Organization (WHO) Classification of Tumours (WCT): a Living Evidence Gap Map by Tumour Type' applies the Evidence and Gap Map (EGM) methodology to systematically synthesise evidence to support iterative refinements of the WCT. As part of this work, we created the first EGM for the WCT of Urinary and Male Genital Tumors, covering clear cell renal cell carcinoma, non-clear cell renal cell carcinoma, glandular neoplasms of the prostate and urothelial tumours.

Open Peer Review

Approval Status ? ?

	1	2
version 2 (revision) 29 Sep 2026		
version 1 30 Jun 2026	? view	? view

1. **David Parada**, Universitat Rovira i Virgili, Reus, Spain

2. **B. Vishal Rao**, Basavatarakam Indo-Americal Cancer Hospital and Research

Methods

Systematic searches in PubMed and Ovid Embase were conducted and harmonised to a 2020 to 2024 publication time period. Following title and abstract screening, reviewers coded and extracted data from the included records for tumour type, characteristic and study design using EPPI-Reviewer®. Levels of evidence were assigned applying the adapted Hierarchy of Research Evidence for Tumour Pathology and visualised with EPPI-Mapper®.

Results

Of 65,686 deduplicated records, 4,138 studies met inclusion criteria, generating 6,117 EGM entries across 30 tumour types. Evidence on histopathology (28.9%) and prognosis (28.2%) dominated the entries, while prediction was underrepresented (1.6%) or absent in several tumour types. High-level evidence accounted for 17.3% of entries overall, while low-level evidence prevailed across all groups (40.9% in urothelial tumours to 68.0% in non-clear cell renal cell carcinomas). Overall, of 180 tumour-type-by-characteristic EGM cells, 24.4% showed absolute gaps, 42.8% relative gaps, 2.8% synthesis gaps and 30.0% solid evidence.

Conclusions

The available evidence for urinary tract and male genital tumours is heterogeneous, with robust research in clear cell renal cell carcinoma contrasting with sparse predictive evidence and absolute gaps for rare and recently introduced tumour types. This map highlights targeted systematic reviews, cancer registry-based case accumulation and standardised use of WCT nomenclature as priorities for strengthening urological tumour classification.

Plain Language summary

Tumours of the urinary tract and male reproductive organs, including kidney, prostate and bladder cancers, are classified according to World Health Organization (WHO) Classification of Tumours. This classification enables consistent diagnosis and treatment of different types of tumours by pathologists and clinicians worldwide. To keep it accurate and up-to-date, researchers must identify where scientific evidence is strong and where gaps remain.

In this study, we systematically searched relevant scientific articles published between 2020 and 2024 to create an evidence and gap map: a visual tool that shows where research exists and how well it addresses various tumour characteristics, such as diagnostic features and patient outcomes. Out of nearly 66,000 studies identified, over 4,000 met the relevance criteria and were included in the final map.

Institute, Hyderabad, India

Any reports and responses or comments on the article can be found at the end of the article.

Given the large volume of publications, the analysis was limited to title and abstracts, rather than the full text.

The map shows that the evidence varies greatly depending on the tumour type. Common types of kidney and bladder cancer have a relatively solid research base, while rarer and newly discovered tumour types are supported by very few studies. A recurring problem, especially with prostate and bladder cancer, is that many studies do not use the defined WHO tumour type nomenclature in their abstracts, making it impossible to capture them in the map. Another gap is the lack of research on how to predict whether drugs developed for a specific tumour type will be effective, which can be partly explained by the focus on accurate diagnosis rather than treatment in this project.

The generated EGM will directly inform future editions of the WHO classification and help prioritize future research. Ultimately, better-targeted research leads to more accurate cancer diagnoses and, consequently, improved treatment decisions for patients worldwide.

Keywords
Evidence-based medicine, Evidence and Gap Map, EGM, glandular neoplasms of the prostate, tumours of the kidney, urothelial tumours, WCT EVI MAP, WHO Classification of Tumours

HE

This article is included in the Horizon Europe gateway.

Corresponding author: Stefan Holdenrieder (s.holdenrieder@tum.de)

Author roles: Gruber S: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Worf K**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Trulson I**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Knoblauch L**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Gruber T**: Software, Writing – Review & Editing; **Schwamborn K**: Formal Analysis, Writing – Review & Editing; **Campbell F**: Conceptualization, Methodology, Resources, Writing – Review & Editing; **Colling R**: Conceptualization, Methodology, Resources, Writing – Review & Editing; **Inskip A**: Data Curation, Resources, Writing – Review & Editing; **Plans-Beriso E**: Conceptualization, Methodology, Resources, Writing – Review & Editing; **Craciun OM**: Conceptualization, Methodology, Resources, Writing – Review & Editing; **Chechlińska M**: Conceptualization, Methodology, Writing – Review & Editing; **Kowalewska M**: Conceptualization, Methodology, Writing – Review & Editing; **Maslova K**: Conceptualization, Methodology, Writing – Review & Editing; **Tan PH**: Conceptualization, Methodology, Writing – Review & Editing; **Wijesinghe H**: Formal Analysis, Writing – Review & Editing; **Cree IA** : Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Resources, Writing – Review & Editing; **Lokuhetty D**: Conceptualization, Formal Analysis, Methodology, Project Administration, Resources, Writing – Review & Editing; **Indave Ruiz BI**: Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Writing – Review & Editing; **Holdenrieder S**: Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: This work was funded by the European Commission within the WCT EVI MAP project [HORIZON grant number 101057127].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Gruber S *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License , which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Gruber S, Worf K, Trulson I *et al.* **Mapping the evidence base of kidney, prostate and urinary tract tumours to inform the WHO classification of tumours [version 2; peer review: 2 approved with reservations]** Open Research Europe 2026, **6** :224 https://doi.org/10.12688/openreseurope.23926.2

First published: 30 Jun 2026, **6**:224 https://doi.org/10.12688/openreseurope.23926.1

REVISED **Amendments from Version 1**

In the updated version, the methodological transparency and interpretation of the Evidence and Gap Map (EGM) was substantially strengthened. The manuscript now more clearly distinguishes between mapped, retrievable evidence and the complete body of potentially available evidence, and explicitly characterises the EGM as a descriptive and exploratory prioritisation tool rather than a definitive assessment of methodological quality or clinical certainty.

We expanded the Methods and Discussion regarding the pragmatic use of title- and abstract-based screening and coding, including its implications for incomplete retrieval and misclassification. The potential influence of terminology heterogeneity is now discussed in greater detail, with qualitative examples from the screening process, and the limitations of interpreting multi-coded EGM entries as independent studies are explicitly addressed.

The explanatory legends of the EGM figures were revised to provide clearer definitions of evidence-gap categories, colour coding and EGM entries. The rationale and consensus-based nature of the gap-classification thresholds are now more clearly described.

Methodological reproducibility was further strengthened through additional information on Priority Screening, quality-control and consensus procedures, ambiguous coding, automated level-of-evidence assignment, software and dependencies. The Data Availability statement was expanded to clarify the publicly available underlying data, code, methodological materials and reproducibility resources.

Finally, the Discussion and Conclusion were revised to distinguish evidence availability, evidence level, methodological quality and clinical certainty, and to highlight terminology bias, abstract-only screening, future full-text review and formal risk-of-bias assessment as important considerations for future EGM iterations.

Any further responses from the reviewers can be found at the end of the article

Introduction

The World Health Organization (WHO) Classification of Tumours (WCT), published by the International Agency for Research on Cancer (IARC) and commonly referred to as the 'WHO Blue Books', serves as a global standard for the classification of cancers. It also provides standardised terminology that facilitates consistent communication among healthcare professionals worldwide, thereby forming the basis for accurate diagnosis, effective treatment planning, leading to improved patient outcomes and robust international research collaboration.[1] Additionally it impacts accurate epidemiological data collection by cancer registries.

Tumours of the urinary system and male genital organs represent a significant global health burden and are among the most frequently diagnosed malignant tumours.[2] Classification of these tumours has evolved considerably, e.g. due to advances in molecular profiling, which have revealed distinct genetic and epigenetic alterations in morphologically similar entities. The 5th edition of the WCT of Urinary and Male Genital Tumours[3] fundamental transformation of diagnostic approaches towards more accurate stratification and increasingly personalised treatment based on molecular tumour characteristics by incorporating multidimensional information generated through advanced technologies, expanding biological knowledge and evolving diagnostic tools and therapeutic strategies.[3,4] Nevertheless, the evidence supporting the diagnostic and prognostic value of specific tumour characteristics varies considerably between different entities, and the systematic assessment of this evidence landscape remains limited.[1]

Evidence and Gap Maps (EGMs) provide a structured visual overview of the research and highlight both robust evidence and critical gaps.[5,6] Unlike traditional systematic reviews, they can capture a broad spectrum of study designs and tumour characteristics, thereby revealing both areas lacking evidence and clusters of low-quality data.[7] The project 'Mapping the Evidence for the WHO Classification of Tumours: a Living Evidence Gap Map by Tumour Type' (WCT EVI MAP), funded by the European Commission (HORIZON grant No. 101057127), applies this methodology to tumour classification. Led by IARC and developed within the framework of the International Collaboration for Cancer Classification and Research (IC3R), the project systematically maps the evidence for all tumour types and presents the results in an interactive decision support online tool, the WCT Evidence Explorer.[7] As part of this framework, the Hierarchy of Research Evidence for Tumour Pathology (HETP) was developed using a Delphi consensus procedure to provide a tailored evidence hierarchy for tumour pathology research.[8] The EGM methodology has already been successfully applied to several tumour types included in the WHO Blue Books.[9–11]

In this study, we present the first EGM for renal cell carcinomas, subdivided into clear cell renal cell carcinoma and non-clear cell renal cell carcinomas, as well as for glandular neoplasms of the prostate and urothelial tumours. These tumour groups illustrate the molecular shift in urological tumour classification.[12,13] By systematically reviewing the available evidence for these tumour groups, we characterise the current research landscape, identify areas requiring further

investigation and highlight areas where, although extensive research exists, methodological quality remains low. Ultimately, this work will help to strengthen the evidence base of the WCT, promote rigorous research standards in tumour pathology, establish reporting standards and improve patient management through more accurate, evidence-based tumour classification in urological pathology.

Methods

To create this EGM, the WCT EVI MAP project consortium developed a common framework based on the PICO criteria (Population, Intervention, Comparison and Outcome), which was adapted for evidence mapping in tumour pathology and incorporated the expertise of pathologists, epidemiologists, molecular biologists and other specialists.[8] The column headings of the map reflect the tumour groups and the corresponding tumour types in line with the WCT of Urinary and Male Genital Tumours (see Supplementary Table S1).[14] A detailed description of the methods of the study can be found in the protocols registered for each tumour group on the Open Science Framework (OSF).[15–18]

Search strategy and screening process

Comprehensive literature searches on the tumour groups of interest were conducted in PubMed and Ovid Embase, by a trained information specialist. The search strategies were developed using a combination of specific terms tailored to each tumour type within each tumour group and terminology describing their characteristics.[19] A more operational than analytical difference between the literature searches for the four tumour groups was that they were conducted at different time points. To ensure methodological consistency and comparability in the analysis of results across all tumour groups, the retrieved records were harmonised to a unified publication period from 1 January 2020 to 31 December 2024. Consequently, records published in 2018 and 2019 were excluded from the clear cell renal cell carcinoma search, while records published in 2025 were excluded from the remaining tumour groups. A minor residual difference remains, as the search for clear cell renal cell carcinoma was conducted on 5 December 2024, whilst the other searches were conducted between 26 February and 12 March 2025. Publications from late 2024 that were indexed in Pubmed or Ovid Embase only after 5 December 2024 are therefore included for the other tumour groups, but not for clear cell renal cell carcinoma. Detailed information on the search strategies for each tumour group, including search queries and restrictions applied, can be found in Supplementary Tables S2A-S5B.[14]

The search results were imported as RIS files into the web-based software EPPI-Reviewer® Version 6.[20] In this project, due to the volume of included records and the scope of the map, screening was restricted exclusively to title and abstract data and the specific reasons for excluding individual studies were not documented. Applying title- and abstract-based screening enabled a standardized and scalable workflow across all participating WCT EVI MAP groups while maintaining consistency in the mapping process. All retrieved records were individually screened using predefined eligibility criteria.

The detailed eligibility criteria and definitions are provided in the study protocols published on OSF.[15–18] In summary, records were excluded if one of the following criteria applied: (i) title and/or abstract were not in English or not available, (ii) studies identified as corrigendum records or studies that have been retracted (iii) study source was non-peer reviewed (including grey-literature, opinion pieces and conference abstracts), (iv) study protocols and surgical descriptions, (v) preclinical mechanistic research, with an exception for studies on pathogenesis or prediction, according to the definitions on WCT tumour characteristics, see Supplementary S6,[14] (vi) all kind of reviews that did not meet predefined criteria for systematic reviews,[15–18] (vii) studies focusing on the evaluation of a treatment intervention, (viii) studies focusing solely on the evaluation of software applications, (ix) studies that did not report on the predefined tumour characteristics or exact WCT 5th edition tumour types of the corresponding tumour group, see Supplementary Table S1,[14] (x) studies that did not provide disaggregated data for the tumour types, (xi) studies reporting on metastasis or recurrence, (xii) studies with data derived from omics databases (e.g. The Cancer Genome Atlas Program or Gene Expression Omnibus) only and (xiii) studies that were unclear according to predefined criteria.[15–18]

According to the screening criteria for clear cell renal cell carcinoma studies addressing metastases, recurrences, therapy relations without biomarkers and omics repository data, were included, which was not the case for the other tumour groups. These broader inclusion criteria for clear cell renal cell carcinoma were applied at an early stage of the project, before the eligibility standards were refined and established.

To improve efficiency, the 'Priority Screening' tool of EPPI-Reviewer® was applied following the manual review of approximately 300 studies to provide an initial training set for the active-learning algorithm underlying the 'Priority Screening' function. This active-learning algorithm is used to iteratively reprioritise unscreened records according to their predicted relevance, while all inclusion and exclusion decisions continued to be made manually by reviewers. More information is provided by Eppi-Reviewer®: https://eppi.ioe.ac.uk/CMS/Portals/35/Machine%20learning%20in%20EPPI-Reviewer%20v_10%20web%20version.pdf

Screening was stopped after 50 consecutive records had been excluded, indicating that no further highly relevant records were being prioritised. To check the 'Priority Screening' function for renal cell carcinoma and glandular neoplasms of the prostate, the EPPI-Reviewer® 'Check Screening' function was applied. This tool identifies records assumed as excluded by the 'Priority Screening' algorithm that are most similar to included studies. The 'Check Screening' function uses the same underlying machine-learning approach as the 'Priority Screening' algorithm, thereby allowing reviewers to identify potential false exclusions. The manual providing more information can be found online: https://eppi.ioe.ac.uk/CMS/Portals/35/Manuals/Check%20Screening.pdf

Excluded records flagged with a probability of over 70% to be included were manually reassessed. For clear cell renal cell carcinoma and urothelial tumours, the predefined stopping criterion of 50 consecutive records was accepted without additional verification using the 'Check Screening' function, since this function was not yet available at this time.

For further quality assurance of the screening process and to keep the number of missed studies as low as possible, a random subset of 10% of all records, or a maximum of 200 studies (whichever was greater), was independently screened by a second reviewer. Interrater agreement was assessed using percentage agreement, with a discrepancy of less than 10% considered acceptable.

Ambiguities and disagreements were resolved through discussions between the reviewers. If consensus could not be reached, a third reviewer participated in the discussion and the final decision was based on majority agreement. The selection process followed the PRISMA 2020 guidelines,[21] with adopted flowcharts documenting the number of records at each stage (see Figure 2, Supplementary Figure S1).[14]

Data extraction and coding

The included records were coded based on the title and abstract according to the following categories: 1. Tumour type(s), see Supplementary Table S1,[14] 2. Tumour characteristic(s), see Supplementary S6,[14] and 3. Study design. As described above, this pragmatic approach of title- and abstract-based coding only, was adopted to ensure that the project remained feasible given the anticipated breadth of the literature across multiple tumour groups and the absence of reliable estimates of the number of records that would be retrieved within this project.

Study designs were classified using the WCT EVI MAP decision tree diagram and glossary,[22,23] which were developed to categorise a wide range of methodological approaches related to tumour pathology. If multiple tumour types or characteristics were reported, all relevant codes were assigned. For records with multiple study designs, the design representing the highest level of evidence (LoE) for the coded tumour characteristics according to the adapted HETP (see Supplementary Table S7), [14,24] refined during the project, was assigned.

Following the same procedure as described above, a second reviewer independently coded 10% of the included studies, or a maximum of 200 records, to ensure consistency of 90% agreement. Records with ambiguous or incomplete information in the title or abstract were independently reviewed and discussed by the reviewers. Where consensus could not be achieved, a third reviewer was consulted and the final decision was reached by majority vote.

The final coded results were then downloaded from EPPI-Reviewer® in JSON format. To ensure consistency across all tumour groups in the final map, additional tumour characteristics of clear cell renal cell carcinoma, which had been coded in the early project phase, were subsequently excluded to align the dataset with the common coding framework. This enables a focused, in-depth analysis of the diagnostic evidence landscape. While these decisions improve methodological consistency within the map, their implications for cross-group comparisons and the scope of the evidence captured are discussed in the section on limitations.

LoE were determined using an updated version of the HETP (see Supplementary Table S7),[14,24] with evidence divided into five levels, where Level One (P1) represents the highest quality of evidence (e.g. systematic reviews) and Level Five (P5) the lowest (e.g. case reports). A custom Python script was developed to automatically assign LoE to the coded studies in the downloaded JSON file linking the coded study design and tumour characteristic(s) to the corresponding LOE in accordance with the adapted HETP framework.[14] Records addressing multiple tumour types and/or multiple tumour characteristics were systematically duplicated within the dataset so that each tumour type-characteristic combination was represented as a separate record. This ensured that the appropriate LOE could be assigned independently to each combination according to its corresponding study design and characteristic. This processing step was necessary to make sure no information about multiple tumour types and/or multiple characteristics would be missed in the visualisation step. The script applied the adapted HETP framework deterministically and did not perform any automated interpretation beyond the manually assigned codes for tumour types, tumour characteristics and study designs. The script was

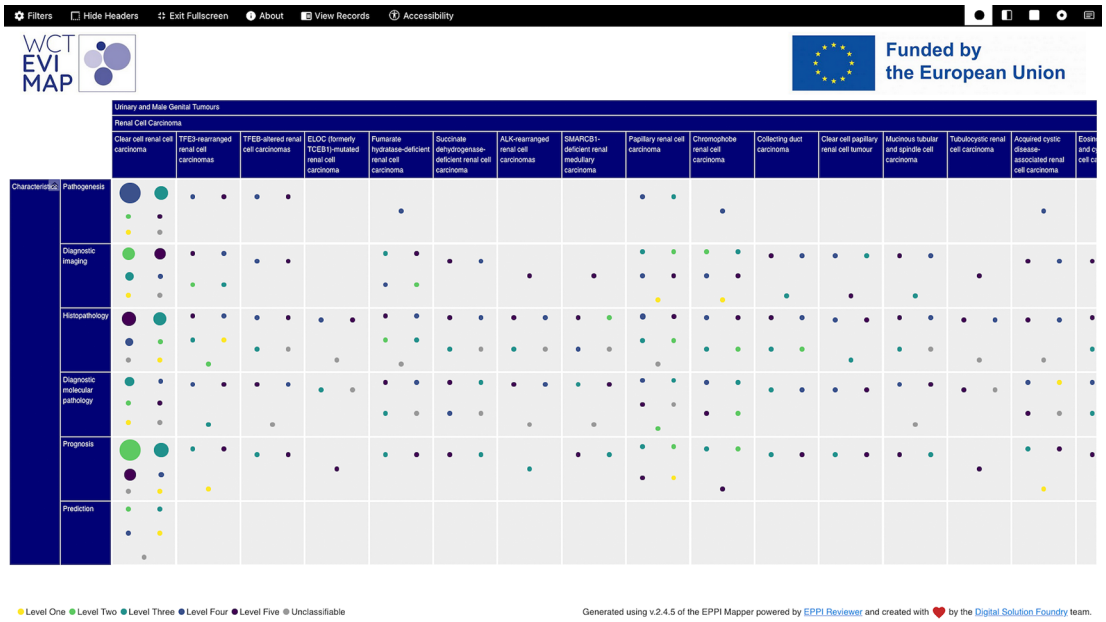


Figure 1. Sample evidence and gap map for WHO Classification of Tumors - Urinary and Male Genital Tumours. Rows depict tumour characteristics and columns represent specific tumour types and groups. Circle size reflects the number of included studies, while colours indicate the assigned Level of Evidence: yellow (Level One), light green (Level Two), turquoise (Level Three), blue (Level Four), dark purple (Level Five) and grey (Unclassifiable). The map was created using EPPI-Mapper®.

implemented in Python version 3.13 and the source is publicly available via the study's OSF repository, which is linked to the corresponding GitHub repository to facilitate access to the source code and version history.[14]

Data analysis and visualisation

The EPPI-Mapper® v.2.4.5 software was used to create an interactive EGM. Each cell in the EGM shows the intersection between a specific tumour type (column) and a tumour characteristic (row), whereas the LoE are represented as colour-coded circles within the cell. The size of the circles corresponds to the number of studies included, and the colours indicate the highest to the lowest LoE available for these studies: from yellow (Level One), light green (Level Two), turquoise (Level Three), blue (Level Four) to dark purple (Level Five) and grey (Unclassifiable). Studies reporting on multiple tumour types and/or tumour characteristics are visualised multiple times on the map. Thus, the number of entries reflects the evidence coverage across the various dimensions of the map and not the total number of independent studies. Absolute gaps (research areas with no evidence), relative gaps (research areas with less than 20 entries and no high-level (Level One and Two) studies), synthesis gaps (available research provides more than 20 entries, but none at Level One and Two) and areas with solid evidence (systematic reviews and high-quality primary studies, Levels One-Two) can be easily identified using this visualisation; see Figure 1 for a sample of the map. This gap classification into the four gap categories was established by consensus among the participating WCT EVI MAP groups prior to analysis. These thresholds were applied uniformly across all tumour groups to ensure methodological consistency within the WCT EVI MAP project. They should therefore be interpreted as pragmatic operational definitions for evidence mapping in this project rather than empirically validated cut-offs.

Figures and tables in this article were created using custom Python scripts, unless otherwise stated. Data analyses were descriptive and exploratory in nature, using counts, proportions, and graphical summaries to characterise the distribution of evidence across tumour groups, tumour characteristics, and levels of evidence. No inferential statistical analyses were performed, as the primary objective was evidence mapping rather than hypothesis testing.

Results

Systematic literature searches in PubMed and Ovid Embase yielded 65,686 records after deduplication (7,748 for clear cell renal cell carcinoma, 10,454 for non-clear cell renal cell carcinoma, 33,143 for glandular neoplasms of the prostate and 14,341 for urothelial tumours). At screening, 9,460 were excluded by human reviewers and 50,598 by the Priority Screening function of EPPI-Reviewer®, and 5,628 studies met the inclusion criteria. Additional 833 studies were excluded during coding, leaving 4,795 coded studies. Of these, 4,138 fell within the harmonised 2020 to 2024 publication

window and were retained for the EGM; the remaining 657 records published in 2018, 2019, 2025 or with no specified year were excluded. Due to the multiple coding of tumour characteristics and types, 43.0% of the studies appear multiple times in the EGM, resulting in 6,117 total entries (Table 1, Figure 2). Per-group PRISMA flowcharts are provided in Supplementary Figure S1.[14]

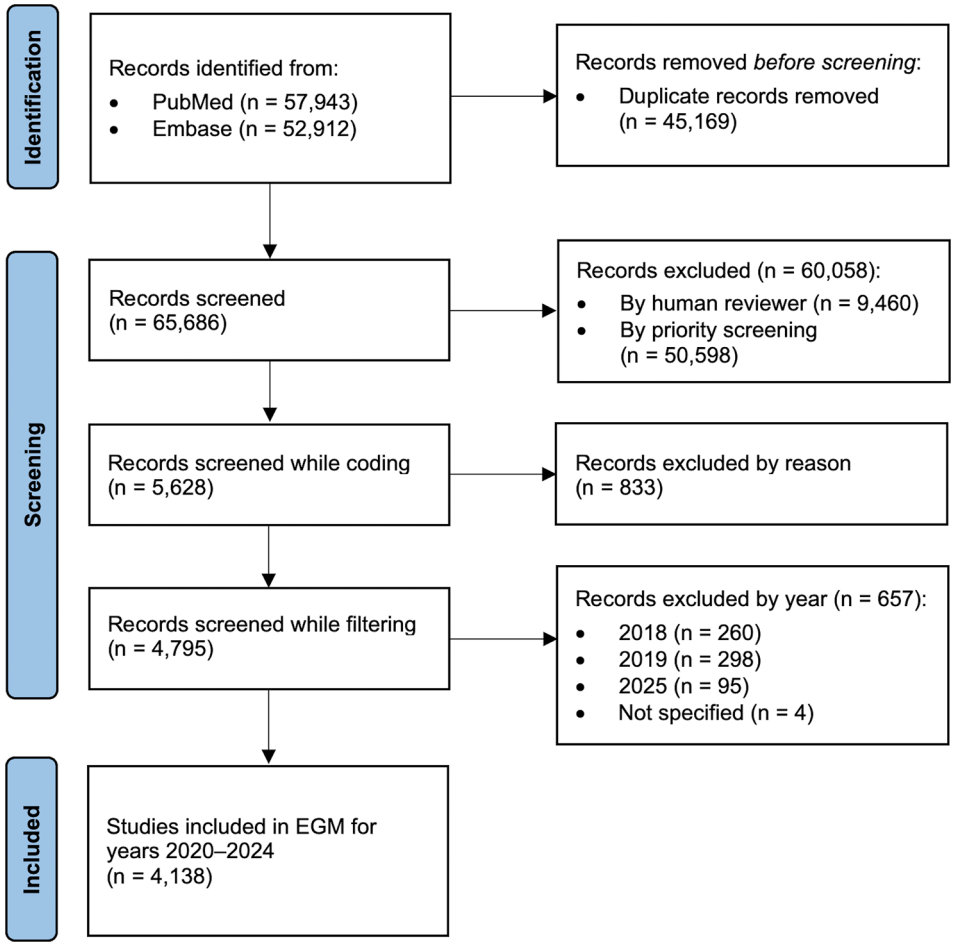


Figure 2. Adapted PRISMA flowchart for the complete evidence and gap map (EGM). Step-by-step presentation of the records identified during the systematic searches, leading up to the studies included in the EGM on tumours of the kidney, prostate and urinary tract, in accordance with the WHO Classification of Tumours of Urinary and Male Genital Tumours.

Table 1. Details of the evidence and gap map (EGM). Number and percentage of studies and entries included in the EGM, by tumour group and in total.

	Clear cell renal cell carcinoma	Non-clear cell renal cell carcinoma	Glandular neoplasms of the prostate	Urothelial tumours	Total (complete EGM)
Studies in EGM (n)	1,900	628	245	1,365	4,138
Systematic reviews (n (%))	18 (0.95%)	6 (0.96%)	4 (1.63%)	46 (3.37%)	74 (1.79%)
Primary studies (n (%))	1,882 (99.05%)	622 (99.04%)	241 (98.37%)	1,319 (96.63%)	4,064 (98.21%)
Multi-coded tumour type studies (n (%))	0 (0.00%)	116 (18.47%)	30 (12.24%)	76 (5.57%)	222 (5.36%)

Table 1. *Continued*

	Clear cell renal cell carcinoma	Non-clear cell renal cell carcinoma	Glandular neoplasms of the prostate	Urothelial tumours	Total (complete EGM)
Multi-coded characteristic studies (n (%))	653 (34.37%)	242 (38.54%)	67 (27.35%)	656 (48.06%)	1,618 (39.10%)
Multi-coded studies in total (n (%))	653 (34.37%)	340 (54.14%)	89 (36.33%)	700 (51.28%)	1,782 (43.06%)
Entries in the EGM (n)	2,719	916	320	2,162	6,117
Entries per tumour type (median (range))	2,719 (2,719 to 2,719)	37 (5 to 337)	79 (5 to 96)	58 (16 to 2,065)	45 (5 to 2,719)

Findings from included studies

The four tumour groups contributed unevenly to the EGM (Table 1). Clear cell renal cell carcinoma had the largest evidence base with 1,900 coded studies (2,719 entries), followed by urothelial tumours (1,365 studies, 2,162 entries), non-clear cell renal cell carcinoma (628 studies, 916 entries) and glandular neoplasms of the prostate (245 studies, 320 entries). Entries per individual tumour type ranged from five (renal cell carcinoma NOS and treatment-related neuroendocrine prostatic carcinoma) to 2,719 (clear cell renal cell carcinoma). Of all included articles, 74 (1.8%) were systematic reviews coded at the highest LoE (Level One) and 4,064 (98.2%) were primary studies distributed across Levels Two to Five. Urothelial tumours had the highest proportion of systematic reviews at 3.4%.

The number of studies per publication year remained relatively stable over the harmonised time period, fluctuating between 781 (18.9%), corresponding to 1,174 entries in the EGM in 2020, and 875 (21.2%), corresponding to 1,270 entries in 2024, with a decline in 2023 that was mainly attributable to studies on urothelial carcinomas and non-clear cell renal cell carcinomas. There were no other notable temporal trends within or between tumour groups (Figure 3).

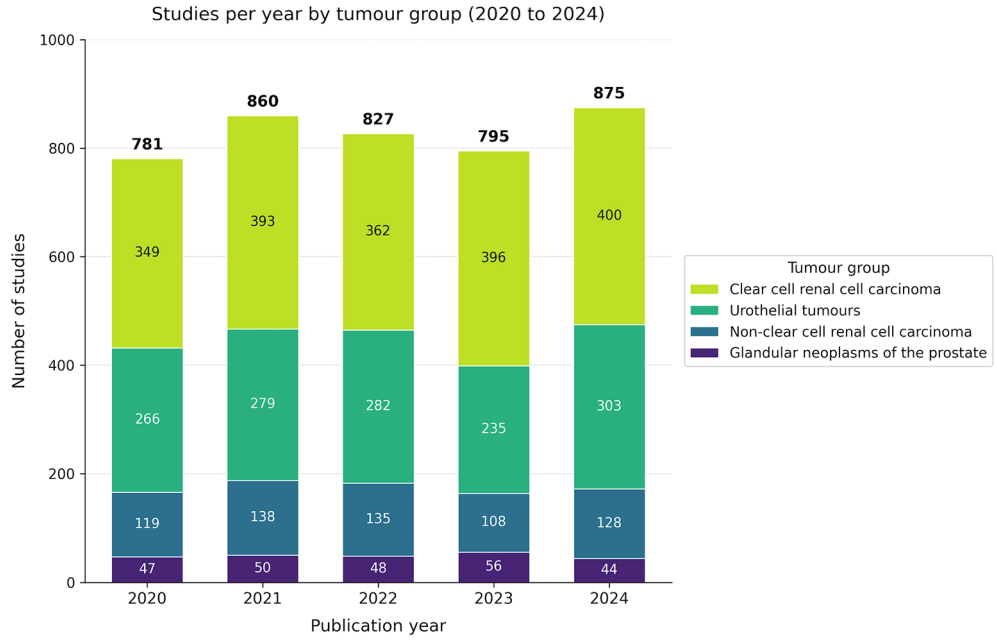


Figure 3. Barplot showing the year of publication of included studies, overall and stratified by tumour group. The height of each stacked bar represents the total number of studies (x-axis) published in a given year (y-axis), with coloured segments indicating the contribution of each tumour group, ordered from the lowest number at the bottom to the highest number at the top: glandular neoplasms of the prostate (purple), non-clear cell renal cell carcinoma (blue), urothelial tumours (turquoise) and clear cell renal cell carcinoma (light green). Numbers within the bars indicate the number of studies per tumour group and bold numbers above the bars indicate the total number of included studies of each publication year.

Evidence distribution across tumour characteristics

The evidence base was concentrated in two characteristics: histopathology (n = 1,765; 28.9%) and prognosis (n = 1,726; 28.2%). Pathogenesis (n = 1,061; 17.4%), diagnostic imaging (n = 771; 12.6%) and diagnostic molecular pathology (n = 694; 11.4%) followed, while prediction was markedly underrepresented (n = 100; 1.6%) and accounted for ≤2% of entries in every tumour group (Figure 4, Supplementary Table S8).[14]

Evidence profiles according to tumour characteristics

The LoE varied considerably both across tumour groups and characteristics (Figure 5, Table 2). Across all four groups, low-level evidence (Levels Four to Five) was the dominant tier, accounting for 40.9% of urothelial tumour entries, 47.0% of clear cell renal cell carcinoma entries, 57.5% of glandular neoplasms of the prostate entries and 68.0% of non-clear cell renal cell carcinoma entries. High-level evidence (Levels One to Two) was correspondingly sparse, ranging from 6.9% in non-clear cell renal cell carcinoma to 23.8% in clear cell renal cell carcinoma (Table 2).

Within tumour groups, the proportion of high-level evidence varied substantially by characteristic (Figure 5, detailed proportion per tumour type provided in Supplementary Figure S2).[14] Prognosis carried the largest absolute number of high-level studies in every group and reached 43.5% high-level evidence in clear cell renal cell carcinoma (337 of 775), but only 8.1% in non-clear cell renal cell carcinoma (15 of 186). Diagnostic imaging showed a similarly uneven pattern, with high-level evidence ranging from 17.1% in glandular prostate neoplasms (7 of 41) to 38.1% in clear cell renal cell carcinoma (164 of 431). Histopathology, despite being the most heavily populated characteristic overall (1,765 entries), held the smallest share of high-level evidence in three of the four groups (1.7% to 10.7%), reflecting the dominance of case reports, case series and other low-level study designs in this domain. Pathogenesis was almost entirely classified as Level Four mechanistic evidence, with high-level proportions of 0.0% to 8.2% across groups. Prediction was the sparsest characteristic across the EGM, with 100 entries in total (Figure 4); within this small pool, clear cell renal cell carcinoma stood out with 65.5% high-level evidence (36 of 55), while glandular prostate neoplasms had no high-level prediction studies (0 of 2) and non-clear cell renal cell carcinoma had no prediction entries.

The complete interactive EGM for renal cell carcinomas, glandular neoplasms of the prostate and urothelial tumours is available online under WCT EVI MAP: Preliminary Results.

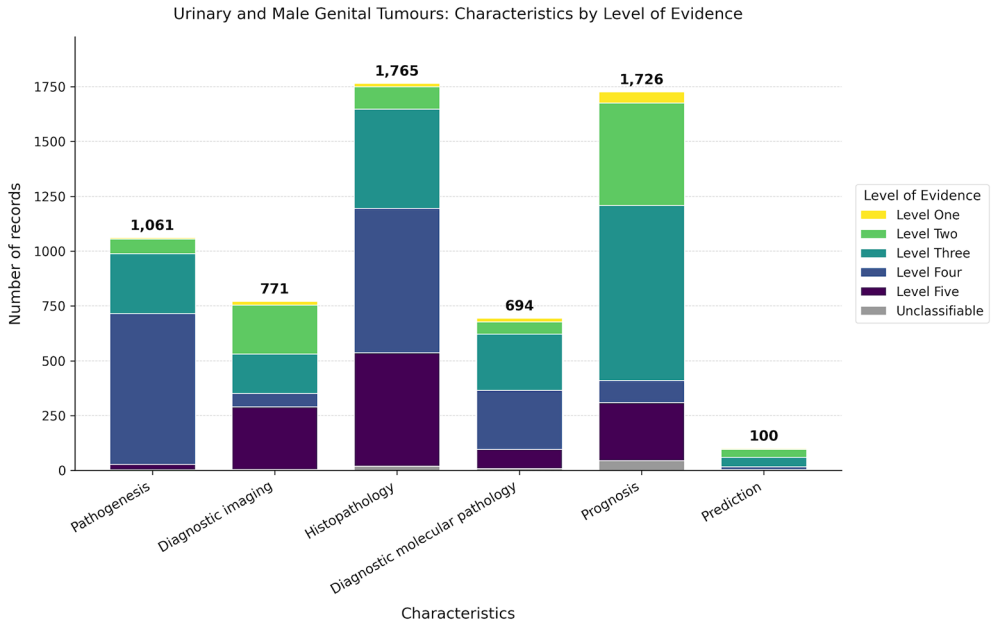


Figure 4. Barplot showing the number of records (entries) identified for each tumour characteristic. The height of each stacked bar represents the total number of records (x-axis) assigned to a given tumour characteristic (y-axis), while colours indicate the level of evidence: Level One (yellow), Level Two (light green), Level Three (turquoise), Level Four (blue), Level Five (purple) and Unclassifiable (grey). Records that contributed to multiple tumour characteristics were counted more than once accordingly. Consequently, the totals shown represent evidence and gap map entries rather than independent studies.

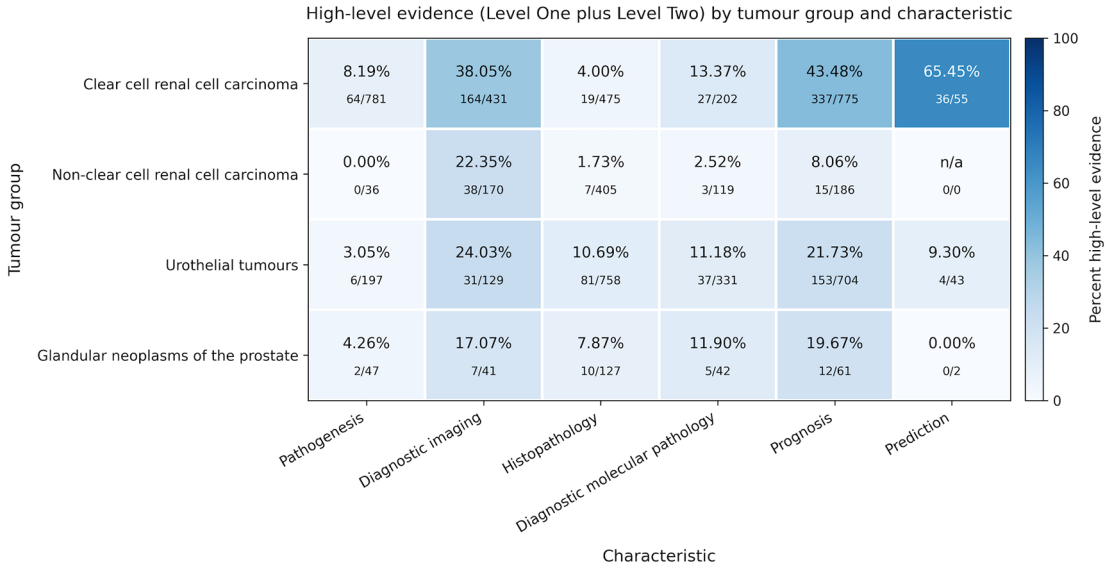


Figure 5. Heatmap showing the proportion of high-level evidence (Level One–Two) by tumour group and characteristic. Each cell displays the percentage of evidence and gap map entries classified as Level One or Level Two, the values represent the numerator (number of records with high-level evidence) in relation to the denominator (total number of records for the respective tumour group and characteristic, irrespective of the level of evidence). The numerator and denominator are shown below the percentage in the cell. Colour intensity reflects the percentage on a fixed scale from 0% to 100%, with a darker shading indicating a higher proportion of Level One-Two evidence. Cells marked 'n/a' indicate combinations with no data records (denominator = 0).

Table 2. Distribution of evidence levels by tumour group. Each row shows the number and percentage within each evidence level category: high-level evidence (Levels One–Two), moderate-level evidence (Level Three), low-level evidence (Levels Four–Five) and Unclassifiable (records that could not be assigned a level of evidence).

Tumour group	Total (n)	High – Levels one to two (n (%))	Moderate – Level three (n (%))	Low – Levels four to five (n (%))	Unclassifiable (n(%))
Clear cell renal cell carcinoma	2,719	647 (23.80%)	752 (27.66%)	1,279 (47.04%)	41 (1.51%)
Non clear cell renal cell carcinoma	916	63 (6.88%)	226 (24.67%)	623 (68.01%)	4 (0.44%)
Glandular neoplasms of the prostate	320	36 (11.25%)	99 (30.94%)	184 (57.50%)	1 (0.31%)
Urothelial tumours	2,162	312 (14.43%)	926 (42.83%)	884 (40.89%)	40 (1.85%)
Total	**6,117**	**1,058 (17.30%)**	**2,003 (32.74%)**	**2,970 (48.55%)**	**86 (1.41%)**

Tumour group-specific evidence patterns and identified gaps

Evidence distribution patterns differed substantially across tumour groups, reflecting variation in disease prevalence, classification maturity and search precision. A complete entity-level breakdown of evidence distribution and gap classification is displayed in the per-group heatmap in Figure 6, with additional detail provided in Supplementary Tables S8-S9B.[14] In Figure 6, each tumour entity is evaluated across predefined characteristics and each corresponding cell is assigned to one of four evidence categories: (i) absolute gaps indicate the complete absence of records; (ii) relative gaps denote fewer than 20 records, with none meeting Level One or Level Two evidence; (iii) synthesis evidence where at least 20 records exist but none is at Level One or Level Two; and (iv) solid evidence indicates the presence of at least one study classified as Level One or Level Two. Of the 180 cells, 44 (24.4%) are absolute gaps, 77 (42.8%) are relative gaps, 5 (2.8%) are synthesis gaps and 54 (30.0%) show solid evidence.

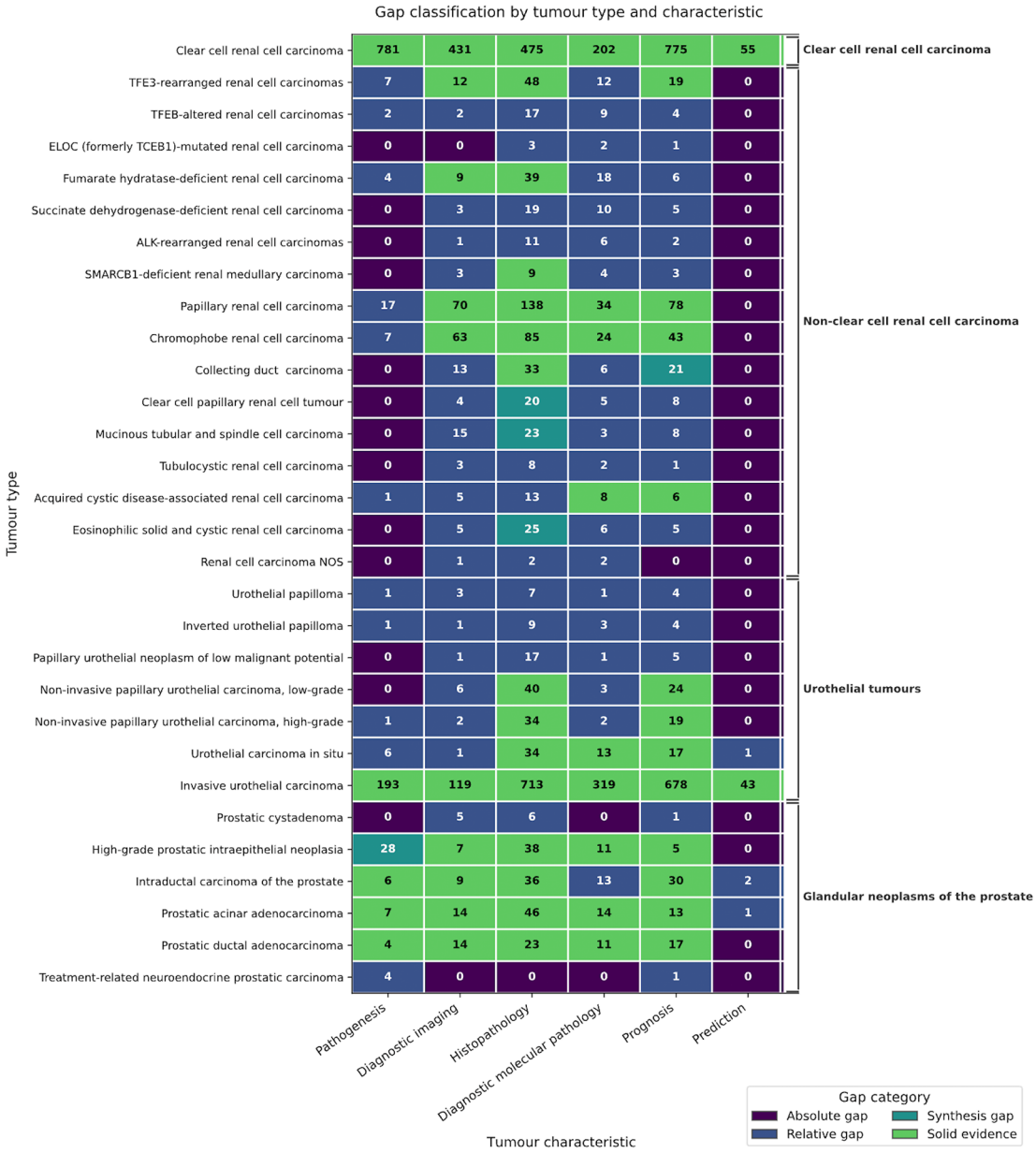


Figure 6. Gap classification matrix by tumour type and characteristic in the evidence and gap map (EGM). Rows represent the 30 individual tumour types (in EGM order, grouped by tumour group), and columns represent the six tumour characteristics. The number in each cell indicates the number of EGM entries reflecting the assigned combination of tumour type and characteristic. Cell colours indicate the evidence-gap classification: absolute gaps (complete absence of records, purple), relative gaps (fewer than 20 records, none meeting Level One or Level Two evidence, blue), synthesis evidence (at least 20 records exist, none at Level One or Level Two, turquoise), and solid evidence (at least one study classified as Level One or Level Two, light green).

Renal cell carcinomas

Clear cell renal cell carcinoma had the most balanced and comprehensive evidence base, with high-level evidence present in every characteristic. Notably, 28.7% of entries addressed pathogenesis, considerably more than the 3.9% to 14.7% observed for other groups and 43.5% of prognostic entries reached at least Level Two. Clear cell renal cell carcinoma was also the only entity with high-level predictive evidence. Among the remaining 16 renal cell carcinoma entities, the combined search yielded substantially less specific results and evidence was sparse. Non-clear cell renal cell carcinoma was dominated by histopathological research (44.2% of entries) but with the highest proportion of low-level evidence for histopathology (87.7%). Papillary and chromophobe renal cell carcinoma carried most of the high-level evidence and

showed solid coverage in diagnostic imaging, histopathology, diagnostic molecular pathology and prognosis. Five additional entities (*TFE3*-rearranged renal cell carcinoma, fumarate hydratase–deficient renal cell carcinoma, acquired cystic disease–associated renal cell carcinoma, collecting duct carcinoma and *SMARCB1*-deficient renal medullary carcinoma) also showed solid evidence in histopathology, with at least three of these entities further supported by solid evidence in diagnostic imaging or prognosis. Solid evidence was identified for four entities, whereas the remaining entities were characterized primarily by relative and absolute gaps, particularly in pathogenesis. *ELOC*-mutated renal cell carcinoma and renal cell carcinoma NOS were the most severely underrepresented entities, with absolute gaps in three characteristics including pathogenesis. Overall, prediction represented an absolute gap for all renal cell carcinoma subtypes except clear cell renal cell carcinoma.

Glandular neoplasms of the prostate

Despite the high global incidence of prostate cancer, the evidence base for individual tumour types was limited, resulting in an inclusion rate of <1% based on abstracts.

Prostatic cystadenoma and treatment-related neuroendocrine prostatic carcinoma were the most severely underrepresented entities, with predominantly absolute gaps in three and four out of six characteristics, respectively. The remaining tumour types demonstrated a relatively balanced evidence base for most characteristics, with the notable exception of prediction. Prediction evidence was confined to two studies on intraductal carcinoma of the prostate and one on prostatic acinar adenocarcinoma, for all other prostate types, there was an absolute gap. A relative gap was identified in the evidence on diagnostic molecular pathology in intraductal carcinoma of the prostate and a synthesis gap was observed in pathogenetic research on high-grade prostatic intraepithelial neoplasia.

Overall, despite the limited evidence base, relatively robust evidence was available for most entities.

Urothelial tumours

The group of urothelial tumours had the highest proportion of systematic reviews (3.4%) and a balanced distribution across non-predictive characteristics, although predictive research remained minimal at 2.0% with absolute gaps in five of the seven entities. Screening exclusion was high (90.1%). Invasive urothelial carcinoma was the only urothelial entity with solid evidence across all tumour characteristics.

Urothelial carcinoma in situ showed solid evidence in histopathology, diagnostic molecular pathology and prognosis, but exhibited relative gaps in the other three tumour characteristics. While both, high-grade and low-grade, non-invasive papillary urothelial carcinoma were supported by solid histopathological and prognostic evidence, the remaining characteristics and the other entities were mostly characterised by relative gaps, with absolute gaps particularly evident in pathogenesis and prediction.

Discussion

This EGM provides the first systematic overview of the research landscape of the WCT of Urinary and Male Genital Tumours, integrating four separate searches into a single megamap covering 4,138 included studies across 30 tumour entities. The findings reveal a markedly heterogeneous evidence base in dominance of prognostic and histopathological research, persistent synthesis and solid evidence gaps and a near-complete absence of predictive evidence sitting alongside domain-specific patterns shaped by disease prevalence, classification maturity and search precision.

Principal findings and clinical relevance

The overall pattern observed here, with prognosis and histopathology dominating the evidence base (>57% of all EGM entries), systematic reviews accounting for only 1 to 3% of included studies and predictive research representing no more than 2% of entries in any tumour group, is consistent with previously published maps of evidence cited in the WHO Blue Books for breast and thoracic tumours, developed within the WCT EVI MAP project.[9–11] The urological map both confirms these cross-tumour patterns and reveals domain-specific deviations. With only 30% solid evidence, 24% absolute gaps and 43% relative gaps, the extent of unmet evidential need across the landscape is highlighted. The pathogenesis literature for clear cell renal cell carcinoma constitutes a notable exception, likely reflecting decades of translational research in VHL/HIF pathway biology,[25] while the absence of predictive studies is strongly pronounced throughout the whole map.

Importantly, the clinical relevance of the evidence base may differ substantially across professional disciplines. For pathologists, studies addressing diagnosis and differential diagnosis may reflect the most immediately applicable in daily

practice and supposed 'low level' evidence may meaningfully inform morphological classification. In contrast, clinicians in urology and oncology may prefer to rely on middle to high level evidence concerning therapy, prognosis and imaging. This disciplinary divergence means that evidence gaps identified in the map carry different practical consequences depending on the professional context in which the map is used. It should also be noted that prognostic and predictive studies are often guided by therapeutic possibilities and needs, which means that they sometimes refer to tumour families rather than tumour types or subtypes that are not explicitly mentioned in the abstract. Consequently, they may not have been included in the current systematic, abstract-based screening approach, a fact that is also mirrored in the high exclusion rates for some tumours.

Clear cell renal cell carcinoma dominates the evidence base. This reflects both its prevalence as the most common renal cell carcinoma type (~75% of cases)[26] and the methodological advantage of its tumour type-specific search strategy, which yielded substantially higher screening inclusion rates of 34% compared to 1–10% for the combined searches used for the other tumour groups. Clear cell renal cell carcinoma and invasive urothelial carcinoma were the only two entities with high-level across all six tumour characteristics, including high-level predictive evidence, a finding that reflects its comparatively rich therapeutic infrastructure: among the entities mapped, they have the largest portfolio of available targeted agents and therefore the most developed prerequisites for predictive marker validation.[13,27]

For non-clear cell renal cell carcinoma, the concentration of evidence in histopathology combined with the highest proportion of low-level evidence per tumour group indicates predominantly descriptive research without the methodological infrastructure required for evidence-based classification. Absolute gaps for *ELOC*-mutated renal cell carcinoma, renal cell carcinoma NOS and tubulocystic renal cell carcinoma partly reflect the recent introduction of these entities in the fifth edition of the WCT in 2022.[3]

For glandular neoplasms of the prostate, the limited overall evidence base, with 254 studies feeding 320 EGM entries and a screening inclusion rate of less than 1%, contrasts sharply with the high global incidence of prostate cancer.[2] High screening exclusion rates partly reflect the widespread use of generic terminology ('prostate cancer') in primary literature without specification of the WCT-defined tumour group or differentiation of the explicit tumour type and, particularly at the level of abstracts, it was observed that clinical studies for glandular neoplasms of the prostate commonly fail to distinguish between ductal and acinar adenocarcinoma, providing this information in the full text only. As a result, studies relevant to a specific entity are not retrieved under entity-specific search terms, contributing to an apparent evidence deficit that does not reflect the actual publication landscape. Beyond this, research appears to be heavily concentrated on prostatic acinar adenocarcinoma at the expense of rarer entities. Prostatic cystadenoma and treatment-related neuroendocrine prostatic carcinoma were severely underrepresented with absolute gaps across three and four of the six characteristics, a particularly notable finding for the latter given its growing clinical relevance as an androgen-deprivation resistance mechanism.[28]

For urothelial tumours, the overall screening exclusion rate was high, once again reflecting the pervasive use of generic terminology as 'bladder cancer' in primary publication. Nevertheless, the group achieved the highest proportion of systematic reviews and compared to non-clear cell renal cell carcinoma a balanced distribution across non-predictive characteristics, with solid evidence across all six characteristics for invasive urothelial carcinoma. This, however, should be interpreted with caution: as the most general term within the group, 'invasive urothelial carcinoma' captures studies that do not further specify the specific type and includes various synonyms, likely inflating its apparent evidence base relative to the more narrowly defined entities. The relatively few studies on urothelial papilloma and inverted urothelial papilloma presumably reflect their benign clinical course rather than a genuine research deficit.

Two structural findings carry particular clinical relevance across all tumour groups. First, the predominance of low-level evidence in histopathology reflects the inherent reliance of morphological classification on descriptive observational research. For rare or recently described entities, prospective multicentre validation is rarely feasible and low evidence levels should be interpreted accordingly, rather than as an indication of poor scientific quality. These findings should be interpreted within the methodological scope of the EGM, which provides a descriptive and exploratory overview of the distribution of tumour characteristics and level of evidence retrievable within the predefined search, screening and coding strategy rather than a definitive assessment of methodological quality or clinical certainty. Accordingly, the EGM serves as a prioritisation tool that identifies areas where evidence is concentrated or limited and where further evidence synthesis or higher-level evidence, where feasible, may be warranted. Accordingly, differences in evidence density should be interpreted in the context of the methodological scope and limitations of the approach, as discussed in more detail below, and should not be taken to represent the complete body of available evidence. Second, the marked underrepresentation of predictive biomarker studies (≤2% across all groups except clear cell renal cell carcinoma) reflects a structural mismatch between the conventions of predictive research and those of tumour classification. Predictive studies are typically

embedded in clinical trials with therapeutic endpoints and their reporting conventions are not oriented toward tumour-type-specific classification. As molecularly targeted therapies, antibody-drug conjugates and checkpoint inhibitors increasingly shape urological oncology,[29–31] integrating predictive evidence into classification frameworks will require either revised eligibility criteria for future EGM iterations or new study designs that explicitly link therapeutic and classificatory endpoints.

Seven entities (*ELOC*-mutated renal cell carcinoma, tubulocystic renal cell carcinoma, renal cell carcinoma NOS, urothelial papilloma, inverted urothelial papilloma, prostatic cystadenoma and treatment-related neuroendocrine prostatic carcinoma) had fewer than 20 eligible studies, precluding meaningful evidence synthesis. For these entities, the 6th edition of the WCT will need to rely on expert consensus and explicitly grading this in the classification (analogous to evidence grading in clinical guidelines) would improve transparency.

Strengths and limitations

The strengths of this work include the integration of four parallel searches from two major literature databases into a single megamap, the comprehensive coverage of the important and more frequently encountered urological tumours within the defined time frame and the application of the standardised HETP framework[8] for evidence grading. The methodological approach is consistent with contemporary guidance for evidence and gap mapping, which emphasises systematic searching, transparent eligibility criteria, structured coding and descriptive presentation of the evidence base.[6,32,33] Furthermore, the systematic and interdisciplinary approach, which involves experts from the fields of epidemiology, systematic searches, systematic reviews, pathology, molecular biology and laboratory medicine, uses a predefined and published methodological protocol. This serves all aspects of screening, coding and mapping, as well as manual quality controls between reviewers at all levels and regular exchange between the groups, to ensure a defined quality of the presented EGMs and comparability of the results with other tumour types.

However, the complexity and scope of the approach led to the necessity to focus on essential information and procedures in this first round of EGM production, which entails several limitations: The most important of these is the fact that screening and coding were based exclusively on titles and abstracts. This approach was adopted as a pragmatic methodological choice to facilitate a large-scale evidence mapping exercise across multiple tumour entities. Although it enabled the systematic application of a uniform screening and coding strategy, it carries the risk that relevant studies were not captured if titles and abstracts did not contain sufficient methodological or tumour-specific information. Furthermore, exact reproduction of the screening and coding process may depend on access to identical bibliographic records and database indexing at the time the searches were performed. There was also a risk of misclassification if abstracts lack methodological details or if standards for abstract reporting are not adhered to, a situation frequently observed by the reviewers. When abstracts reported more than one study design, only the primary design was coded. This may have led to conservative evidence-level assignments in methodologically mixed studies. Future updates of the EGMs may benefit from targeted full-text review for selected evidence clusters or areas identified as potential gaps, as well as formal risk-of-bias assessments to strengthen the evaluation of the quality and reliability of the underlying evidence.

Cross-group comparisons should be interpreted cautiously: clear cell renal cell carcinoma was processed under broader inclusion criteria and a longer search time period (2018 to 2024 versus 2020 to 2025 for other groups), with 2018 to 2019 entries representing approximately 22% of clear cell renal cell carcinoma EGM entries. The harmonisation to a 2020 to 2024 window addresses this only partially: a residual asymmetry remains because the clear cell renal cell carcinoma search was executed on 5 December 2024 while the other searches extended into early 2025, meaning publications from late 2024 indexed only after 5 December 2024 may appear in the other groups but not in clear cell renal cell carcinoma. In addition, the visualized EGM entries do not directly correspond to the number of individual studies. A single record included for coding may contribute to multiple EGM entries by reporting on more than one tumour type and/or tumour characteristic. Consequently, the density of mapped entries should not be interpreted as representing independent studies but rather the distribution of retrieved evidence across the mapped dimensions.

Furthermore, the thresholds used for evidence-gap classification represent consensus-based operational definitions developed for this project. While they enabled a standardized approach across tumour groups, alternative threshold definitions are conceivable and may result in different classifications.

The strict application of WCT 5th edition terminology as the inclusion standard resulted in screening exclusion rates of 68–94% across three of four tumour groups, largely due to heterogeneous terminology in the primary literature. This heterogeneity leads to partial underrepresentation of entities whose designations changed between editions, newly recognised entities and benign entities. To some extent, this limitation was addressed by including MeSH terms alongside historical and current nomenclature in the search strategy. It is acknowledged, however, that the map is anchored to the

5th edition WCT framework by design and is intended to inform the classification process rather than to anticipate emerging entities. These limitations are likely to have the greatest impact on recently introduced tumour entities, entities whose nomenclature has changed between WCT editions and tumour characteristics that are inconsistently reported in titles and abstracts. During screening, it was noted that pathogenetic studies on clear cell renal cell carcinoma frequently used alternative designations such as "KIRC" (Kidney Renal Clear Cell Carcinoma), while studies on glandular neoplasms of the prostate often referred only to "prostate cancer" without specifying the corresponding WCT entity. Although these observations were not quantified, they illustrate how inconsistent terminology and the limited information available in titles and abstracts may contribute to incomplete evidence retrieval and should therefore be considered when interpreting apparent evidence gaps. Terminology-related challenges may be further amplified in low- and middle-income countries, where limited access to advanced molecular diagnostics reinforces reliance on histopathological classification, thereby exacerbating variability in terminology and reporting.

Several additional methodological boundaries shape the interpretation of evidence gaps. Pathogenesis was likely undercaptured: scoping decisions excluded epidemiology, including molecular epidemiology with pathogenetic overlap, and preclinical or mechanistic studies were excluded unless they explicitly addressed pathogenesis in the abstract. For rare tumour entities, even substantial increases in study volume may not translate into higher-level evidence, since the prospective multicentre designs required to reach Levels One to Three are often not feasible. Finally, neither publication bias nor the methodological quality of individual studies was formally assessed, consistent with EGM methodology but meaning that the map may disproportionately reflect positive findings, a methodological limitation that could be addressed in future EGM updates.

Conclusions, implications and future directions

This EGM systematically locates and characterises the variation in mapped research evidence for urinary tract and male genital tumours, identifying both areas where evidence is concentrated as well as areas where it remains sparse. It provides a transparent evidence-base for the development of the future editions of the WCT and serves as a priorisation tool by identifying areas where further evidence synthesis or higher-level evidence may be warranted. By distinguishing between areas where evidence is concentrated and those where it remains limited, it helps to inform where classification decisions can be supported by available evidence and where expert consensus must remain central.

Three priorities emerge for the research community. First, the gap identified across all four tumour groups may represent the most promising opportunity: commissioning targeted systematic reviews for tumour types with substantial primary literature but absent synthesis would translate existing studies into a classification-ready format with relatively little effort. Second, for entities where primary literature remains critically scarce, registry-based case accumulation within international pathology networks could offer a feasible route to higher-level evidence. Third, standardised adoption of the WCT nomenclature in primary publications and adherence to reporting guidelines (STARD, STROBE, REMARK, as promoted by the EQUATOR Network), including principles of clear and informative abstract reporting,[34] would substantially improve search precision and evidence retrievability in future EGM iterations. Greater alignment between tumour classification terminology and established standards for structured diagnostic pathology reporting, such as those developed by the International Collaboration on Cancer Reporting (ICCR), could improve the consistency and completeness of pathology reporting and thereby facilitate the identification and retrieval of relevant evidence in future evidence maps.[35] Furthermore, the gap between evidence grading criteria and the conventions of diagnostic pathology research suggests that the field may benefit from greater alignment with report standards established in clinical oncology, to ensure that high-quality diagnostic studies are recognised and retrieved as such in future evidence maps.

Assessing study quality beyond the assignment of study type ultimately requires full-text screening and data extraction and, where feasible, risk-of-bias assessment, pointing to an important area for methodological refinements in subsequent EGM iterations. This could also be expanded to other clinical domains, including epidemiology, clinical manifestation, anatomical localisation, aetiology and risk factors, to provide a more comprehensive picture of the urological cancer evidence landscape and better serve the full range of stakeholders who rely on the WCT as a classification standard.

Recent developments in large language model-based approaches, piloted in the present project, offer a promising route towards both goals: by enabling full-text screening at scale and more granular study classification, they may substantially reduce the resource requirements of future EGM iterations while improving the depth and precision of evidence retrieval.

Ethical approval and informed consent

This study did not involve human participants or animals; ethical approval and informed consent were not required.

Disclaimer

The content of this article represents the personal views of the authors and does not represent the views of the authors' employers and associated institutions. Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

The authors used Claude Sonnet 4.6, Claude Opus 4.7 and DeepL Pro to edit and improve the readability of the text. The AI tools were used solely to provide linguistic support. The authors created and verified all scientific content, interpretations, analyses and conclusions.

No role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Data availability

Underlying data

The underlying data is available online under WCT EVI MAP: Preliminary Results and on OSF.

Repository name: Open Science Framework ("Mapping the Evidence Base of Kidney, Prostate and Urinary Tract Tumours to Inform the WHO Classification of Tumours – Underlying and Extended Data") https://doi.org/10.17605/OSF.IO/M34RS.[14]

The project contains the following underlying data:

- Underlying_Data_EGM_urinary_and_male_genital_tumours.html (HTML file containing the interactive Evidence and Gap Map (EGM), available for download)

Extended data

The extended data is available online on OSF.

Repository name: Open Science Framework ("Mapping the Evidence Base of Kidney, Prostate and Urinary Tract Tumours to Inform the WHO Classification of Tumours – Underlying and Extended Data") https://doi.org/10.17605/OSF.IO/M34RS.[14]

The project contains the following underlying data:

- Extended_Data_Mapping_the_Evidence_Base_of_Kidney_Prostate_and_Urinary_Tract_Tumours_ to_Inform_the_WHO_Classification_of_Tumours.pdf (PDF file containing the adapted PRISMA flowcharts, high-level evidence heatmaps across tumour types, detailed terminology per tumour group, search strategies for Ovid Embase and PubMed, adapted Hierarchy of Evidence for Tumour Pathology (HETP), detailed distribution of evidence levels across tumour characteristics by tumour group, detailed distribution of EGM entries across tumour characteristics and levels of evidence by tumour type)

- automated_coding_loe.py (custom Python script automatically assigning LoE to the coded studies)

- .gitignore (configuration file specifying files and directories excluded from version control)

- README.md (documentation describing the purpose of the repository, dependencies and instructions for running the custom Python script)

- study_design_to_level.json (custom JSON file including the LoE assignments for each tumor characteristic and study design)

- Tests (folder containing test datasets and unit tests)

All data underlying the results are available as part of the article and its extended data. No additional datasets were generated or analysed. Together, the deposited underlying and extended data provide the materials required to reproduce the descriptive analyses, figures, and tables presented in this manuscript.

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Acknowledgements

Preliminary data was presented at the ISOBM 2025 conference in Murnau, Germany [Abstracts of the 47th Symposium of the ISOBM – A New Era of Biomarkers in Oncology. *Tumor Biology.* 2025;47. doi:10.1177/14230380251386526]. The authors would like to thank the International Agency for Research on Cancer (IARC) for supporting and leading the project, especially Cécile Monnier for her ongoing administrative support as project assistant. We would also like to thank Anastasia Koryakina from the EPPI Centre for her helpful guidance on EPPI-Reviewer® and her prompt support throughout the project, as well as the whole WCT EVI MAP group for the excellent collaboration.

References

1. Cree IA, Indave Ruiz BI, Zavadil J, *et al.*: **The International Collaboration for Cancer Classification and Research.** *Int. J. Cancer.* 2021; **148**: 560–571.
PubMed Abstract | Publisher Full Text | Free Full Text

2. Sung H, Ferlay J, Siegel RL, *et al.*: **Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.** *CA Cancer J. Clin.* 2021; **71**: 209–249.
PubMed Abstract | Publisher Full Text

3. *WHO Classification of Tumours Editorial Board, Who Classification of Tumours Editorial. Urinary and Male Genital Tumours.* WHO Classification of Tumours; 2022.

4. Kench JG, Amin MB, Berney DM, *et al.*: **WHO Classification of Tumours fifth edition: evolving issues in the classification, diagnosis, and prognostication of prostate cancer.** *Histopathology.* 2022; **81**: 447–458.
PubMed Abstract | Publisher Full Text | Free Full Text

5. Campbell F, Tricco AC, Munn Z, *et al.*: **Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): the same but different- the "Big Picture" review family.** *Syst. Rev.* 2023; **12**: 45.
PubMed Abstract | Publisher Full Text | Free Full Text

6. Fredlund M, Rogers M, Orr N, *et al.*: **What are the methodological characteristics of evidence and gap maps? A systematic review and evidence and gap map.** *Cochrane Evid. Synth. Methods.* 2024; **2**: e12096.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Indave BI, Colling R, Campbell F, *et al.*: **Evidence-levels in pathology for informing the WHO classification of tumours.** *Histopathology.* 2022; **81**: 420–425.
PubMed Abstract | Publisher Full Text

8. Colling R, Indave I, Del Aguila J, *et al.*: **A New Hierarchy of Research Evidence for Tumor Pathology: A Delphi Study to Define Levels of Evidence in Tumor Pathology.** *Mod. Pathol.* 2024; **37**: 100357.
PubMed Abstract | Publisher Full Text

9. Wong CJW, Md Nasir ND, Koh VCY, *et al.*: **Mapping the cited evidence of ductal carcinoma in situ from the 5th edition of the World Health Organisation classification of tumours of the breast.** *Histopathology.* 2024; **85**: 510–520.
PubMed Abstract | Publisher Full Text

10. Md Nasir ND, Koh VCY, Cree IA, *et al.*: **Phyllodes tumour evidence gaps mapped from the 5th edition of the WHO classification of tumours of the breast.** *Histopathology.* 2023; **82**: 704–712.
PubMed Abstract | Publisher Full Text

11. Giesen C, Del Águila MJ, Armon S, *et al.*: **Exploratory evidence maps for the WHO Classification of Tumours 5th edition for lung and thymus tumors.** *Virchows Arch.* 2024; **485**: 869–878.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Moch H, Amin MB, Berney DM, *et al.*: **The 2022 World Health Organization Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours.** *Eur. Urol.* 2022; **82**: 458–468.
PubMed Abstract | Publisher Full Text

13. Netto GJ, Amin MB, Berney DM, *et al.*: **The 2022 World Health Organization Classification of Tumors of the Urinary System and Male Genital Organs-Part B: Prostate and Urinary Tract Tumors.** *Eur. Urol.* 2022; **82**: 469–482.
PubMed Abstract | Publisher Full Text

14. Gruber S, Worf K, Trulson I, *et al.*: **Mapping the evidence base of kidney, prostate and urinary tract tumours to inform the WHO classification of tumours - extended data.** *OSF.* 2026.
Publisher Full Text

15. Holdenrieder S, Trulson I, Knoblauch L, *et al.*: **Study protocol for the development of an evidence gap map of the available evidence for diagnostic tumour characteristics associated with clear cell renal cell carcinoma.** 2025.
Reference Source

16. Holdenrieder S, Trulson I, Knoblauch L, *et al.*: **Study protocol for the development of an evidence gap map of the available evidence for diagnostic tumour characteristics associated with renal cell carcinoma.** 2025.
Reference Source

17. Holdenrieder S, Trulson I, Knoblauch L, *et al.*: **Study protocol for the development of an evidence gap map of the available evidence for diagnostic tumour characteristics associated with urothelial tumours.** 2025.
Reference Source

18. Holdenrieder S, Trulson I, Knoblauch L, *et al.*: **Study protocol for the development of an evidence gap map of the available evidence for diagnostic tumour characteristics associated with glandular neoplasms of the prostate.** 2025.
Reference Source

19. Cree I, Inskip A, Lokuhetty D, *et al.*: **Search strategies to obtain evidence for cancer classification [version 1; peer review: 1 approved].** *Open Res. Europe.* 2025; **5**: 369.
Publisher Full Text

20. Thomas J, Graziosi S, Brunton J, *et al.*: *EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis.* EPPI Centre, UCL Social Research Institute, University College London; 2023.

21. Page MJ, McKenzie JE, Bossuyt PM, *et al.*: **The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.** *Rev. Esp. Cardiol (Engl Ed).* 2021; **74**: 790–799.
PubMed Abstract | Publisher Full Text

22. Craciun OM, García-Ovejero E, Montes Mota M, *et al.*: **Decision tree and glossary for classifying study designs: interactive version (WCT EVI MAP Project).** 2024 [cited 2025 Nov 18].
Reference Source

23. Craciun OM, García-Ovejero E, Campbell F, *et al.*: **Development of a decision tree diagram for classifying study designs in tumour pathology research: a multidisciplinary approach.** *J. Pathol. Clin. Res.* 2026; **12**: e70056.
PubMed Abstract | Publisher Full Text | Free Full Text

24. **Abstracts of the 47th symposium of the ISOBM – A New Era of biomarkers in oncology.** *Tumour Biol.* 2025; **47**: 47.
Publisher Full Text

25. Nabi S, Kessler ER, Bernard B, *et al.*: **Renal cell carcinoma: a review of biology and pathophysiology.** *F1000Research.* 2018; **7**: 307.
PubMed Abstract | Publisher Full Text | Free Full Text

26. Feng X, Zhang L, Tu W, *et al.*: **Frequency, incidence and survival outcomes of clear cell renal cell carcinoma in the United States from 1973 to 2014: A SEER-based analysis.** *Medicine (Baltimore).* 2019; **98**: e16684.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Gkolfinopoulos S, Psyrri A, Bamias A: **Clear-cell renal cell carcinoma - A comprehensive review of agents used in the contemporary management of advanced/metastatic disease.** *Oncol. Rev.* 2021; **15**: 530.
Publisher Full Text

28. Imamura J, Ganguly S, Muskara A, *et al.*: **Lineage plasticity and treatment resistance in prostate cancer: the intersection of genetics, epigenetics, and evolution.** *Front. Endocrinol. (Lausanne).* 2023; **14**: 1191311.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

29. Cornford P, van den Bergh RCN, Briers E, *et al.*: **EAU-EANM-ESTRO-ESUR-ISUP-SIOG guidelines on prostate cancer-2024 update. Part I: Screening, diagnosis, and local treatment with curative intent.** *Eur. Urol.* 2024; **86**: 148–163.
PubMed Abstract | **Publisher Full Text**

30. Bedke J, Ghanem YA, Albiges L, *et al.*: **Updated European Association of Urology guidelines on the use of adjuvant immune checkpoint inhibitors and subsequent therapy for renal cell carcinoma.** *Eur. Urol.* 2025; **87**: 491–496.
PubMed Abstract | **Publisher Full Text**

31. Alfred Witjes J, Max Bruins H, Carrión A, *et al.*: **European Association of Urology guidelines on muscle-invasive and metastatic bladder cancer: Summary of the 2023 guidelines.** *Eur. Urol.* 2024; **85**: 17–31.
PubMed Abstract | **Publisher Full Text**

32. White H, Albers B, Gaarder M, *et al.*: **Guidance for producing a Campbell evidence and gap map.** *Campbell Syst. Rev.* 2020; **16**: e1125.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

33. Khalil H, Welch V, Grainger M, *et al.*: **Methodology for mapping reviews, evidence maps, and gap maps.** *Res. Synth. Methods.* 2025; **16**: 786–796.
PubMed Abstract | **Publisher Full Text**

34. Maslova K, Kowalewska M, Colling R, *et al.*: **From Data to Impact: How to Write an Effective Abstract for Cancer Research.** 2026. [under review].

35. ICCR: *Standardising structured pathology reporting of cancer.* ICCR; 2017.
Reference Source

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 16 July 2026

https://doi.org/10.21956/openreseurope.25903.r76785

© **2026 Rao B.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

B. Vishal Rao
[1] Basavatarakam Indo-Americal Cancer Hospital and Research Institute, Hyderabad, India
[2] Basavatarakam Indo-Americal Cancer Hospital and Research Institute, Hyderabad, India

From the standpoint of tumour classification, this paper represents a significant methodological advance. It introduces a transparent framework for identifying where evidence exists and where additional research may be needed. However, its conclusions should be interpreted with caution because the methodology measures **retrievable literature indexed through predefined search terms**, not necessarily the complete body of scientific evidence supporting a tumour entity.
In other words, the Evidence Gap Map is best viewed as a **research prioritization tool**, not as a definitive measure of the strength of evidence underlying WHO classification decisions. This distinction is essential when using the map to inform future editions of the WHO Blue Books.
The paper is valuable because it is the **first systematic evidence map** for urinary tract tumours. However, **the map should not be interpreted as a map of evidence—it is actually a map of retrievable evidence using a predefined search strategy.**

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 21 Aug 2026

Sophie Gruber

We thank the reviewer for this thoughtful and constructive comment and fully agree with this important distinction. In response, we revised the Discussion and Conclusion to consistently emphasize that the EGM represents a descriptive and exploratory overview of the mapped, retrievable evidence within the predefined search, screening and coding strategy. We now explicitly describe the EGM as a prioritisation tool intended to identify areas where evidence is concentrated or limited and where further evidence synthesis or higher-level evidence may be warranted, rather than as a definitive assessment of methodological quality, clinical certainty, or the complete body of available evidence. We also expanded the limitations section to discuss the influence of terminology, title- and abstract-based screening, and evidence retrieval on the interpretation of apparent evidence gaps.

Competing Interests: No competing interests were disclosed.

Reviewer Report 13 July 2026

https://doi.org/10.21956/openreseurope.25903.r76786

© 2026 Parada D. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

David Parada
[1] Universitat Rovira i Virgili, Reus, Spain
[2] Universitat Rovira i Virgili, Reus, Spain

Summary:
The authors present an Evidence and Gap Map (EGM) addressing urinary and male genital tumors, focusing on clear cell renal cell carcinoma, non-clear cell renal cell carcinomas, glandular neoplasms of the prostate, and urothelial tumors. Literature searches were performed in PubMed and Ovid Embase, limited to publications from 2020 to 2024. Records were coded by tumor type, tumor characteristics, and study design. Levels of evidence were determined using an adapted Version of the Hierarchy of Evidence for Tumor Pathology (HETP).

1) Is the work clearly and accurately presented and does it cite the current literature?

The manuscript is generally well written and logically organized. The rationale for developing an Evidence and Gap Map (EGM) to inform future updates of the WHO Classification of Tumors is clearly articulated. The study is contextualized within a relevant and contemporary body of literature on tumor classification, evidence mapping, and urological pathology. The authors reference appropriate methodological and clinical sources, including WHO classification documents, EGM methodology, PRISMA guidelines, and previous WCT EVI MAP publications.
However, several aspects of the presentation require clarification. The manuscript would benefit from a more explicit distinction between genuine evidence gaps and those arising from retrieval or terminology limitations. Studies employing generic terms such as “prostate cancer,” “bladder cancer,” or “renal cancer” may have been excluded if they did not utilize specific WHO terminology in their titles or abstracts. Although this limitation is acknowledged, its potential impact on the distribution of evidence warrants more prominent discussion.
Additionally, certain figures and the thresholds used for gap classification require clearer explanation for readers who may be unfamiliar with EGM methodology. While this is somewhat incomprehensible, the definitions and interpretations of “solid evidence,” “relative gaps,” and “synthesis gaps” should be made more accessible.
Recommended improvements:
- Clarify the distinction between the absence of evidence and failures in evidence retrieval.
- Provide a clearer explanation of how terminology limitations may influence the construction and interpretation of the map.
- Enhance the explanatory legends accompanying the EGM figures.
- Provide justification for the thresholds applied in classifying evidence gaps.
- Strengthen the discussion of contemporary literature on evidence mapping and standards for diagnostic pathology reporting.

2) Is the study design appropriate and does the work have academic merit?

The study design is consistent with the objective of mapping the evidence base, rather than conducting a traditional systematic review or meta-analysis. The Evidence Gap Map (EGM) approach effectively identifies clusters of evidence, gaps, and areas requiring further review. Given the breadth of the WHO Classification of Tumors and the diversity of tumor entities, the EGM represents a methodologically sound and valuable approach.
This study demonstrates academic merit by addressing a significant gap in evidence-informed tumor classification. It provides a valuable resource for WHO classification panels, pathologists, oncologists, urologists, and researchers. The inclusion of over 4,000 studies and 6,000 EGM entries further enhances its scope and relevance.
While the study design is appropriate, its implementation presents certain limitations. Exclusive reliance on title and abstract screening may result in misclassification and incomplete capture of evidence. Although these limitations do not undermine the overall study design, they may reduce confidence in specific findings.
Recommended improvements:
- Clarify that the EGM functions as a prioritization tool rather than a definitive assessment of evidence quality.
- Provide a more comprehensive justification for the use of title and abstract-only screening.
- Validate a representative sample through full-text review to improve reliability.

- Discuss the implications of the study design for interpreting findings related to both common and rare tumor entities.

3) Are sufficient details of methods and analysis provided to allow replication by others?

The manuscript presents a generally transparent description of the methods, detailing the databases searched, the harmonized publication period, the use of EPPI-Reviewer and EPPI-Mapper, the coding framework, the adapted HETP evidence hierarchy, and the availability of underlying and extended data.
Further clarification of certain methodological details is necessary to ensure reproducibility. The manuscript should provide a more detailed description of how the EPPI-Reviewer Priority Screening tool was used. Additionally, it is important to explain how the stopping rule based on 50 consecutive excluded records was validated, whether any eligible records were missed, and how reviewer oversight was managed.
The manuscript reports inter-rater agreement using percentage-based thresholds. For greater reproducibility, inter-rater agreement should be reported using Cohen's kappa or a comparable statistic. Additionally, clarification is needed regarding the handling of multi-coded studies, the resolution of ambiguous abstracts, and the process by which the automated Python script assigned levels of evidence.
Recommended improvements:
- Include additional detail regarding the validation and quality control procedures for Priority Screening.
- Report inter-reviewer agreement using Cohen's kappa or an equivalent robust statistic.
- Clarify the procedures for coding ambiguous or incomplete abstracts.
- Offer further explanation of the automated script used for level-of-evidence assignment.
- Ensure that sufficient detail is included to enable another research group to reproduce the coding and evidence-classification process.

4) If applicable, is the statistical analysis and its interpretation appropriate?

The manuscript primarily employs descriptive statistics, with minimal application of inferential methods. Descriptive analyses, including counts, proportions, and visual summaries, are suitable for an Evidence Gap Map (EGM). Tables, heatmaps, and bar plots effectively summarise the distribution of evidence across tumor groups, characteristics, and levels.
Greater caution is warranted in interpreting the descriptive data. The number of mapped studies or entries does not reflect methodological robustness or clinical certainty. In the absence of a formal risk-of-bias assessment, the study can identify evidence but cannot assess its reliability.
The rationale for the thresholds used to define absolute gaps, relative gaps, synthesis gaps, and solid evidence requires further justification. Implementing sensitivity analyses using alternative thresholds would enhance the robustness of the interpretation of the results.
Recommended improvements:
- Avoid suggesting that the quantity of evidence equates to its quality.
- Provide justification for the numerical thresholds applied in gap classification.
- Consider conducting sensitivity analyses with alternative gap thresholds.
- Clarify that the analysis is both descriptive and exploratory in nature.
- Include a caution regarding the interpretation of multi-coded entries, as these entries do not represent independent studies.

5) Are all the source data underlying the results available to ensure full reproducibility?

The provision of underlying and extended data via OSF, along with the interactive EGM and supporting materials, is a significant strength that enhances transparency.
However, achieving full reproducibility necessitates not only data availability but also the capacity to replicate screening, coding, and evidence-level assignment decisions. The manuscript would benefit from more comprehensive documentation of coding rules, decision trees, procedures for resolving reviewer disagreements, the automated Python script, and the relationship between exported EPPI-Reviewer data and the final EGM.
Since screening and coding were based solely on titles and abstracts, reproducibility is contingent upon access to identical bibliographic records and indexing status at the time of the search. This limitation should be acknowledged more explicitly in the manuscript.
Recommended improvements:
- Ensure that all search strategies, coding rules, and scripts are fully accessible and version-controlled.
- Provide more detailed instructions for reproducing the EGM from the raw exported data.
- Include documentation of excluded records and, where feasible, the reasons for their exclusion.
- Clarify whether the deposited data are sufficient to enable reproduction of all figures and tables.
- Provide detailed information on software versions and dependencies required for the Python scripts.

6) Are the conclusions drawn adequately supported by the results?

The main conclusions are generally supported by the results. The manuscript demonstrates that the evidence base varies across tumor groups, that histopathology and prognosis are predominant topics in the literature, that predictive evidence is limited, and that rare or recently introduced tumor entities have comparatively less mapped evidence.
Certain conclusions require more cautious presentation. The study design identifies mapped evidence gaps but does not establish a definitive absence of evidence. Because literature retrieval relied primarily on WHO terminology and title or abstract reporting, some gaps may reflect incomplete retrieval rather than an actual lack of research.
As no formal risk-of-bias assessment was performed, the authors should avoid implying that areas with a high volume of studies necessarily offer strong or reliable evidence. Conclusions should clearly differentiate between:
- Evidence volume,
- Evidence level as determined by study design,
- Methodological quality,
- Clinical applicability,
- Evidence retrievability.
Recommended improvements:
- Reframe conclusions to emphasize the mapped evidence gaps rather than to assert an absolute absence of research.
- Avoid overstating the robustness of evidence. Clearly distinguish between the availability of evidence and its quality. Evidence quality.
- Discuss which findings are most susceptible to terminology bias and limitations of abstract-only screening.
- Clarify the need for future full-text analyses and risk-of-bias assessments.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Uropathology, Kidney cancer, Prostate Cancer, Urothelial Carcinoma.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 21 Aug 2026

Sophie Gruber

We sincerely thank the reviewer for the careful evaluation of our manuscript and for the constructive comments. We appreciate the positive assessment of the study design and its academic merit. We have carefully considered all suggestions and revised the manuscript accordingly. Our responses to the individual recommendations are provided below. 1) Is the work clearly and accurately presented and does it cite the current literature?

- Clarify the distinction between the absence of evidence and failures in evidence retrieval.

We agree and revised the Discussion and Conclusion to distinguish mapped evidence gaps from a definitive absence of research. The manuscript now explicitly states that the EGM reflects the evidence retrievable within the predefined search, screening and coding strategy and should be interpreted within the methodological scope of the approach.

- Provide a clearer explanation of how terminology limitations may influence the construction and interpretation of the map.

We expanded the limitations section to discuss terminology-related bias and retrieval limitations.

- Enhance the explanatory legends accompanying the EGM figures.

We revised the legends of the EGM figures to improve readability by explaining the gap categories, colour coding, percentages, EGM entries versus independent studies and the interpretation of multi-coded entries where appropriate.

- Provide justification for the thresholds applied in classifying evidence gaps.

We clarified that the thresholds are consensus-based operational definitions developed by the WCT EVI MAP consortium to ensure a standardized approach across tumour groups. Their pragmatic nature are now discussed.

- Strengthen the discussion of contemporary literature on evidence mapping and standards for diagnostic pathology reporting.

We strengthened the discussion by incorporating contemporary evidence-mapping methodology and by highlighting established standards for structured diagnostic pathology reporting, including those developed by the ICCR, as relevant to improving evidence retrieval in future EGM iterations.

2) Is the study design appropriate and does the work have academic merit?

- Clarify that the EGM functions as a prioritization tool rather than a definitive assessment of evidence quality.

We revised the Discussion and Conclusion to explicitly describe the EGM as a descriptive and exploratory prioritisation tool rather than an assessment of methodological quality or clinical certainty.

- Provide a more comprehensive justification for the use of title- and abstract-only screening.

We expanded the Methods and Discussion to explain that title- and abstract-based screening was a pragmatic methodological decision made before screening to ensure the feasibility of a large-scale evidence mapping exercise while maintaining a standardized workflow throughout the whole WCT EVI MAP project. We agree that validation through full-text review would further strengthen the analysis. However, systematic full-text validation was beyond the scope of the present project because of feasibility considerations associated with the large number of records, including limitations in full-text availability; AI-supported tools that might have facilitated such large-scale processing were not available to the project at that time. We now discuss targeted full-text review and formal risk-of-bias assessment as important directions for future EGM iterations.

- Validate a representative sample through full-text review to improve reliability.

We agree that this would strengthen the methodology; however, systematic full-text validation was beyond the scope of the present project. We now discuss targeted full-text review together with formal risk-of-bias assessment as future methodological developments and know that other groups of the project are planning a publication on exactly this comparison.

- Discuss the implications of the study design for interpreting findings related to both common and rare tumour entities.

We expanded the Discussion to explain that low evidence levels for rare tumour entities

should not be interpreted as poor scientific quality, while high evidence volumes for common tumour entities should not be interpreted as proportionally stronger evidence. 3) Are sufficient details of methods and analysis provided to allow replication by others?

- Include additional detail regarding the validation and quality control procedures for Priority Screening.

We expanded the Methods by providing additional detail on the use of Priority Screening, reviewer oversight, validation and quality-control procedures.

- Report inter-reviewer agreement using Cohen's kappa or an equivalent robust statistic.

We thank the reviewer for this valuable suggestion. Formal inter-reviewer agreement statistics such as Cohen's kappa were not calculated. Initial screening and coding were performed according to the predefined protocol. During the second screening and coding step for quality assurance, records with disagreements or uncertain classifications were discussed among the reviewers with the participation of a third reviewer until consensus was reached. Accordingly, quality assurance relied on structured consensus rather than formal statistical assessment of inter-reviewer agreement. Given the descriptive and exploratory character of the present EGM, a formal statistical comparison of inter-reviewer agreement was considered beyond the scope of the current analysis and could be addressed in a future methodological study. We have clarified this reviewer consensus process in the Methods.

- Clarify the procedures for coding ambiguous or incomplete abstracts.

We expanded the Methods to describe how ambiguous or incomplete abstracts were handled and how reviewer disagreements were resolved.

- Offer further explanation of the automated script used for level-of-evidence assignment.

We expanded the Methods section to describe the automated Python script and its supporting documentation.

- Ensure that sufficient detail is included to enable another research group to reproduce the coding and evidence-classification process.

We expanded the methodological description and added material in the extended data on the definitions of the tumour characteristics and on the tumour types. In addition, all underlying data, extended data and methodological materials are publicly available through the associated repository. The overarching methodology of the project has also been described in dedicated methodological publications, providing further information on the development and application of the evidence-mapping approach. Together, these resources provide transparent access to the data and methods underlying the EGM and facilitate reproduction and further evaluation of the coding and evidence-classification process. 4) If applicable, is the statistical analysis and its interpretation appropriate?

- Avoid suggesting that the quantity of evidence equates to its quality.

We revised the Discussion and Conclusion to clearly distinguish evidence volume, evidence level and methodological quality.

- Provide justification for the numerical thresholds applied in gap classification.

We clarified that these thresholds are consensus-based operational definitions developed for methodological consistency across tumour groups.

- Consider conducting sensitivity analyses with alternative gap thresholds.

We agree that sensitivity analyses would be informative. However, they were beyond the scope of the current study and are acknowledged as an important direction for future methodological work.

- Clarify that the analysis is both descriptive and exploratory in nature.

We revised the Methods and Discussion to explicitly describe the analyses as descriptive and exploratory.

- Include a caution regarding the interpretation of multi-coded entries, as these entries do not represent independent studies.

We added this clarification to the Discussion and the relevant figure legends, explaining that EGM entries do not necessarily represent independent studies. 5) Are all the source data underlying the results available to ensure full reproducibility?

- Ensure that all search strategies, coding rules, and scripts are fully accessible and version-controlled.

We clarified the availability of the search strategies, coding framework, scripts and repository documentation within the OSF repository.

- Provide more detailed instructions for reproducing the EGM from the raw exported data.

We expanded the documentation referenced in the Data Availability section and clarified the workflow from exported data to the final EGM.

- Include documentation of excluded records and, where feasible, the reasons for their exclusion.

We considered this suggestion. Individual excluded records were not retained as a project output; however, the screening workflow and exclusion process are now described in greater detail.

- Clarify whether the deposited data are sufficient to enable reproduction of all figures and tables.

We revised the Data Availability statement to explicitly state that the deposited data and code are sufficient to reproduce the descriptive analyses, figures and tables presented in the manuscript.

- Provide detailed information on software versions and dependencies required for the Python scripts.

We clarified software dependencies and information on the automated script in the Methods and explicitly refer to the accompanying OSF repository. 6) Are the conclusions drawn adequately supported by the results?

- Reframe conclusions to emphasize the mapped evidence gaps rather than to assert an absolute absence of research.

We revised the Discussion and Conclusion to consistently refer to mapped evidence gaps and to avoid implying an absolute absence of research.

- Avoid overstating the robustness of evidence. Clearly distinguish between the availability of evidence and its quality.

We revised the Discussion and Conclusion to distinguish evidence availability, evidence level, methodological quality and clinical certainty and removed wording that could imply an

assessment of evidence robustness.

- Discuss which findings are most susceptible to terminology bias and limitations of abstract-only screening.

We expanded the limitations section in Discussion with concrete examples observed during screening to illustrate which findings are most susceptible to terminology bias and abstract-only screening.

- Clarify the need for future full-text analyses and risk-of-bias assessments.

We now explicitly discuss targeted full-text review and formal risk-of-bias assessment as important future methodological developments.

Competing Interests: No competing interests were disclosed.