International Journal of Creative and Open Research in Engineering and Management
ISSN: 3108-1754 (Online)
Volume 02 Issue 09 September-2026 | Impact Factor: 3.5

CancerGeneHub: An Evidence-Linked Bidirectional Web Portal for Centralised Exploration of Cancer–Gene Associations

Gulnaaz Parveen[1], Alam Zia[2], Shams Alam[2], Ayushi Sharma[3]*

[1] Department of Biotechnology, Hindustan College of Science and Technology

[1,2,3]Uttar Pradesh, India

Corresponding Author Email: ayushi.sharma515@gmail.com

How to Cite this Article:

Parveen, G., Zia, A., Alam, S. & Sharma, A. (2026). CancerGeneHub: An Evidence-Linked Bidirectional Web Portal for Centralised Exploration of Cancer–Gene Associations. International Journal of Creative and Open Research in Engineering and Management, 02(9), 1-9. https://doi.org/10.55041/ijcope.v2i9.127

https://doi.org/10.55041/ijcope.v2i9.127

Abstract—

Cancer is a major global health challenge characterised by genetic and molecular alterations that disrupt normal cellular growth, regulation, and survival. The rapid expansion of cancer genomics has generated extensive information concerning cancer-associated genes; however, this information remains distributed across scientific publications and specialised biological databases, making efficient retrieval difficult, particularly for students and early-stage researchers. This study presents CancerGeneHub, a web-based Cancer Gene Information Portal developed to provide centralised and evidence-linked exploration of gene–cancer associations. The portal organises 8,077 gene–cancer associations involving 3,177 genes and 130 cancer types distributed across 11 body-system classifications. A bidirectional search mechanism allows users to investigate genes associated with a selected cancer type and, conversely, cancer types associated with a selected gene. Each association is linked to a PubMed Identifier, enabling users to trace the reported relationship to supporting scientific literature. The system was implemented using React and Tailwind CSS for the frontend, Express for backend services, and MongoDB for data storage and management. The resulting platform combines structured biological information, searchable many-to-many relationships, and literature traceability within a single web interface. The developed system provides an accessible resource for cancer-genetics education and exploratory research. It establishes a scalable foundation for future integration of genomic databases, evidence scoring, mutation information, functional annotations, and interactive analytical capabilities.

Keywords— Cancer genomics; cancer-associated genes; gene–cancer association; bioinformatics; PubMed; biomedical information retrieval.

I. INTRODUCTION

Cancer is a highly complex disease that continues to impose a substantial burden on global health and biomedical research. Its development is associated with the gradual accumulation of genetic, epigenetic, and molecular alterations that disturb the normal mechanisms responsible for cellular growth, differentiation, survival, and tissue homeostasis [1], [2]. Alterations affecting genes involved in cell-cycle regulation, DNA damage response, apoptosis, signal transduction, and genome maintenance can provide cells with a selective advantage and contribute to malignant transformation [1], [3]. Advances in next-generation sequencing and large-scale genomic analysis have substantially improved the ability of researchers to characterize these alterations and investigate their contribution to tumor initiation and progression. Comprehensive analyses of

cancer genomes have demonstrated that tumors contain a combination of driver alterations that contribute to cancer development and passenger alterations that accumulate during tumor evolution without providing a direct growth advantage [3], [4]. In addition, the genetic landscape is not uniform across cancer types; different tumors may exhibit distinct molecular profiles while sharing alterations in important biological pathways [4], [5]. Large-scale initiatives such as The Cancer Genome Atlas (TCGA) have consequently generated extensive molecular information across multiple tumor types, providing valuable resources for investigating genomic variation, molecular pathways, and potential cancer-associated genes [5]. Similarly, the Catalogue of Somatic Mutations in Cancer (COSMIC) provides curated information on somatic mutations identified in human cancers, while the Cancer Gene Census provides systematically curated information about genes with established roles in cancer development [6], [7]. The availability of these resources has significantly expanded the knowledge base of cancer genetics; however, it has also created a highly distributed information environment in which relevant biological information may be located across different databases, publications, and specialized repositories. NCBI resources, including PubMed and other biological databases, further provide extensive access to scientific literature and molecular information [8], [9]. Although these resources are valuable individually, obtaining a complete view of a particular gene–cancer relationship may require users to perform multiple searches and interpret information presented in different formats. This challenge is particularly relevant for students, early-career researchers, and users who require a straightforward method for exploring cancer-associated genetic information without navigating several specialized platforms.

The fragmentation of cancer-genetics information also becomes more apparent when the relationship between genes and cancer types is considered as a many-to-many association. A single gene may be implicated in several cancer types, whereas a single cancer type may involve numerous genes with different biological functions and levels of evidence [3], [7]. Consequently, restricting information retrieval to only a cancer-centered or gene-centered perspective can make it difficult to understand the broader relationship between these biological entities. Existing integrated resources such as DisGeNET and Open Targets demonstrate the value of connecting genes, diseases, targets, and supporting evidence within structured knowledge systems [16]–[18]. However, these platforms serve broader biomedical and drug-discovery purposes and may not provide the focused, simplified interaction required for direct exploration of cancer-to-gene and gene-to-cancer relationships by general academic users. To address this gap, the present study introduces CancerGeneHub, a centralized web-based Cancer Gene Information Portal designed specifically for the organized exploration of cancer–gene associations. The proposed portal enables users to approach the information from either direction: a cancer type can be selected to identify its associated genes, or a gene can be searched to identify the cancer types with which it has been reported to be associated. In addition, each association is linked with a PubMed Identifier (PMID), allowing users to move from a structured database record to the corresponding scientific literature and independently examine the supporting publication [8], [9]. The principal objectives of the study are therefore to organize gene–cancer relationships within a structured and searchable dataset, implement bidirectional cancer-to-gene and gene-to-cancer retrieval, provide literature traceability through PMID integration, classify cancer types according to body-system categories, and establish an accessible and extensible platform for cancer-genetics exploration. The resulting system is intended to complement established biomedical databases rather than replace them by providing a focused information-access layer that simplifies the initial exploration of gene–cancer relationships while retaining a connection to the underlying scientific evidence.

II. LITERATURE REVIEW

A. Cancer Genomics and Cancer-Associated Genes

Cancer genomics has fundamentally changed the understanding of cancer by demonstrating that malignant transformation is closely associated with the accumulation of genetic and molecular

alterations that affect important cellular processes. The hallmarks of cancer framework proposed by Hanahan and Weinberg provided a biological basis for understanding how alterations in cellular signaling, proliferation, apoptosis, genome stability, and other processes contribute to tumor development [1], [2]. Subsequent genomic studies have shown that cancer is not driven by a single molecular event but by combinations of alterations that may vary according to tumor type, tissue of origin, disease stage, and individual patient. Stratton *et al.* emphasized the importance of systematically characterizing cancer genomes to identify recurrent genetic alterations and understand their contribution to tumor biology [3]. Vogelstein *et al.* further demonstrated that cancer genomes contain both driver and passenger alterations, with driver alterations providing selective advantages that contribute to tumor development and progression [4]. This distinction is important for cancer-genetics research because not every mutation observed in a tumor has the same biological significance.

Large-scale cancer-genomics initiatives have subsequently generated increasingly comprehensive molecular datasets. The International Cancer Genome Consortium established an international framework for systematic cancer genome analysis and promoted coordinated investigation of genomic alterations across cancer types [10]. Similarly, The Cancer Genome Atlas (TCGA) Pan-Cancer initiative integrated molecular information across tumor lineages and examined genomic, transcriptomic, epigenomic, and proteomic characteristics to identify common and cancer-specific molecular patterns [5]. The TCGA project demonstrated the value of integrating information across tumor types rather than examining each cancer independently, thereby providing a broader understanding of shared molecular mechanisms and differences between cancers [5]. Such large-scale projects have substantially increased the volume of publicly available cancer information, but the resulting datasets are complex and are primarily designed for detailed genomic analysis rather than simple relationship-based information retrieval. The growing volume and diversity of cancer genomic information therefore creates a parallel requirement for systems capable of organizing these relationships in a form that can be readily explored by researchers, students, and other academic users.

B. Cancer Genomic Databases and Biomedical Information Resources

The expansion of cancer genomics has been accompanied by the development of specialized databases designed to collect, curate, and distribute genomic information. Among these, the Catalogue of Somatic Mutations in Cancer (COSMIC) is a major resource for investigating somatic alterations in human cancer. COSMIC integrates information from scientific publications and large-scale cancer studies and provides curated information concerning coding and non-coding mutations, gene fusions, copy-number alterations, and other mechanisms through which genomic changes can contribute to cancer [6]. The breadth of COSMIC illustrates both the value and complexity of modern cancer-genomics resources. Its Cancer Gene Census (CGC) provides an additional layer of expert curation by identifying genes with established roles in cancer development and describing their functional contribution to oncogenesis [7]. The CGC has consequently become an important reference for researchers studying cancer-driving genes and genetic dysfunction [7].

The National Center for Biotechnology Information (NCBI) provides another major component of the biomedical information infrastructure. NCBI maintains a collection of resources including PubMed, Gene, GenBank, RefSeq, GEO, and other databases that support access to biological records and scientific literature [8], [9]. PubMed is particularly relevant to the present study because scientific publications represent an important source of evidence for reported gene–cancer relationships. However, PubMed is fundamentally a literature retrieval system rather than a dedicated gene–cancer association database. Consequently, users may need to interpret individual publications and manually establish relationships between genes, diseases, and evidence.

Additional resources provide complementary biological information. The Human Protein Atlas offers tissue-, cell-, and protein-level information that can help researchers understand gene and

protein expression in normal and disease contexts [11]. UniProt provides curated information concerning proteins, including sequence, function, and biological annotations [12]. Gene Ontology provides a standardized vocabulary for describing molecular functions, biological processes, and cellular components, thereby supporting consistent functional interpretation of genes [13], [14]. ClinVar, in contrast, focuses on relationships between human genetic variants and clinically relevant interpretations [15]. Collectively, these resources demonstrate that cancer-genetics research often requires information from several complementary sources rather than a single database.

The diversity of these resources is valuable from a scientific perspective, but it also introduces an information-integration challenge. Different databases may use different identifiers, disease classifications, terminology, evidence models, and interfaces. A researcher interested in a particular gene–cancer relationship may therefore need to move between mutation databases, literature repositories, gene-function databases, and disease-association resources. The availability of information does not necessarily guarantee its efficient accessibility, particularly when the research question is relatively simple and relationship-oriented.

C. Integrated Gene–Disease and Evidence-Based Resources

The limitations associated with fragmented biomedical information have encouraged the development of integrated knowledge platforms. DisGeNET is an important example because it was specifically developed to integrate and standardize information concerning genes, genomic variants, and human diseases from multiple sources [16]. The platform combines expert-curated repositories, genomic resources, and scientific literature and provides standardized annotations and prioritization measures for gene–disease relationships [16]. Earlier work on DisGeNET explicitly identified fragmentation, heterogeneity, and differences in the conceptualization of biomedical data as challenges that limit effective reuse of disease-associated genomic information [21]. These observations are particularly relevant to the present study because they demonstrate that simply making large amounts of biomedical data available does not eliminate the need for effective organization and retrieval mechanisms.

Open Targets provides another example of evidence-oriented integration. The platform was developed to support systematic identification and prioritization of therapeutic targets by connecting targets with diseases and integrating multiple forms of biological and genetic evidence [17], [18]. Such platforms demonstrate the value of connecting biological entities to supporting evidence rather than presenting associations as isolated records. However, their primary objective is broader biomedical discovery and therapeutic target prioritization rather than straightforward exploration of cancer-to-gene and gene-to-cancer relationships.

The evolution of these platforms demonstrates an important trend in biomedical informatics: the movement from isolated databases toward integrated knowledge systems. Nevertheless, comprehensive platforms can also introduce complexity because users must navigate multiple data dimensions, evidence categories, filters, and specialized terminology. For users whose primary requirement is to understand a specific gene–cancer relationship, a more focused interface may provide a simpler entry point into the available scientific information.

D. Literature-Based Evidence and Traceability

Scientific literature remains an important component of cancer-genetics knowledge because many biological associations originate from experimental studies, clinical investigations, sequencing projects, and systematic analyses reported in peer-reviewed publications. COSMIC, for example, relies extensively on literature curation to characterize somatic alterations and cancer-associated genes [6]. Similarly, integrated resources such as DisGeNET combine curated databases with information extracted from the scientific literature [16], [21]. These approaches emphasize that the provenance of a biological association is important when evaluating or interpreting biomedical information.

For a user exploring a gene–cancer relationship, the ability to identify the underlying publication can therefore provide an important level of

transparency. A structured record stating that a particular gene is associated with a particular cancer provides useful information, but access to the associated publication allows the user to investigate the context in which the relationship was reported. This may include the cancer subtype, experimental system, mutation or alteration involved, study population, and type of evidence. Consequently, linking structured associations with PubMed identifiers can serve as a practical bridge between database-level information and the underlying scientific literature.

The proposed CancerGeneHub follows this evidence-traceability principle by retaining PMID information with the corresponding gene–cancer association wherever available. The purpose is not to independently assess or experimentally validate the conclusions of each publication, but to provide users with a direct route to the scientific source associated with a database record.

E. Comparative Analysis of Existing Resources

The reviewed resources differ substantially in their intended purpose and information structure. TCGA primarily supports large-scale molecular characterization of tumors [5], whereas COSMIC focuses extensively on somatic mutations and cancer-associated genomic alterations [6]. The Cancer Gene Census provides expert-curated information about cancer-driving genes [7]. NCBI and PubMed provide broad access to biological records and scientific literature [8], [9]. UniProt and Gene Ontology provide complementary protein and functional information [12]–[14], while ClinVar emphasizes clinically relevant variant interpretation [15]. DisGeNET integrates gene–disease and variant–disease relationships across multiple sources [16], and Open Targets connects biological targets with diseases and evidence for therapeutic research [17], [18].

This diversity demonstrates that the biomedical information ecosystem is rich but highly specialized. Each resource solves a particular information problem, and researchers frequently obtain the most complete understanding by combining information from several sources. However, this approach can be inefficient when the immediate research question is simply to identify the genes associated with a cancer or the cancers associated with a gene. The challenge is therefore not the absence of information, but the accessibility and organization of information for a focused relationship-level query.

F. Research Gap

The literature indicates that extensive information concerning cancer-associated genes, genomic alterations, diseases, and supporting scientific evidence is already available through established resources [5]–[18]. The principal limitation is therefore not a lack of biomedical data but the fragmentation of information across resources with different purposes, data models, terminology, and interfaces. Large genomic resources such as TCGA and COSMIC provide substantial biological depth, while platforms such as DisGeNET and Open Targets demonstrate the benefits of integrating relationships and evidence [5], [6], [16]–[18]. Nevertheless, users investigating a specific gene–cancer relationship may still need to move between several resources to obtain a consolidated understanding of the association and locate the relevant supporting publication.

The many-to-many nature of gene–cancer relationships further emphasizes this problem. A single gene may be associated with several cancer types, while a single cancer may involve numerous genes with different biological roles and levels of evidence [3], [4], [7]. Therefore, a system that supports only one direction of retrieval may provide an incomplete view of the relationship. In practical terms, a user asking “Which genes are associated with this cancer?” and another asking “Which cancers are associated with this gene?” are exploring two directions of the same underlying biological relationship. Providing these two perspectives through a common interface can reduce unnecessary information switching and make exploratory research more efficient.

Based on these observations, CancerGeneHub was developed as a focused information-access platform rather than as a replacement for established genomic databases. The proposed system combines structured gene–cancer association data with bidirectional search, centralized retrieval, body-system classification, and PMID-based evidence traceability. The portal allows users to begin with either a cancer type or a

gene and retrieve the corresponding associations through a single interface. At the same time, the inclusion of PMID information provides a pathway from the structured association to the underlying scientific literature. In this respect, the proposed work builds upon the integration principles demonstrated by resources such as COSMIC, DisGeNET, and Open Targets [6], [16]–[18], while narrowing the interaction model to the specific task of accessible cancer–gene exploration. The contribution of CancerGeneHub therefore lies primarily in simplifying access to an existing body of biomedical knowledge, presenting gene–cancer relationships from both directions, and preserving traceability to supporting literature.

III. METHODOLOGY

A. Research and System Design

The study adopted a structured software-development and biomedical information-integration approach for the design and implementation of the Cancer Gene Information Portal. The methodology was organized as a sequential yet interconnected process beginning with the collection and organization of gene–cancer association data, followed by data cleaning and normalization to improve consistency across biological entities. Gene names, cancer types, body-system classifications, and PubMed identifiers were systematically mapped to establish meaningful relationships between the records and their supporting scientific literature.

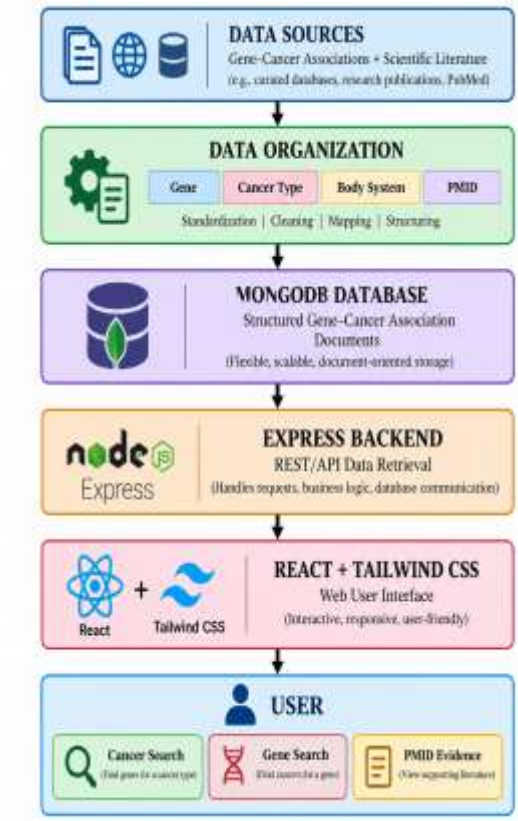

Figure 1. Overall architecture of the Cancer Gene Information Portal

The prepared data were then structured for database implementation using MongoDB, while the application layer was developed using Express and application programming interface (API) services to manage data requests and communication between the database and user interface. The frontend was implemented using React and Tailwind CSS to provide an interactive, responsive, and user-friendly environment for information retrieval. A key component of the methodology was the implementation of bidirectional search functionality, enabling users to retrieve genes associated with a selected cancer type as well as cancer types associated with a selected gene. Finally, functional validation was performed to verify the major search, retrieval, result-display, and PMID-access operations of the system. Through the integration of these stages, the methodology transformed structured biomedical association data into a centralized web-based information platform while preserving a traceable connection between individual gene–cancer records and their associated scientific evidence. The overall system architecture and interaction among the data, database, backend, frontend, and user layers are illustrated in Figure 1.

B. Dataset

The current dataset comprises **8,077 gene–cancer associations**, representing **3,177 unique genes** and **130 cancer types** organized into **11 body-system classifications**. The dataset was structured to support both cancer-centered and gene-centered information retrieval, allowing the same underlying information to be explored from complementary perspectives. The principal data entities include the **gene, cancer type, body-system classification, PubMed Identifier (PMID), and gene–cancer association**. The relationship between genes and cancer types is inherently many-to-many, since a single gene may be associated with multiple cancer types, while a particular cancer type may involve multiple genes. This relationship can be represented as $A = \{(G_i, C_j)\}$, where $G_i$denotes the i-th gene, $C_j$denotes the j-th cancer type, and Arepresents the collection of documented gene–cancer associations. To improve evidence traceability, an association can additionally be connected to a corresponding scientific publication through its PMID, represented as $(G_i, C_j) \rightarrow PMID_k$, where $PMID_k$identifies the publication associated with the respective gene–cancer relationship. This data representation enables the portal to retrieve associations efficiently while maintaining the connection between structured biological information and the scientific literature supporting the reported relationship.

C. Data Organization and Normalization

The data were systematically organized according to **gene names, cancer types, body-system classifications, gene–cancer relationships, and PubMed identifiers (PMIDs)** to establish a consistent structure for storage and retrieval within the web application.

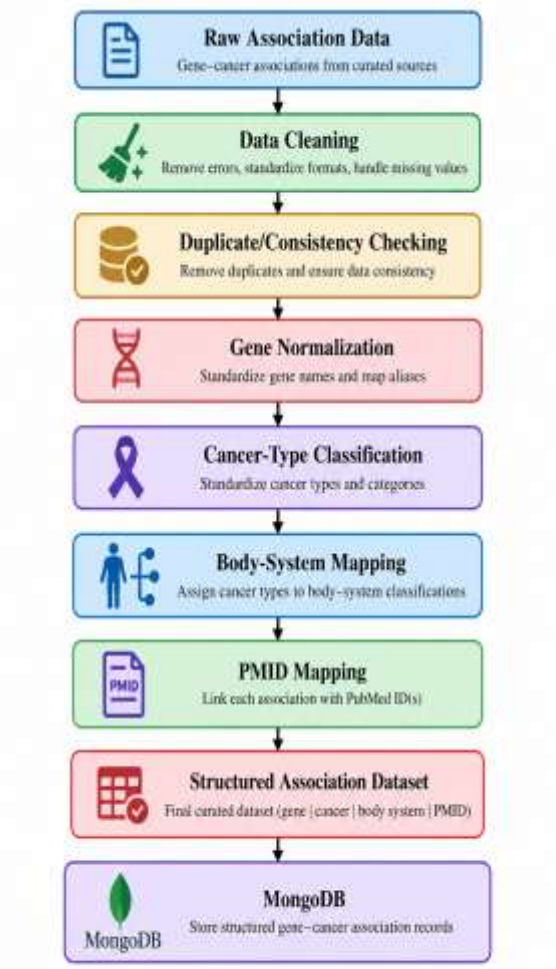

Figure 2. Data preparation and database-construction workflow

During this stage, gene and cancer names were identified and standardized to maintain consistency across individual records, while each cancer type was mapped to its corresponding body-system category to support organized classification and future filtering. The gene–cancer relationships were then preserved as individual association records so that multiple genes could be linked to the same cancer type and, conversely, a single gene could be associated with multiple cancer types. Where supporting literature was available, the corresponding PMID was retained with the association to maintain a traceable connection between the structured record and its scientific source. Following these organization and mapping steps, the processed records were prepared in a structured format suitable for database storage and subsequent retrieval through the application programming interface (API). The complete data-organization and preprocessing workflow is illustrated in **Figure 2**.

D. Web Application Development

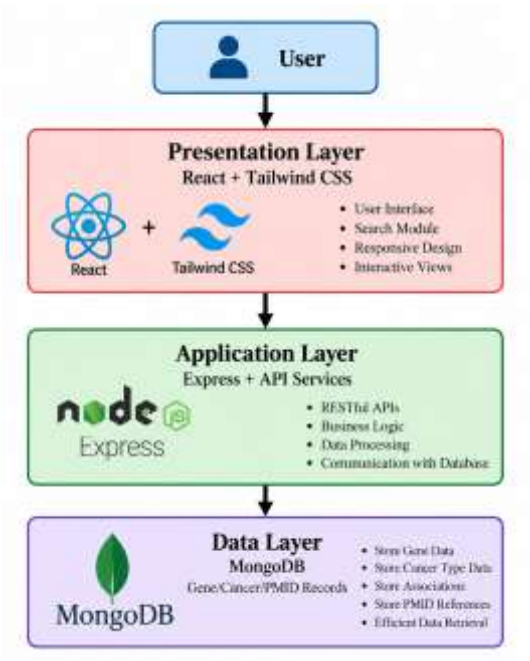

Figure 3. Three-layer block diagram of the CancerGeneHub implementation

The frontend was developed using React. React provides a component-based approach for developing interactive web interfaces [19]. Tailwind CSS was used for interface styling and responsive layout. Express was used to implement backend services and provide communication between the frontend and database.

MongoDB was selected for data storage because of its document-oriented architecture and flexible schema model [20]. The three principal system layers are shown in **Figure 3**.

E. Bidirectional Search

1) Cancer-to-Gene Search

For a selected cancer C_j, the system retrieves the set of genes associated with that cancer:

$$Genes(C_j) = \{G_i \mid (G_i, C_j) \in A\}$$

The retrieved records are then displayed to the user. The cancer-to-gene workflow is shown in **Figure 4**.

Figure 4. Cancer-to-gene search workflow

2) Gene-to-Cancer Search

For a selected gene G_i, the system retrieves the associated cancer types:

$$Cancers(G_i) = \{C_j \mid (G_i, C_j) \in A\}$$

The workflow is illustrated in **Figure 5**.

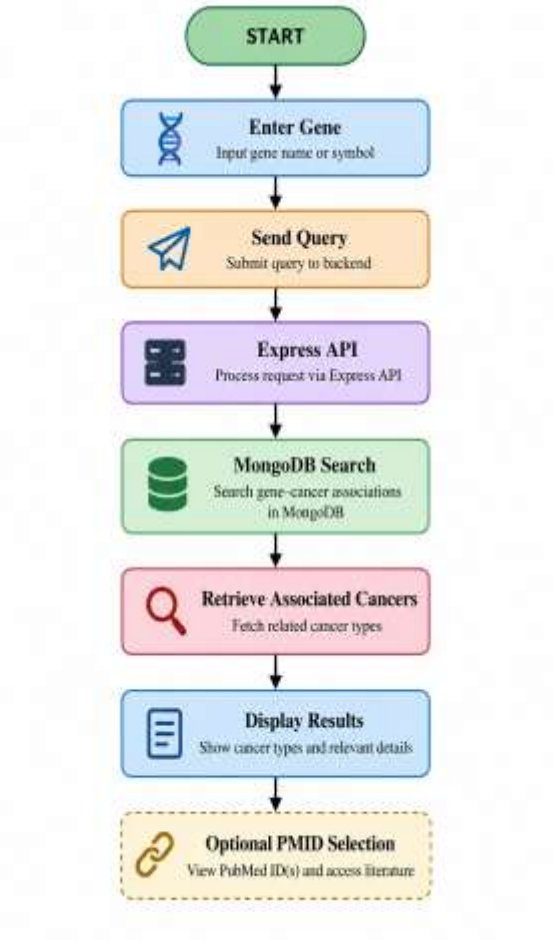

Figure 5. Gene-to-cancer search workflow

F. PubMed Evidence Integration

A PubMed Identifier (PMID) was retained with the corresponding gene–cancer association wherever supporting literature was available. This evidence-linking approach enables users to move directly from a structured gene–cancer record to the relevant biomedical publication, thereby improving the traceability and transparency of the information presented in the portal. By preserving the connection between an association and its source publication, users can independently examine the underlying scientific evidence and obtain additional context beyond the summarized database record. The overall evidence-linking process is illustrated in **Figure 6**.

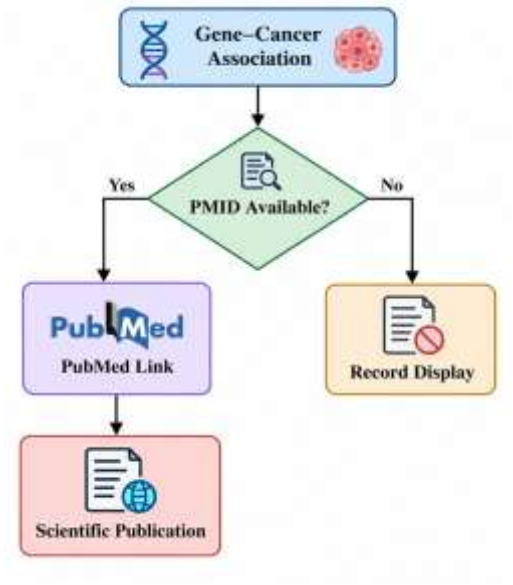

Figure 6. PMID-based evidence traceability workflow

G. User Interface

The user interaction process consists of selecting a search mode, entering a cancer or gene query, retrieving the relevant association records, and optionally examining the associated PMID.

The complete user workflow is shown in **Figure 7**.

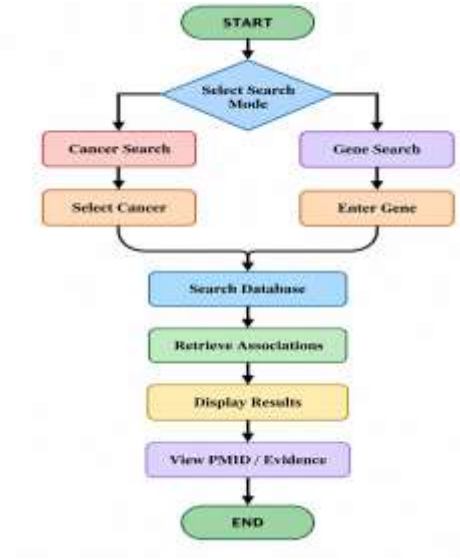

Figure 7. End-to-end user interaction flowchart

H. Functional Validation

Functional validation focused on the major user operations of the portal.

Table I: Functional Validation Test Cases

Test Case	Expected Result
Valid cancer search	Associated genes displayed
Valid gene search	Associated cancers displayed
Cancer with multiple associations	Multiple genes returned
Gene with multiple associations	Multiple cancers returned
PMID selection	Supporting literature accessible
Invalid query	Appropriate no-result response
Repeated query	Consistent result retrieval

This validation approach evaluates the information-retrieval functionality rather than experimentally validating the underlying biological associations.

I. Ethical Considerations

The portal is intended as an educational and research-oriented information resource. It does not independently establish clinical diagnoses, therapeutic recommendations, or patient-specific medical decisions.

The PMID-based evidence mechanism is therefore intended to support literature exploration rather than replace professional biomedical or clinical interpretation.

IV. RESULTS AND DISCUSSION

A. Dataset Summary

The developed portal currently contains 8,077 gene–cancer associations involving 3,177 genes and 130 cancer types across 11 body-system classifications.

Table II: Summary Of The Cancergenehub Dataset

Dataset Component	Number
Gene–cancer associations	8,077
Unique genes	3,177
Cancer types	130
Body systems	11

The dataset provides sufficient breadth to demonstrate many-to-many relationships between genes and cancer types.

B. Portal Homepage

The Cancer Gene Information Portal is publicly deployed as a web application at:

The homepage provides the primary entry point for cancer and gene searches. The actual homepage screenshot should be inserted as **Figure 8**.

Figure 8. Homepage of the Cancer Gene Information Portal

The homepage is intended to reduce the number of steps required for users to initiate cancer-genetics information retrieval.

C. Cancer-to-Gene Search Results

The cancer-to-gene functionality allows a user to begin with a cancer type and retrieve the associated genes stored within the dataset.

Figure 9 should present the actual cancer-to-gene search interface and corresponding results.

Figure 9. Cancer-to-gene search results displayed by the portal

This functionality provides a direct answer to a cancer-centered query and can be particularly useful for educational exploration and preliminary research.

D. Gene-to-Cancer Search Results

The reverse search functionality begins with a gene and retrieves cancer types associated with that gene.

Figure 10. Gene-to-cancer search results displayed by the portal

The two search directions provide complementary perspectives of the same underlying many-to-many relationship.

E. PMID-Based Evidence Display

The PMID integration provides users with a pathway from a database association to scientific literature.

Figure 11. PMID-based evidence display for a selected gene–cancer association

This approach improves transparency because users can independently inspect the publication associated with an individual relationship.

However, the presence of a PMID should not be interpreted as independent experimental validation by the portal. Rather, it provides a traceable route to the literature supporting the reported association.

F. Comparison With Existing Resources

The comparison with established biomedical resources indicates that CancerGeneHub has a more focused scope than comprehensive genomic and biomedical databases, which typically provide broader information covering mutations, gene functions, disease associations, clinical variants, molecular profiles, or therapeutic targets [5]–[18]. Rather than attempting to replicate the extensive functionality of these specialized resources, CancerGeneHub concentrates on simplifying the exploration of gene–cancer relationships through a unified interface. Its principal contribution lies in combining bidirectional search, structured association retrieval, and PMID-based literature traceability within a single platform. This focused design allows users to begin their investigation from either a cancer type or a gene and subsequently access the corresponding associations and supporting literature without having to perform separate searches across multiple resources. Thus, CancerGeneHub is positioned as a complementary information-access platform that improves the accessibility and initial exploration of cancer–gene knowledge while retaining links to established scientific evidence**.**

Table III: Comparison With Selected Biomedical Resources

Resource	Main Focus	Gene–Cancer Information	Literature/Evidence	Bidirectional Search
PubMed	Biomedical literature	Indirect	Yes	No
TCGA	Cancer genomics	Yes	Yes	Limited
COSMIC	Somatic mutations	Yes	Yes	Yes
Cancer Gene Census	Cancer genes	Yes	Yes	Limited
DisGeNET	Gene–disease associations	Yes	Yes	Yes
Open Targets	Target–disease evidence	Yes	Yes	Yes
CancerGeneHub	Gene–cancer exploration	Yes	PMID-linked	Yes

G. Discussion

The results demonstrate that the proposed system can organize a relatively large collection of biological relationships into a searchable web interface. The first major contribution is **centralization**. Existing resources such as COSMIC, TCGA, NCBI, ClinVar, UniProt, DisGeNET, and Open Targets each address different biological information requirements [5]–[18]. CancerGeneHub provides a focused environment for exploring one specific relationship: the association between genes and cancer types. The second contribution is **bidirectional retrieval**. Because cancer–gene relationships are many-to-many, supporting both directions is important. A cancer-centered researcher can identify candidate genes, while a

gene-centered researcher can examine associated cancers.

The third contribution is **evidence traceability**. Linking records to PMIDs provides an important transparency mechanism. Rather than requiring users to independently search PubMed after retrieving a gene–cancer relationship, the portal provides a direct route to the associated publication. The fourth contribution is **accessibility**. Specialized genomic databases can contain extensive information but may require familiarity with complex terminology and database structures. CancerGeneHub emphasizes straightforward search and presentation, making it potentially useful for students and early-stage researchers.

The architecture also provides a foundation for future expansion. Additional resources could be incorporated into the same system while maintaining the core gene–cancer relationship model.

H. Limitations

The current implementation has several limitations.

First, the coverage of the portal depends on the underlying dataset. The 8,077 associations represent the current dataset and should not be interpreted as an exhaustive representation of all known cancer–gene relationships.

Second, the portal primarily provides association retrieval rather than biological interpretation. It does not currently provide comprehensive mutation frequencies, gene-expression profiles, pathway analysis, survival analysis, or treatment-response prediction.

Third, the system links associations to PubMed identifiers but does not independently evaluate the methodological quality or biological strength of each publication.

Fourth, cancer terminology and classification can vary among databases. Differences in cancer names, synonyms, subtypes, and hierarchical classifications may create interoperability challenges.

Finally, the current implementation does not include automated evidence scoring or continuous database synchronization.

I. Future Scope

Future development can include integration with COSMIC, ClinVar, TCGA-derived information, UniProt, Gene Ontology, DisGeNET, and Open Targets [5]–[18].

An evidence-scoring framework could be introduced to categorize associations according to publication count, evidence type, database agreement, experimental support, and recency.

The portal could also incorporate gene-function information, protein annotations, pathways, mutation-level information, genomic coordinates, and clinical variant interpretations.

Interactive network visualization could further represent the relationship:

$$Gene \leftrightarrow Cancer \leftrightarrow Evidence$$

Such visualization could help users identify highly connected genes, cancer clusters, and evidence relationships.

Automated data-update pipelines could also be developed to periodically retrieve new associations, normalize terminology, remove duplicates, and update the database.

V. CONCLUSION

This study presented CancerGeneHub, a web-based Cancer Gene Information Portal developed to provide centralized and accessible exploration of gene–cancer associations.

The current implementation contains 8,077 gene–cancer associations involving 3,177 genes and 130 cancer types organized across 11 body-system classifications. The system combines a structured biological dataset with a React-based user interface, Express backend services, MongoDB storage, and PMID-linked literature access.

The most important functional contribution is the ability to explore associations in both directions. Users can begin with a cancer type and identify associated genes, or begin with a gene and identify

associated cancer types. The PMID integration further provides a mechanism for tracing database records to scientific publications.

The proposed system should be regarded as an information-access and educational research resource rather than a clinical decision-support system or an independent biological validation framework.

Future development involving additional genomic databases, evidence scoring, mutation information, functional annotation, interactive network visualization, and automated data updates could substantially increase the scientific utility of the platform.

The study demonstrates that focused biomedical information systems can simplify exploration of complex many-to-many biological relationships while maintaining links to underlying scientific evidence.

REFERENCES

[1] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: The next generation," *Cell*, vol. 144, no. 5, pp. 646–674, 2011, doi: 10.1016/j.cell.2011.02.013.

[2] D. Hanahan, "Hallmarks of cancer: New dimensions," *Cancer Discovery*, vol. 12, no. 1, pp. 31–46, 2022, doi: 10.1158/2159-8290.CD-21-1059.

[3] M. R. Stratton, P. J. Campbell, and P. A. Futreal, "The cancer genome," *Nature*, vol. 458, no. 7239, pp. 719–724, 2009, doi: 10.1038/nature07943.

[4] B. Vogelstein *et al.*, "Cancer genome landscapes," *Science*, vol. 339, no. 6127, pp. 1546–1558, 2013, doi: 10.1126/science.1235122.

[5] The Cancer Genome Atlas Research Network, "The Cancer Genome Atlas Pan-Cancer analysis project," *Nature Genetics*, vol. 45, no. 10, pp. 1113–1120, 2013, doi: 10.1038/ng.2764.

[6] J. G. Tate *et al.*, "COSMIC: The Catalogue Of Somatic Mutations In Cancer," *Nucleic Acids Research*, vol. 47, no. D1, pp. D941–D947, 2019, doi: 10.1093/nar/gky1015.

[7] Z. Sondka *et al.*, "The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers," *Nature Reviews Cancer*, vol. 18, no. 11, pp. 696–705, 2018, doi: 10.1038/s41568-018-0060-1.

[8] NCBI Resource Coordinators, "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Research*, vol. 46, no. D1, pp. D8–D13, 2018, doi: 10.1093/nar/gkx1095.

[9] NCBI Resource Coordinators, "Database resources of the National Center for Biotechnology Information," *Nucleic Acids Research*, vol. 49, no. D1, pp. D10–D17, 2021, doi: 10.1093/nar/gkaa892.

[10] T. J. Hudson *et al.*, "International network of cancer genome projects," *Nature*, vol. 464, no. 7291, pp. 993–998, 2010, doi: 10.1038/nature08987.

[11] M. Uhlén *et al.*, "Proteomics. Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, 2015, doi: 10.1126/science.1260419.

[12] UniProt Consortium, "UniProt: the Universal Protein Knowledgebase in 2023," *Nucleic Acids Research*, vol. 51, no. D1, pp. D523–D531, 2023, doi: 10.1093/nar/gkac1052.

[13] M. Ashburner *et al.*, "Gene ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, pp. 25–29, 2000, doi: 10.1038/75556.

[14] The Gene Ontology Consortium, "The Gene Ontology resource: enriching a GOld mine," *Nucleic Acids Research*, vol. 49, no. D1, pp. D325–D334, 2021, doi: 10.1093/nar/gkaa1113.

[15] M. J. Landrum *et al.*, "ClinVar: improving access to variant interpretations and supporting evidence," *Nucleic Acids Research*, vol. 46, no. D1, pp. D1062–D1067, 2018, doi: 10.1093/nar/gkx1153.

[16] J. Piñero *et al.*, "DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants," *Nucleic Acids Research*, vol. 47, no. D1, pp. D845–D855, 2019, doi: 10.1093/nar/gky1111.

[17] G. Koscielny *et al.*, "Open Targets: a platform for therapeutic target identification and validation,"

Nucleic Acids Research, vol. 45, no. D1, pp. D985–D994, 2017, doi: 10.1093/nar/gkw1055.

[18] K. Ochoa *et al.*, "Open Targets Platform: supporting systematic drug–target identification and prioritisation," *Nucleic Acids Research*, vol. 49, no. D1, pp. D1302–D1310, 2021, doi: 10.1093/nar/gkaa1027.

[19] Meta Open Source, "React: A JavaScript library for building user interfaces," Meta Open Source, 2026. [Online]. Available: https://react.dev/

[20] MongoDB Inc., "MongoDB documentation," MongoDB Documentation, 2026. [Online]. Available: https://www.mongodb.com/docs/

[21] P. A. Futreal *et al.*, "A census of human cancer genes," *Nature Reviews Cancer*, vol. 4, no. 3, pp. 177–183, 2004.

[22] S. Bailey *et al.*, "Comprehensive characterization of cancer driver genes and mutations," *Cell*, vol. 173, no. 2, pp. 371–385.e18, 2018, doi: 10.1016/j.cell.2018.02.060.

[23] L. B. Alexandrov *et al.*, "The repertoire of mutational signatures in human cancer," *Nature*, vol. 578, pp. 94–101, 2020, doi: 10.1038/s41586-020-1943-3.

[24] L. B. Alexandrov *et al.*, "Signatures of mutational processes in human cancer," *Nature*, vol. 500, pp. 415–421, 2013, doi: 10.1038/nature12477.

[25] C. Kandoth *et al.*, "Mutational landscape and significance across 12 major cancer types," *Nature*, vol. 502, pp. 333–339, 2013, doi: 10.1038/nature12634.

[26] L. A. Lawrence *et al.*, "Mutational heterogeneity in cancer and the search for new cancer-associated genes," *Nature*, vol. 499, pp. 214–218, 2013, doi: 10.1038/nature12213.

[27] N. Cancer Genome Atlas Research Network, "Comprehensive molecular portraits of human breast tumours," *Nature*, vol. 490, pp. 61–70, 2012, doi: 10.1038/nature11412.

[28] J. N. Weinstein *et al.*, "The Cancer Genome Atlas Pan-Cancer analysis project," *Nature Genetics*, vol. 45, no. 10, pp. 1113–1120, 2013.

[29] M. Gerstberger, H. Jiang, and K. Ganesh, "Metastasis," *Cell*, vol. 186, no. 8, pp. 1564–1579, 2023, doi: 10.1016/j.cell.2023.03.003.

[30] C. S. Davis *et al.*, "The Cancer Genome Atlas: a resource for cancer research and biomedical discovery," *Nature Reviews Cancer*, vol. 24, 2024.