TITLE: A standard operating procedure to enable global, regionally-guided precision oncology knowledge bases

AUTHORS: Brendan Reardon[1,2,3], Sabrina Y. Camp[1,2], Grainne O'Mahoney[3], Matías Vergara[1,2], Mark Lawler[4], Eliezer M. Van Allen[1,2,†], Aedin C. Culhane[3,†]

AFFILIATIONS: Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA[1]; Broad Institute of MIT and Harvard, Cambridge, MA, USA[2]; School of Medicine, Limerick Digital Cancer Research Centre, Health Research Institute, University of Limerick, Limerick, Ireland[3]; Patrick G Johnston Centre for Cancer Research, Faculty of Medicine, Health and Life Sciences, Queen's University Belfast, Belfast, UK[4]

[†]**email**: Aedin C. Culhane: aedin.culhane@ul.ie; Eliezer M. Van Allen: EliezerM_VanAllen@dfci.harvard.edu

[H1] Abstract

Precision oncology knowledge bases are central to interpreting cancer genomic variants and associating them with clinical implications, guiding therapeutic selection and molecular tumor board deliberations. Although clinical evidence informing these resources are broadly globally applicable, drug approvals and patient access to therapies are governed by regional regulatory and health system authorities. Currently, publicly available knowledge bases primarily represent regulatory approvals from the United States (US) Food and Drug Administration (FDA). Consequently, they may recommend inaccessible therapies or omit regionally approved options, limiting clinical utility in non-US settings. This disparity highlights an urgent need for region-specific precision oncology knowledge bases that account for local regulation and access.

Here, we outline essential steps for establishing region-specific precision oncology knowledge bases by curating cancer drug approvals that have received local market authorization or reimbursement within their local health system. We do so in a project-agnostic manner to encourage distributed deployment, while also recommending the adoption of representation standards to support global interoperability. We envision a federated ecosystem of regionally contextualized knowledge bases that enable comparative analyses, which may provide a means to identify regional differences in treatment eligibility and pharmacoeconomic determinants of access. Building such frameworks will advance equitable precision oncology and ensure that genomic-driven cancer care benefits patients globally.

[H1] Keywords

Precision oncology, Knowledge bases, Tumour Biomarkers

[H1] Introduction

Precision oncology knowledge bases (for example, OncoKB, CIViC, Cancer Genome Interpreter, and the Molecular Oncology Almanac, among others) serve a substantial role in the interpretation of cancer variants(1–7). They make relationships between biomarkers, cancer types, and therapies programmatically accessible, enabling the annotation of biomarkers for clinical and biological relevance, and facilitating interpretation for actionability in the context of an individual patient's unique cancer. Although these resources are publicly available and globally used, they largely only curate regulatory approvals from the United States (US) Food and Drug Administration (FDA).

Globally, heterogenous market authorizations and health system access decisions result in differential access to companion diagnostics and cancer medicines across countries(8–11). While the results in supporting clinical trials are clear, agencies can have differing interpretations of the efficacy achieved through the trial, resulting in diverging approvals(12,13). Thus, implementations of precision oncology require regional personalization to efficiently identify cancer drugs that can be administered to patients with cancer, based on their local regulatory and health system landscapes. Relying on knowledge resources that only contain

approvals from other regions ultimately slows the clinical interpretation of cancer variants, and contributes to precision oncology's interpretation bottleneck(14).

Here, we describe a procedure for curating cancer drug regulatory approval data to develop region-specific precision oncology knowledge bases and related resources. We do so through several parts of the process: defining the scope of the curation effort, identifying and organizing relevant documents, recording approved indications and uses of drugs and treatment regimens, and deriving precision oncology relationships. When beginning to curate region-specific approvals, we recommend creating a document that addresses points raised within this proposed procedure that can be shared and iterated upon by collaborators and project stakeholders.

[H1] Define scope

We recommend first defining the scope of the project, specifically including the scope of what will or will not be curated. At minimum, this will involve identifying relevant organizations specific to the region of interest. This may include medicine licensing agencies and separate authorities responsible for funded access within a health system. Some countries or regions concentrate these roles in a single ministry of health, while others distribute them across multiple agencies. Furthermore, access to cancer treatments may be managed by separate organizations, depending on if a country or region has socialized medicine, private insurance, or both.

Market authorization, or licensing, of cancer drugs and treatment regimens are often dictated by governmental regulatory agencies(12). Market authorization is required for a cancer drug or treatment regimen to be legally administered to patients with cancer within the region for specific clinical contexts that have been approved. These contexts that a medicine is approved for is referred to as treatment indications. For example, oncology approvals within the US are dictated by the FDA, while both the European Medicines Agency (EMA) and national registers (such as the Swedish Medical Products Agency) may authorize medicines within the European Union (EU) and member states.

Health Technology Assessment (HTA) is a structured way for health systems to evaluate which technologies to introduce, reimburse, or phase out, including medicines(15). Private insurers in some countries are either legally mandated or have elected to cover cancer treatments that have received market authorization, as is the case within the United States due to the Affordable Care Act. However, in other regions, health payers, whether public or private, must evaluate cancer drugs following market authorization to decide upon access within their respective health system. The evaluation examines the cost-benefit for each approved indication based on the monetary burden of the treatment relative to the estimated benefit(16–19). The delay between authorization of treatments and approval of access can be variable, referred to as time to reimbursement, and a source of disparity for cancer care(20). One study of EU countries observed a median delay of 407 days, with a maximum observed as 2320 days(21). Both private and public health systems undertake HTA assessment, and thus delays

in time to reimbursement is a source of disparity in cancer care between and within regions(9,11,15).

Once the source(s) of approved indications are identified, the types of clinical implications of interest should be defined in advance of an initial pilot. For example, will the project curate all approved indications for every cancer drug? Or only approved indications involving at least one biomarker? Is there a focus on solid tumors? Current publicly available precision oncology knowledge bases largely only cover approvals that require least one biomarker for both solid tumors and hematological malignancies. This constraint has been chosen due to the knowledge bases' intent of supporting the interpretation of genomic data from patients with cancer. Alternatively, choosing to curate cancer drug approvals that do not require the presence or absence of a biomarker(s) may require additional resource investment to perform an initial curation and maintenance, due to the increased quantity of approvals. Furthermore, market authorizations and HTA approvals primarily cover clinical implications of therapeutic response for sensitivity, but increasingly molecular definitions are being deployed to diagnose types of cancer and may be included within an agency's documentation. Relationships related to resistance to therapeutic response and disease prognosis are often described within clinical guidelines from professional organizations and journal articles rather than from regulatory bodies(22–24).

[H1] Identify and organize relevant documents

Regardless of an organization's procedure for publishing and sharing their regulatory approvals, curation of their supporting documents will follow a consistent set of steps, where each document can be systematically identified, organized, screened, and cited. Performing an initial review of documents is the most labor-intensive component of launching an oncology knowledge base. It requires manually accessing, organizing, and reviewing hundreds, if not thousands, of documents to comprehensively identify those relevant to the defined scope. Completing a thorough characterization of source materials before proceeding will provide assurance that the chosen landscape is fully captured. With this foundation, subsequent steps can be performed in parallel by curators for the defined set of documents.

[H2] Identification

While the specifics of how each agency hosts documents may differ, there appears to be a few models: look up by drug or regimen name, search by criteria, or reference a curated table. For example, the FDA electronically publishes product labels for each approved Biologic License Application (BLA) and maintains a publicly accessible database of these, versioned by publication date. This database, Drugs@FDA, allows look up by drug name, active ingredient, or BLA number. In contrast, the EMA publicly shares the current version of each product's summary of product characteristics (label as "European Public Assessment Report [EPAR]") and has separate documents for detailing each modification over time (labeled as "EPAR Assessment report - Variation"). The EMA provides a complete index of medicines sorted by modification date and allows users to search and filter results by several categories such as

“therapeutic area (cancer)” and date last updated. Ireland’s public health system agency, the Health Service Executive, maintains a curated table of cancer drugs approved for reimbursement on their website that links to specific treatment regimen documents. Note that medicine brand (trade) names are often region specific, even when the underlying active ingredient and formulation are identical. This can be due to trademarks, regulatory rules, language, or marketing. Therefore record both brand and generic names, and therapy name normalization services may be helpful(25).

[H2] Organization

Document source metadata should be organized and systematically referenced into either a spreadsheet or database for tracking and data provenance purposes. If the organization makes documents available through a look up, similar to Drugs@FDA, it may be worthwhile generating a list of expected treatments to attempt to find. Other oncology knowledge bases and journal articles that detail specific cancer drugs, such as Falzone et al. 2023 and Suehnholz et al. 2023(6,26), are useful supporting resources. This creates a benchmark to ensure that the process reached complete coverage.

[H2] Screening

Documents for cancer drugs and treatment regimens will contain a section on approved “indications and usage”. We recommend performing an initial screening of each documents’ section on approved “indications and usage” to label each document as “relevant” or “not relevant”, depending on the previously agreed upon scope, before proceeding to knowledge curation (Supplementary Table 1). Furthermore, it may be helpful to track all documents reviewed, regardless of if they are relevant to the chosen scope or not. Doing so has two benefits: it facilitates sharing outcome metrics and statements such as, “Over 500 published product labels were manually reviewed to identify 149 cancer drugs that include at least one approved usage involving a biomarker”, while also serving as a historical record of previously reviewed documents for future clinical variant curators within the project.

[H2] Citation

A goal of oncology knowledge bases is to expediate access to information that is otherwise not structured and, by virtue of not originally creating that knowledge, it is important to directly cite source materials. The citation format for each type of document(s) detailing approved indications and usages should be identified. This should contain details such as the manufacturer, drug name(s), document type, access URL, publication date, and date of access. Agencies may or may not have established guidance on their preferred citation style. If not, we recommend using this as an opportunity to initiate dialogue with them.

[H1] Record approved indications and usages

In a separate database table or spreadsheet from document metadata, approved treatment indications and relevant metadata should be structured and recorded (Supplementary Table 2).

At minimum, this should contain the approved “indications and usage” text, exactly as written, and a reference to the document that the indication originates from. It is worthwhile curating approval-specific metadata. For example, an associated document that initially announces or describes the approval, an initial approval date, or current approval status. Representing the current status enables filtering and eventual deprecation of treatment indications as the regulatory landscape changes (Table 1). Public health systems may provide additional details worthwhile curating such as the relevant reimbursement mechanism.

[H1] Derive genomic knowledge statements

Next, derive precision oncology relationships between biomarkers, cancer types, and therapies for each approved indication to construct a structured, programmatically accessible representation. Specific details of this process may manifest differently depending on the implementation and chosen data model. Previously published standard operating procedures for genomic knowledge bases may inform design decisions(27). We recommend adhering to five best practices: include a human readable description of the approval, detail how the project handles ambiguity, define concept relationships, utilize representation standards, and align with clinical actionability reporting classifications.

First, write a description text to describe genomic knowledge statements derived from an underlying indication (Table 2). Rather than simply restating the “indication and usage” text, the description should provide context such as the originating agency and current approval status. Additionally, the description may rephrase the indication text to align with implementation-specific style preferences such as aligning with the American Society for Clinical Oncology (ASCO)’s Language of Respect(28,29) or localizing vocabulary; for example, American or British English spelling of tumo(u)r.

Second, approvals can include biomarkers, cancer types, or therapies with ambiguity that can be clarified and expanded upon within an accompanying description text for knowledge base use. For example, a therapy may be approved for use in combination with others, including chemotherapy broadly or a therapy strategy such as aromatase inhibition. In these cases, it is a design decision on behalf of the project whether to curate a concept exactly as presented, such as curating aromatase inhibition as is, or deriving separate statements for each possible therapy that satisfies the therapy group. Approvals may also be for categories of genomic variants; such as, *BRAF* V600 or *EGFR* exon 19 deletion variants. If deciding to clarify, we recommend referring to the underlying clinical trial supporting an approval. The trial will often clarify ambiguous components of the approval by stating what choice of therapies were available to prescribe to patients within the trial, or what specific biomarkers were required for enrollment. This information is often listed within a cancer drug’s market authorization document.

We also strongly recommend utilizing a relational database structure to derive relationships from them (Fig. 2). Storing each concept separately will enable curators to reference them by an identifier as they construct knowledge statements, rather than repeatedly detail them. At minimum, each approved indication will be result in at least one knowledge statement consisting

of at least one biomarker, cancer type, and therapy, accompanied by a human readable description of the approval.

Next, we recommend following existing standards for representing derived genomic knowledge statements, such as those from the Global Alliance for Genomics and Health's Genomic Knowledge Standards (GKS) Work Stream(30–32). Their Variant Annotation Specification product is an information model suited to this task and is gaining adoption through several information systems such as ClinVar, MaveDB, and MetaKB(3,33,34). Following existing standards and FAIR (Findable, Accessible, Interoperable, and Reusable) principles enables interoperability between curated precision oncology relationships and other services, while also reducing adoption requirements for data consumers(35,3,36). Two core data elements of GKS schemas are the "Coding" and "Mappable Concept" entities, which facilitate structured representations of concepts (e.g., genes, cancer types, and therapies) as defined by other systems and terminologies. Users of the resulting knowledge base can then interpret content confidently due to concepts referencing pre-existing ontologies (Table 3)(31,37–47). For example, aligning gene symbols with the Hugo Gene Nomenclature Committee reduces the likelihood of presenting outdated or alias gene symbols, and thus information potentially overlooked by users. Several implementation-specific decisions will need to be made for the representation of genomic concepts such as primary reference genome build choice and transcript selection(27). Although the cancer genomics community has not yet uniformly transitioned to GRCh38, the use of Matched Annotation from NCBI and EMBL-EBI (MANE) Select transcripts are recommended for clinical sequencing. These transcripts are defined relative to the GRCh38 reference assembly and mapped coordinates are provided for compatibility with GRCh37 (hg19)(48,49).

Finally, derived precision oncology relationships should further be aligned with clinical actionability reporting standards developed by professional organizations. Specifically, the joint consensus recommendation from the Association for Molecular Pathology, ASCO, and College of American Pathologists (AMP/ASCO/CAP) and the European Society for Medical Oncology's Scale for Clinical Actionability of molecular Targets (ESCAT) provide guidelines for classifying and sharing therapeutic response statements based on their supporting evidence(50,51). In both cases, relationships derived from market authorizations are likely to meet their highest categories: Tier 1 and Level A evidence, and Tier 1 for AMP/ASCO/CAP guidelines and ESCAT, respectively. While market authorization or inclusion within clinical guidelines is sufficient to be classified as the highest strength of evidence for AMP/ASCO/CAP, ESCAT further delineates by considering the results from and type of the underlying clinical trial that supports an approval. If curating approvals that dictate access within a health system, additional investigation may be required to align to these standards if the approval is for an off-label use of a cancer drug.

[H1] Detail a standard operating procedure

We encourage curators to write a standard operating procedure (SOP) that details decisions and procedures related to their organization's knowledge curation. The SOP should contain

details of how cited documents were initially identified and sourced, labeled as relevant or not to the project, what metadata was recorded for each, content update frequency, and testing procedures to assess quality of the knowledge base. This should likewise be performed for each organization of interest.

The SOP should detail the data schema being followed by the curation effort. Required and optional fields for each table should be listed and described, including their expected data type. Using an established specification is recommended and can expediate the generation of this documentation. However, even when doing so, it is worthwhile to separately detail the schema to reduce external dependency and document any modifications made.

Project choices for deriving genomic knowledge statements should be included as well. For example, reference assembly used to represent sequence variants, transcript choice, term localization or normalization, and how ambiguous terms or missing data are interpreted. The external vocabularies used for biomarkers, cancer types, and therapies should also be listed, as well as the procedure for how the project generates mappings for each of them.

Drafting this SOP either prospectively before an initial effort or early on will allow the project to iterate on its procedures as it matures and should facilitate ongoing updates to content curation over time. Ideally, an organization may have a mechanism to review updates relative to a date. For example, the FDA manages a web page and email list for Oncology / Hematological Malignancies Approval Notifications to alert subscribers to new approvals and revisions. Likewise, the EMA allows filtering medicines by date across several criteria: first published, last updated, and market authorization date. Auxiliary code tailored to an agency's knowledge distribution can be written to help scrape, cache, and curate approvals in absence of a formal update mechanism existing. The SOP should thus likewise detail a procedure for staying current with each organization's approvals and how often it will be updated.

Importantly, the knowledge base schema should be versioned separately from knowledge base content. The schema, as a software product, can be managed through Semantic Versioning (SemVer) to detail major, minor, and patch changes. Following versioning conventions provides data consumers with confidence in the reliability of the resulting knowledge base. Communicating content releases does not have a standard as ubiquitous as SemVar but there are well established conventions. For example, maintaining a changelog summarizing changes per data release. Labeling releases by date is reasonable, especially because regulatory approvals are temporally defined. Additionally, a testing suite should be written to reduce the likelihood of curator-introduced errors within the database; for example, a mismatch of URLs included within a citation and the document object. With both schema construction and content population, following FAIR principles and being transparent with chosen policies will improve the utility of the resulting knowledge base and project.

[H1] Discussion

This standard operating procedure details logistical and practical considerations for how to curate and develop region-specific precision oncology knowledge bases. Once curated, this content can be made accessible both programmatically through an application program interface (API) and interactively through a web-based browser. An API can enable high-throughput, programmatic annotation of biomarkers to facilitate interpretation within molecular diagnostic workflows. In contrast, web-based browsers can provide an intuitive, user friendly interface for users to interact with the content. Based on both, the project should iterate with stakeholders and contacts within relevant organizations to ensure that information is being represented accurately and being made accessible in a useful way.

The curation of genomic knowledge from any source of evidence is labor intensive and requires ongoing maintenance to stay current. This is especially true for pilot curation efforts of regulatory approvals, as they require an initial review of hundreds of documents and indications. To date, the field has relied on the manual curation of this information from variant curation scientists. The inclusion of machine learning and large language models into these workflows may improve efficiency of staying current and supporting expansion into new regions(52). Despite the effort, we believe this to be a worthwhile effort and encourage variant scientists to do so.

We acknowledge that there are barriers and limitations to curating region-specific precision oncology genomic knowledge. First, our procedure presumes that these approvals are publicly viewable online, and this may not be the case. Second, it also assumes that these approvals exist within documents that can be cited; agencies may list their approvals as a collection on a single website within a table, for example. Agency contacts may be able to provide a record of historical approvals and transition to their documents to be public facing. Third, while identifying agencies responsible for market authorization is usually straightforward, identifying the appropriate organization(s) for access within a health system is often more complex. In some regions, coverage policy decisions and their publication may be handled by separate agencies; for example, in Northern Ireland the Department of Health is the decision-making agency for public use and the Northern Ireland Drug Formulary publishes those decisions. Depending on a professional's relative entry point with a health system, they may prefer the project to cite a different set of documents. To reduce this friction, we recommend iterating with stakeholders and prospective users to ensure that appropriate evidence items are being used.

Cancer drug approvals are region-specific, and precision oncology knowledge bases should be too. Advancement towards this ideal will enable the promise of precision oncology to be implemented globally, matching patients with therapies tailored to their molecular profiles that they can receive, regardless of where they live.

[H1] Abbreviations

ASCO: American Society for Clinical Oncology

API: Application Program Interface
AMP: Association for Molecular Pathology
BLA: Biologic License Application
CAP: College of American Pathologists
EMA: European Medicines Agency
EPAR: European Public Assessment Report
ESCAT: European Society for Medical Oncology's Scale for Clinical Actionability of molecular Targets
ESMO: European Society for Medical Oncology
EU: European Union
FAIR: Findable, Accessible, Interoperable, and Reusable
FDA**:** Food and Drug Administration
GKS: Genomic Knowledge Standards
HTA: Health Technology Assessment
SemVer: Semantic Versioning
SOP: Standard Operating Procedure
US**:** United States

[H1] Declarations

[H2] Ethics approval and consent to participate

Not applicable

[H2] Consent for publication

Not applicable

[H2] Availability of data and materials

All display items and supplementary information have been made publicly available on GitHub: https://github.com/vanallenlab/region-specific-approvals-sop.

[H2] Competing interests

B. Reardon has filed institutional patents on methods for clinical interpretation. E.M.V.A. has received research support (to institution) from Novartis, BMS, Sanofi, and NextPoint. E.M.V.A. serves as a consultant or on scientific advisory boards of Novartis Institute for Biomedical Research, Serinus Bio, and TracerBio. E.M.V.A. has equity in Tango Therapeutics, Genome Medical, Genomic Life, Enara Bio, Manifold Bio, Microsoft, Monte Rosa, Riva Therapeutics, Serinus Bio, Syapse, and Tracer Bio. E.M.V.A. has received speaking fees from TD Cowen. E.M.V.A. has filed institutional patents on chromatin mutations and immunotherapy response, and methods for clinical interpretation, and has intermittent legal consulting on patents for

Foaley Hoag LLP. E.M.V.A. serves on the editorial board of Science Advances. The remaining authors declare that they have no competing interests.

[H2] Funding

Funded by the European Union's Digital Europe Programme (GDI grant number 101081813), European Union’s Horizon Europe programme (CANDLE grant number 101214368) and Government of Ireland North-South Research Programme, a collaborative scheme funded through the Government’s Shared Island Fund that is administered by the Higher Education Authority (HEA) on behalf of the Department of Further and Higher Education, Research, Innovation and Science.

[H2] Authors’ contributions

B.R. wrote the original draft of the manuscript. All authors read, revised, and approved the final manuscript.

[H2] Acknowledgements

We thank B. Amos, L. Babb, K. Kuzma, D. Puthawala, and A. Wagner for their help and continued engagement in introducing us to GA4GH's Genomic Knowledge Standards products.

[H2] Authors’ information

Aedin C. Culhane and Eliezer M. Van Allen are joint senior authors.

[H1] References

1. Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precision Oncology. 2017 Nov;1:1–16.

2. Griffith M, Spies NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat Genet. 2017 Feb;49(2):170–4.

3. Wagner AH, Walsh B, Mayfield G, Tamborero D, Sonkin D, Krysiak K, et al. A harmonized meta-knowledgebase of clinical interpretations of somatic genomic variants in cancer. Nat Genet. 2020 Apr;52(4):448–57.

4. Reardon B, Moore ND, Moore NS, Kofman E, AlDubayan SH, Cheung ATM, et al. Integrating molecular profiles into clinical frameworks through the Molecular Oncology Almanac to prospectively guide precision oncology. Nat Cancer. 2021 Oct;2(10):1102–12.

5. Krysiak K, Danos AM, Saliba J, McMichael JF, Coffman AC, Kiwala S, et al. CIViCdb 2022: evolution of an open-access cancer variant interpretation knowledgebase. Nucleic Acids Research. 2023 Jan 6;51(D1):D1230–41.

6. Suehnholz SP, Nissan MH, Zhang H, Kundra R, Nandakumar S, Lu C, et al. Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. Cancer Discovery. 2023 Oct 18;14(1):49–65.

7. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med. 2018 Dec;10(1):25.

8. Mateo J, Steuten L, Aftimos P, André F, Davies M, Garralda E, et al. Delivering precision oncology to patients with cancer. Nat Med. 2022 Apr;28(4):658–65.

9. Li M, Ka D, Chen Q. Disparities in availability of new cancer drugs worldwide: 1990-2022. BMJ Glob Health. 2024 Aug;9(9):e015700.

10. Hofmarcher T, Charalambous A, Normanno N, Szmytke E, Wilking N. Access to novel cancer medicines in Europe: inequities across countries and their drivers. ESMO Open [Internet]. 2025 Oct 1 [cited 2025 Dec 19];10(10). Available from: https://www.esmoopen.com/article/S2059-7029(25)01679-5/fulltext

11. Cai JX, Wang SY, Hu H, Ung COL, Li FX, Lin TF, et al. Disparities in the access to immune checkpoint inhibitors approved in the United States, the European Union and mainland China: a serial cross-sectional study. bmjph [Internet]. 2025 Mar 13 [cited 2025 Dec 19];3(1). Available from: https://bmjpublichealth.bmj.com/content/3/1/e001995

12. Kashoki M, Hanaizi Z, Yordanova S, Veselý R, Bouygues C, Llinares J, et al. A Comparison of EMA and FDA Decisions for New Drug Marketing Applications 2014–2016: Concordance, Discordance, and Why. Clinical Pharmacology & Therapeutics. 2020;107(1):195–202.

13. Rohr UP, Iovino M, Rudofsky L, Li Q, Juritz S, Gircys A, et al. A decade comparison of regulatory decision patterns for oncology products to all other non-oncology products among Swissmedic, European Medicines Agency, and US Food and Drug Administration. Clinical and Translational Science. 2023;16(9):1569–81.

14. Good BM, Ainscough BJ, McMichael JF, Su AI, Griffith OL. Organizing knowledge to enable personalization of medicine in cancer. Genome Biology. 2014 Aug 27;15(8):438.

15. Jenei K, Raymakers AJN, Bayle A, Berger-Thürmel K, Cherla A, Honda K, et al. Health technology assessment for cancer medicines across the G7 countries and Oceania: an international, cross-sectional study. The Lancet Oncology. 2023 June;24(6):624–35.

16. Siddiqui M, Rajkumar SV. The High Cost of Cancer Drugs and What We Can Do About It. Mayo Clin Proc. 2012 Oct;87(10):935–43.

17. Lawler M, French D, Henderson R, Aggarwal A, Sullivan R. Shooting for the Moon or Flying Too Near the Sun? Crossing the Value Rubicon in Precision Cancer Care. Public Health Genomics. 2016;19(3):132–6.

18. Goldstein DA, Stemmer SM, Gordon N. The cost and value of cancer drugs – are new innovations outpacing our ability to pay? Isr J Health Policy Res. 2016 Sept 22;5:40.

19. Cherla A, Renwick M, Jha A, Mossialos E. Cost-effectiveness of cancer drugs: Comparative analysis of the United States and England. eClinicalMedicine. 2020 Nov 5;29–30:100625.

20. Post HC, Schutte T, van Oijen MGH, van Laarhoven HWM, Hollak CEM. Time to reimbursement of novel anticancer drugs in Europe: a case study of seven European countries. ESMO Open. 2023 Apr 6;8(2):101208.

21. Post HC, Schutte T, Oijen MGH van, Laarhoven HWM van, Hollak CEM. Time to reimbursement of novel anticancer drugs in Europe: a case study of seven European countries. ESMO Open [Internet]. 2023 Apr 1 [cited 2025 Dec 17];8(2). Available from: https://www.esmoopen.com/article/S2059-7029(23)00433-7/fulltext

22. Pavlidis N, Hansen H, Stahel R. ESMO Clinical Practice Guidelines: development, implementation and dissemination. Annals of Oncology. 2010 May 1;21:v7–8.

23. Birkeland ML, McClure JS. Optimizing the Clinical Utility of Biomarkers in Oncology: The NCCN Biomarkers Compendium. Arch Pathol Lab Med. 2015 May;139(5):608–11.

24. Pluchino LA, D'Amico TA. National Comprehensive Cancer Network Guidelines: Who Makes Them? What Are They? Why Are They Important? Ann Thorac Surg. 2020 Dec;110(6):1789–95.

25. Cannon M, Stevenson J, Kuzma K, Kiwala S, Warner JL, Griffith OL, et al. Normalization of drug and therapeutic concepts with Thera-Py. JAMIA Open. 2023 Dec 1;6(4):ooad093.

26. Falzone L, Bordonaro R, Libra M. SnapShot: Cancer chemotherapy. Cell. 2023 Apr;186(8):1816-1816.e1.

27. Danos AM, Krysiak K, Barnell EK, Coffman AC, McMichael JF, Kiwala S, et al. Standard operating procedure for curation and clinical interpretation of variants in cancer. Genome Medicine. 2019 Nov 29;11(1):76.

28. Duma N, Durani U, Woods CB, Kankeu Fonkoua LA, Cook JM, Wee C, et al. Evaluating Unconscious Bias: Speaker Introductions at an International Oncology Conference. J Clin Oncol. 2019 Dec 20;37(36):3538–45.

29. Dizman N, Agarwal R, Castro DV, Mercier B, Li X, Barragan-Carrillo R, et al. Assessment of Compliance With ASCO Language of Respect Guidelines in Renal Cell Carcinoma Abstracts. JCO Oncol Pract. 2025 Apr;0(0):OP-24-01039.

30. Siu LL, Lawler M, Haussler D, Knoppers BM, Lewin J, Vis DJ, et al. Facilitating a culture of responsible and effective sharing of cancer genome data. Nat Med. 2016 May;22(5):464–71.

31. Wagner AH, Babb L, Alterovitz G, Baudis M, Brush M, Cameron DL, et al. The GA4GH Variation Representation Specification: A computational framework for variation representation and federated identification. Cell Genomics. 2021 Nov 10;1(2):100027.

32. Rehm HL, Page AJH, Smith L, Adams JB, Alterovitz G, Babb LJ, et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. Cell Genomics. 2021 Nov 10;1(2):100029.

33. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Research. 2018 Jan 4;46(D1):D1062–7.

34. Rubin AF, Stone J, Bianchi AH, Capodanno BJ, Da EY, Dias M, et al. MaveDB 2024: a curated community database with over seven million variant effects from multiplexed functional assays. Genome Biology. 2025 Jan 21;26(1):13.

35. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3(1):160018.

36. Barker M, Chue Hong NP, Katz DS, Lamprecht AL, Martinez-Ortiz C, Psomopoulos F, et al. Introducing the FAIR Principles for research software. Sci Data. 2022 Oct 14;9(1):622.

37. Fragoso G, de Coronado S, Haber M, Hartel F, Wright L. Overview and utilization of the NCI Thesaurus. Comparative and Functional Genomics. 2004;5(8):648–54.

38. Liu S, Wei Ma, Moore R, Ganesan V, Nelson S. RxNorm: prescription for electronic drug information exchange. IT Prof. 2005 Sept;7(5):17–23.

39. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006 Jan 1;34(suppl_1):D668–72.

40. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. Human Mutation. 2016;37(6):564–9.

41. Holmes JB, Moyer E, Phan L, Maglott D, Kattman B. SPDI: data model for variants and applications at NCBI. Bioinformatics. 2020 Mar 15;36(6):1902–7.

42. Kundra R, Zhang H, Sheridan R, Sirintrapun SJ, Wang A, Ochoa A, et al. OncoTree: A Cancer Classification System for Precision Oncology. JCO Clinical Cancer Informatics. 2021 Dec;(5):221–30.

43. Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, et al. Genenames.org: the HGNC resources in 2023. Nucleic Acids Research. 2023 Jan 6;51(D1):D1003–9.

44. The UniProt Consortium, Bateman A, Martin MJ, Orchard S, Magrane M, Adesina A, et al. UniProt: the Universal Protein Knowledgebase in 2025. Nucleic Acids Research. 2025 Jan 6;53(D1):D609–17.

45. Belenkaya R, Gurley MJ, Golozar A, Dymshyts D, Miller RT, Williams AE, et al. Extending the OMOP Common Data Model and Standardized Vocabularies to Support Observational Cancer Research. JCO Clinical Cancer Informatics. 2021 Dec;5:12–20.

46. Hart RK, Fokkema IFAC, DiStefano M, Hastings R, Laros JFJ, Taylor R, et al. HGVS Nomenclature 2024: improvements to community engagement, usability, and computability. Genome Medicine. 2024 Dec 20;16(1):149.

47. Ajmal A, Bouissou O, Brash J, Cheeseman S, Banduge PG, Gomes AL, et al. Establishing standards: harmonising coding principles for a minimal cancer dataset in the OMOP Common Data Model. ESMO Real World Data and Digital Oncology. 2025 Sept;9:100179.

48. Morales J, Pujar S, Loveland JE, Astashyn A, Bennett R, Berry A, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. Nature. 2022 Apr;604(7905):310–5.

49. Wright CF, FitzPatrick DR, Ware JS, Rehm HL, Firth HV. Importance of adopting standardized MANE transcripts in clinical reporting. Genetics in Medicine. 2023 Feb 1;25(2):100331.

50. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. J Mol Diagn. 2017 Jan;19(1):4–23.

51. Mateo J, Chakravarty D, Dienstmann R, Jezdic S, Gonzalez-Perez A, Lopez-Bigas N, et al. A framework to rank genomic alterations as targets for cancer precision medicine: the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). Annals of Oncology. 2018 Sept;29(9):1895–902.

52. Reisle C, Grisdale CJ, Krysiak K, Danos AM, Khanfar M, Pleasance E, et al. Evaluating Language Models for Biomedical Fact-Checking: A Benchmark Dataset for Cancer Variant Interpretation Verification [Internet]. bioRxiv; 2025 [cited 2025 Oct 7]. p. 2025.09.10.675443. Available from: https://www.biorxiv.org/content/10.1101/2025.09.10.675443v2

[H1] Display items

[H2] Figure 1

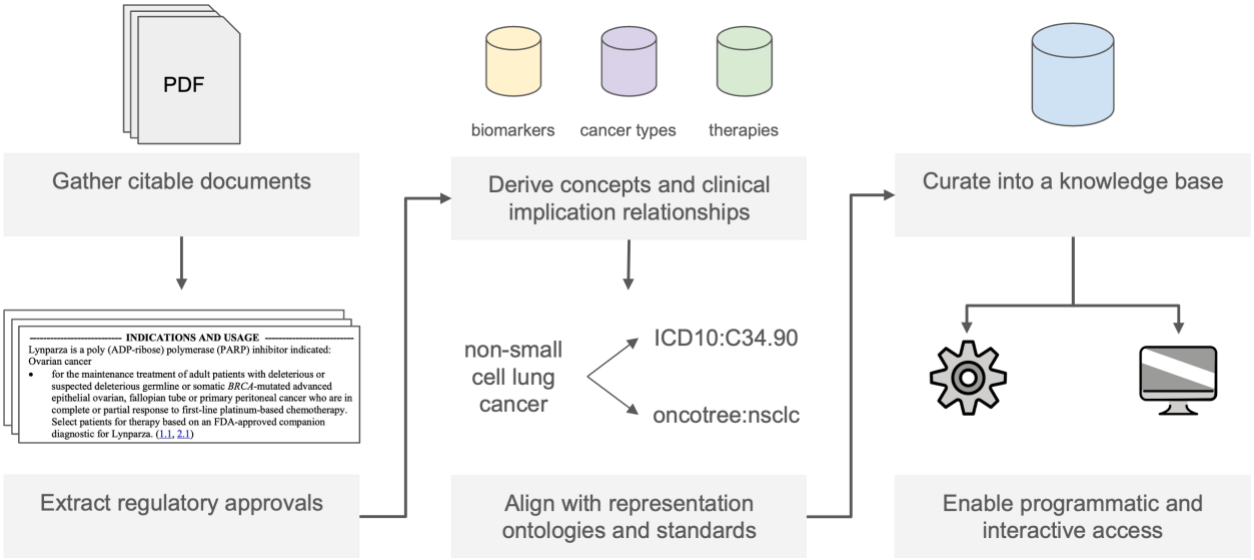

Fig. 1: A standard operating procedure for making region-specific precision oncology approvals programmatically accessible. Once relevant documents from an agency are identified, approvals of interest from each can be extracted. For example, precision oncology

knowledge bases usually select cancer drugs with market authorization that have at least one approved indication involving a biomarker. For each approval, derive genomic knowledge statements by associating the biomarker, cancer type, and therapies involved and align each of these with representation ontologies and standards. These statements can then be compiled into a knowledge base, and made programmatically and interactively accessible.

[H2] Figure 2

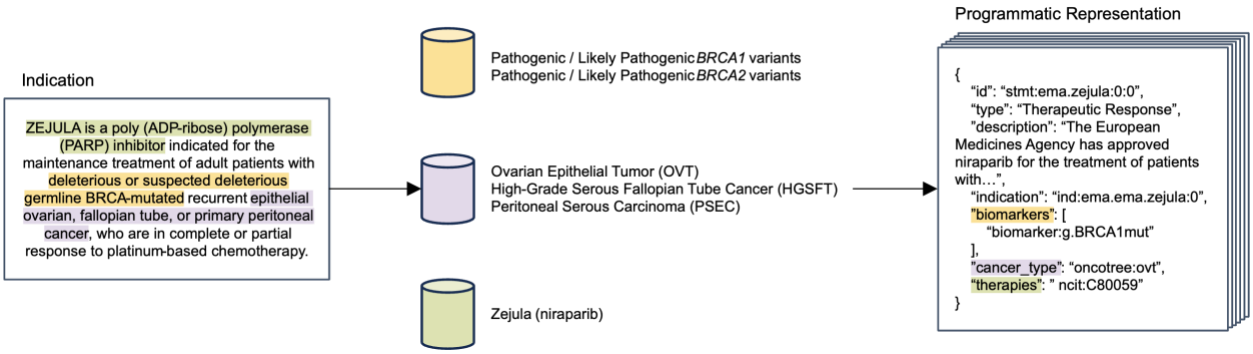

Fig. 2: A single regulatory approval can derive to multiple precision oncology knowledge statements. An approved indication of Zejula (niraparib) involves two biomarkers (pathogenic or likely pathogenic germline *BRCA1/2* variants), three cancer types (ovarian, fallopian tube, and peritoneal cancers), and one therapy (niraparib itself). This results in six separate biomarker, cancer type, and therapy relationships, each of which are derived to their own genomic knowledge statement.

[H2] Table 1

Example statuses of curated regulatory approvals.

Approval status	**Description**
Accelerated / Conditional	The approval has received conditional authorization based on positive, albeit preliminary or incomplete, endpoints, often in underserved disease contexts.
Approved	The treatment can be administered to patients based for the approved indication (‘on label’ usage). This may be a market authorization (approval for sale) or an authorization of use within a health system.
Superseded	The treatment indication’s text has been revised to a new version, often broadening or narrowing of eligibility criteria. For example, removing biomarker requirement(s). The earlier indication text remains historically valid but no longer current.
Withdrawn	The treatment indication is no longer authorized. This may be voluntarily on behalf of the sponsor (for example, due to commercial sales) or involuntarily due to new evidence.

[H2] Table 2

Example treatment indication and derived knowledge statement

Description	Text
Original indication text	Verzenio is a kinase inhibitor indicated in combination with an aromatase inhibitor as initial endocrine-based therapy for the treatment of adult patients with hormone receptor (HR)-positive, human epidermal growth factor receptor 2 (HER2)-negative advanced or metastatic breast cancer.
Description for knowledge base use	The U.S. Food and Drug Administration (FDA) granted approval to abemaciclib in combination with an aromatase inhibitor as an initial endocrine-based therapy for the treatment of adult patients with hormone receptor (HR)-positive, human epidermal growth factor 2 (HER2)-negative advanced or metastatic breast cancer.
Description with additional clarifying details	The U.S. Food and Drug Administration (FDA) granted approval to abemaciclib in combination with an aromatase inhibitor as an initial endocrine-based therapy for the treatment of adult patients with hormone receptor (HR)-positive, human epidermal growth factor 2 (HER2)-negative advanced or metastatic breast cancer. This indication is based on the MONARCH 3 (NCT02246621) clinical trial, which was a randomized (2:1), double-blinded, placebo-controlled, multicenter study. A total of 493 patients were randomized to receive 150 mg abemaciclib or placebo orally twice daily, plus physician's choice of letrozole (80% of patients) or anastrozole (20%).

[H2] Table 3

Illustrative ontologies and standards to represent concepts, non-exhaustive

Concept	Ontology or representation standard
Cancer types	Experimental Factor Ontology, National Cancer Institute (NCI) Thesaurus, OncoTree, Observational Medical Outcomes Partnership (OMOP)
Genetic variation	Human Genome Variation Society Nomenclature, OMOP Genomic, SPDI, Variation Representation Specification
Gene symbols	Hugo Gene Nomenclature Committee, National Center for Biotechnology Information (NCBI)
Protein markers	UniProt, NCBI
Therapies	DrugBank, NCI Thesaurus, OMOP, RxNorm

[H1] Supplementary information

Supplementary table 1

Title: Curating underlying documents for cancer drug approvals

Caption: An example curation of 35 documents for cancer drugs approvals by the European Medicine Agencies. Documents were labeled as “relevant” or “not relevant” based on if the therapy was an oncology therapy and at least one approved indication involved a biomarker. Citations were generated and additional chosen metadata was recorded for each relevant document.

Supplementary table 2

Title: Curating approved indications for cancer drug approvals
Caption: An example curation of 20 approved treatment indications from the US FDA. Approved indications were recorded verbatim from the originating document. The approval text was then rephrased for knowledge base use and chosen metadata of approval date and current approval status. Precision oncology concepts of biomarkers, cancer types, and therapies were also extracted as written from the approval text.