JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 28, Number 12, 2021

Pp. 1248–1257
DOI: 10.1089/cmb.2021.0348

ProCanBio: A Database of Manually Curated Biomarkers for Prostate Cancer

DIKSCHA SAPRA,[1] HARPREET KAUR,[2,i] ANJALI DHALL,[1,ii] and GAJENDRA P.S. RAGHAVA[1,iii]

ABSTRACT

Prostate cancer (PCa) is the second lethal malignancy in men worldwide. In the past, numerous research groups investigated the omics profiles of patients and scrutinized biomarkers for the diagnosis and prognosis of PCa. However, information related to the biomarkers is widely scattered across numerous resources in complex textual format, which poses hindrance to understand the tumorigenesis of this malignancy and scrutinization of robust signature. To create a comprehensive resource, we collected all the relevant literature on PCa biomarkers from the PubMed. We scrutinize the extensive information about each biomarker from a total of 412 unique research articles. Each entry of the database incorporates PubMed ID, biomarker name, biomarker type, biomolecule, source, subjects, validation status, and performance measures such as sensitivity, specificity, and hazard ratio (HR). In this study, we present ProCanBio, a manually curated database that maintains detailed data on 2053 entries of potential PCa biomarkers obtained from 412 publications in user-friendly tabular format. Among them are 766 protein-based, 507 RNA-based, 157 genomic mutations, 260 miRNA-based, and 122 metabolites-based biomarkers. To explore the information in the resource, a web-based interactive platform was developed with searching and browsing facilities. To the best of the authors' knowledge, there is no resource that can consolidate the information contained in all the published literature. Besides this, ProCanBio is freely available and is compatible with most web browsers and devices. Eventually, we anticipate this resource will be highly useful for the research community involved in the area of prostate malignancy.

Keywords: biomarkers, database, prostate cancer, resource, signatures.

[1]Department of Computational Biology, Indraprastha Institute of Information Technology Delhi, New Delhi, India.
[2]Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh, India.
[i]ORCID ID (https://orcid.org/0000-0003-0421-8341).
[ii]ORCID ID (https://orcid.org/0000-0002-0400-2084).
[iii]ORCID ID (https://orcid.org/0000-0002-8902-2876).
An earlier version of this article was posted as a preprint at bioRxiv (DOI: 10.1101/2021.06.06.447247).

1. INTRODUCTION

According to the International Agency for Research on Cancer, prostate cancer (PCa) is the second most prevalent type of cancer in men, accounting for ~13.5% of the total cancer in 2018 (Rawla, 2019). PCa accounts for 1.4 million cases all over the world and ~375,000 deaths (6.3% of deaths in males) in 2018 alone (Sung et al., 2021). It is the sixth leading cause of cancer-related death worldwide (Jemal et al., 2010; Sung et al., 2021). PCa usually occurs at older age, that is, 65 years and above (Droz et al., 2010). The aggressiveness of the disease is measured by Gleason Score (GS) (Verma et al., 2011). GS was introduced in 1966 by Donald F. Gleason. GS is a score between 2 and 10 calculated on the primary and the secondary core of the tumor. The higher the score, the more aggressive is the cancer (Mellinger et al., 1967). Some other noncancerous prostate conditions include benign prostatic hyperplasia (BPH) that is one of the most common prostate-related diseases and is associated with low urinary tract syndrome. It is not a life-threatening disease, but in extreme cases, it can lead to renal failure. Prostatic intraepithelial neoplasia (PIN) is known to be the most common precursor to PCa. It is further divided into high-grade (HGPIN) and low-grade (LGPIN) disease. HGPIN patients are said to be at high risk for PCa (Brawer, 2005).

The most common test for the diagnosis and prognosis of PCa is measurement of prostate specific antigen (PSA) (Atan and Guzel, 2013). PSA is present both in normal and in malignant prostate tissue; however, the range of PSA differs in both cases (Adhyam and Gupta, 2012). PSA is very sensitive in detecting PCa but lacks specificity. Elevated levels of PSA are also associated with benign prostate conditions such as BPH, PIN, and prostatitis that lead to poor specificity (Nadler et al., 1995). Hence more complex forms of PSA such as tPSA (total PSA), fPSA (free PSA), and %fPSA have been used for diagnosis. Although these markers enhance diagnosis and prognosis capabilities to a certain extent, however, there is still a need to identify robust biomarkers and drug targets to further improve the prognosis of PCa patients. The need for new biomarkers was also driven by the need to differentiate PCa from BPH, PIN, and prostatitis. The most common treatment for PCa is androgen deprivation therapy (Huggins and Hodges, 2002), especially in cases of recurrent and advanced PCa (Moul et al., 2011).Over the years there have been a number of studies performed to understand the pathogenesis of this malignancy and to elucidate signatures for PCa (Hutchinson et al., 2005; Hessels et al., 2007; Pretorius et al., 2009; Kosari et al., 2012).

This enormous information generated related to PCa biomarkers including genomics, epigenomics, proteomics, metabolomics, and peptidomics. (Sallam, 2015). But even with the increase in the biomarkers in the literature, it is extremely difficult to analyze this information since it is available in unstructured textual format across diverse platforms. This information is widely scattered and there is no repository to store all this information together in a user-friendly manner. The absence of structured data format calls the need for ProCanBio. In the past, various databases and prediction tools have been developed for maintaining biomarkers for different types of malignancies such as cervical cancer, liver cancer, breast, colorectal cancer, and skin cancer. (Agarwal et al., 2011; Butti et al., 2014; Nalejska et al., 2014; Bhalla et al., 2017; Zhang et al., 2018; Bhalla et al., 2019; Kaur et al., 2019a,b; Perez-Granado et al., 2019; Chu et al., 2020; Dhall et al., 2020; Dingerdissen et al., 2020; Kaur et al., 2020; Terkelsen et al., 2020; Zuo et al., 2020; Kaur et al., 2021). To the best of the author's knowledge, there is no such dedicated database available for maintaining the signatures or biomarkers for PCa. Although Cancer Proteomics Database includes information related to biomarkers for PCa, it only includes proteomics data and was last updated in 2013 (Arntzen et al., 2015). The Early Detection Research Network (ERDN) provides a list of biomarkers for different types of cancers (including PCa), but does not give detailed information about the biomarker, and is not annotated according to individual literature publications.

To fill the gap from previous studies, in this study, we developed ProCanBio (https://webs.iiitd.edu.in/raghava/procanbio/) that provides manually curated detailed information from published research articles that related to various biomarkers of PCa. This is a freely accessible database to help researches analyze biomarkers for further use by the scientific community.

2. MATERIALS AND METHODS

2.1. Data collection

To find all relevant literature related to PCa biomarkers, two keyword searches were performed on PubMed: ''(prostate cancer[title/abstract]) AND biomarkers[title/abstract])'' and ''(prostate cancer[title/abstract]) AND signatures [title/abstract])'' that yielded 3623 and 422 relevant publications, respectively

(as analyzed on May 10, 2019). From a total of 4045 publications, 112 were common from both the keyword searches and they were removed. A total of 3933 publications were then further analyzed to extract information. After carefully reviewing all articles, reviews, and publications that are not available in English language and irrelevant to desired topic were excluded from the study. A total of 412 were left, which were used to extract information about biomarkers. However, the true positive rate (TPR) and false positive rate (FPR) for the relevant research articles is 10.47% and 89.52%, respectively, as calculated by the given equations.

$$\mathrm{TPR} = {}^{\mathrm{TP}}/_{TP + FN}, \tag{1}$$

$$\mathrm{FPR} = {}^{\mathrm{FP}}/_{\mathrm{FP} + \mathrm{TN}}, \tag{2}$$

where true positive (TP) = 412, false positive (FP) = 3933, and false negative (FN) = 3521.

Extensive information about each biomarker has been extracted from each publication including PubMed ID; technical name; biomarker name; biomarker basis; biomolecule; source; subjects; regulation status in cancerous conditions; type of biomarker; cohort used in the study; effect on pathway; experimental conditions; and performance metrics such as sensitivity, specificity, accuracy, ROC-AUC (receiver operating characteristics-area under the curve), hazard ratio (HR), odds ratio (OR), and relative risk (RR); level of significance (p value); degree of validity; clinical trials; and the methods used for analysis. Pathway information associated with the biomarker was obtained from the Enrichr (Kuleshov et al., 2016). Each biomarker entry is linked to its original PubMed article from where the biomarker was taken. The biomarker ID is linked to the GeneCards (version 4.14) (Stelzer et al., 2016). If the biomarker has been assessed in a clinical trial and is registered with the ClinicalTrials (https://clinicaltrials.gov), it is linked with its National Clinical Trial (NCT) number or any other clinical trial.

2.2. Web interface

ProCanBio developed using APACHE HTTP server that is freely available. The backend is maintained using MySQL (v8) as RDBMS. Front end is developed using PHP(v7), HTML(v5), Javascript (v1.8), and CSS (v3). The server was developed on a Linux machine.

3. RESULTS

3.1. Database architecture

The architecture and overall organization for ProCanBio are as described in Figure 1.

3.2. Biomarkers from the literature

There are a total of 2053 entries related to biomarkers extracted from 412 publications out of which 1497 are unique biomarkers. The description of each field/column is as follows: ''PubMed ID'' is linked to the original publication from which the information of biomarker was extracted; ''Year'' represents the year during which the study was published; ''Biomarker'' tells the name of the biomarker/or a group of biomarker that is studied; and ''Biomarker Basis'' gives how the Biomarker was measured—expression, concentration, methylation, or mutation. ''Technical Name'' refers to the actual name of the biomarker derived from GeneCard.org.; ''Biomolecule'' provides the bimolecular basis of the biomarker—RNA, DNA, protein, miRNA, metabolite, etc.; and ''Source'' tells from which part of the body this biomarker was extracted—cell lines, tissue, blood, plasma, serum, urine, bone marrow, or semen.

''Regulation status in Cancerous Conditions'' gives information whether a particular biomarker was observed to be upregulated or downregulated in PCa (or the mentioned experimental conditions) along with fold change (difference in the two experimental conditions, if given in original study); ''Subjects'' tells whether the experiment was performed on humans, mice, or rats. ''Odds Ratio/Hazard Ratio/Relative Risk'' tells the OR, hazard ratio (HR), or relative ratio (RR) between the two mentioned experimental conditions. ''Effect on Pathways'' provides information about the different pathways in which the given biomarker is involved. The pathways information was extracted from the Enrichr (Top 5 pathways sorted by adjusted p-values) (Kuleshov et al., 2016) or GeneCard (pathways with the highest Jaccard Index) or associated publication (Stelzer et al., 2016). ''Experiment'' refers to the different conditions under which

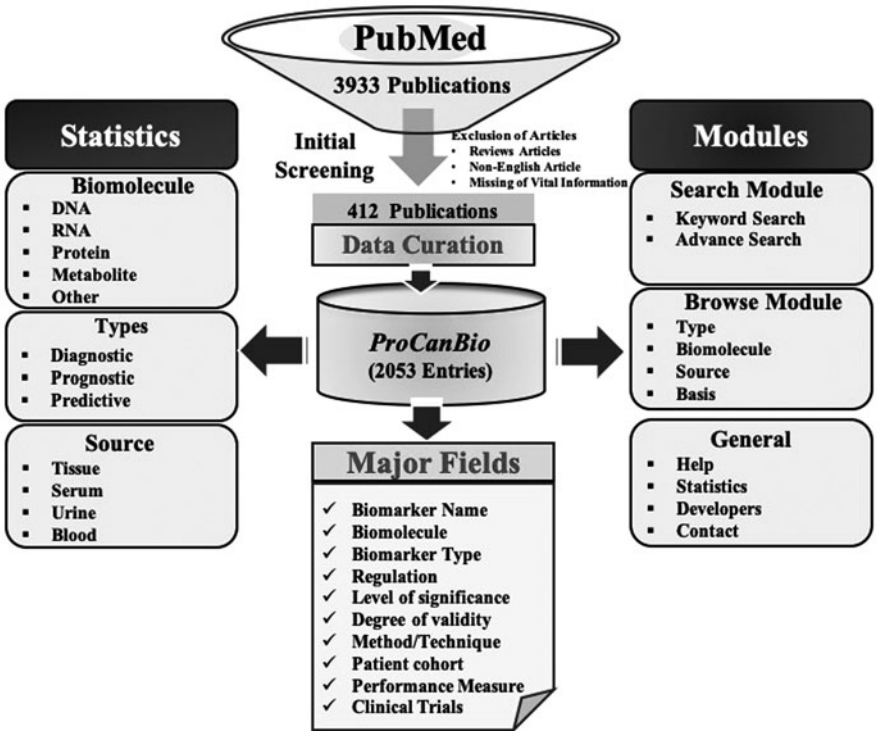

FIG. 1. Architecture of ProCanBio.

the biomarker was evaluated and ''Type of Biomarker'' indicates the utility of given biomarker, that is, diagnostic, prognostic, or predictive. Diagnostic biomarkers handle cases wherein biomarkers are used to segregate PCa from healthy controls, BPH, PIN, and prostatitis. Prognostic biomarkers are used to identify biomarkers that distinguish PCa stage, GS, metastasis, miochemical recurrence, overall and disease-specific survival. Predictive biomarkers include biomarkers that give information about the effects of therapeutic interventions, particularly, biomarkers that were differentially expressed after a certain therapy.

The most common therapies from the database include Docetaxel therapy. ''Cohort'' gives a description of the patient cohort (population) that was chosen for the study. ''sen,'' ''spec,'' ''AUC,'' and ''accuracy'' refer to the sensitivity, specificity, ROC-AUC, and accuracy performance of the biomarkers., respectively, in the given experimental conditions. Level of significance tells the p-value of the experiment. $p \leq 0.05$ was considered as significant. ''Method Used'' provides us the information regarding techniques that were used to perform the experiment in the reported study for biomarker discovery or evaluation such as immunohistochemistry, quantitative polymerase chain reaction (qPCR), real-time reverse transcription-polymerase chain reaction (RT-PCR), fluorescence in situ hybridization, mass spectrometry, enzyme-linked immunosorbent assay (ELISA), Western blot, and matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF). ''Clinical'' informs whether the study was a part of a clinical trial, and ''Clinical Trial Number'' gives the clinical trial ID for the same—either through NCT or other global trial registrations. ''Remarks'' include any additional remarks that are to be made about the biomarker card. ''Degree of Validity'' indicates whether the mentioned biomarker was validated on human patient cohort (in case the experiment is performed on cell lines) and it also provides information whether it was validated on an independent data set or not in the associated study.

In this case, if the signature set or biomarkers consist of multiple genes/proteins/miRNA, then individual entities (gene/miRNA/protein) are separated from each other via comma (‘‘,’’), semi-colon (‘‘;’’), plus-sign (‘‘+’’) within the biomarkers; whereas comma or semi-colon indicates that the biomarkers were evaluated individually and plus-sign represents that the biomarkers constitute multiple entities (genes/proteins/miRNA, etc).

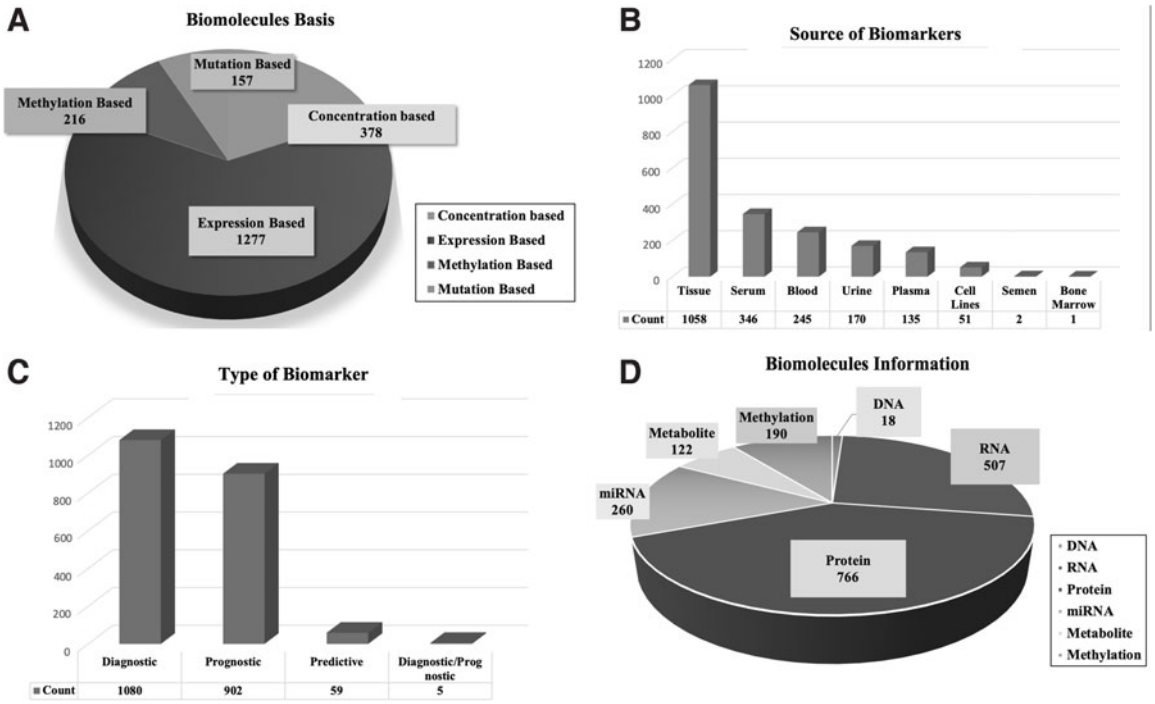

FIG. 2. Statistical representation of entries in PanCanBio: **(A)** biomolecule basis, **(B)** source of biomarker, **(C)** type of biomarker, and **(D)** biomolecules information.

3.3. Statistics

In the database, stratifying on the basis of the expression of Biomarker-1277, 157 biomarkers were mutation based, 216 biomarkers were methylation based, and 378 biomarkers were concentration based (Fig. 2A). Based on the source, 1058 biomarkers were extracted from tissues, 356 from serum, 245 from blood, 170 from urine, 135 from plasma, 51 from cell lines, 2 from semen, and 1 from bone marrow (Fig. 2B). Furthermore, based on the experimental subjects, 2034 biomarkers were extracted from humans, 13 from mice, and 5 from rats. On the basis of biomolecules, 766 were protein biomarkers, 18 were DNA markers, 508 were RNAs, 260 were miRNAs, and 122 were metabolites (Fig. 4D). Out of the total biomarker studies, 327 used RT-PCR, 220 used immunohistochemistry, and 97 used ELISA. The most common biomarker was PCA3 that has 24 entries in the database; Ki-67 appeared 12 times as represented in Table 1.

Table 1. List of Top 10 Biomarker/Signature of Prostate Cancer with the Information Such as Type of Biomolecule and Biomarker, Number of Entries, and Studies Related to the Biomarker

		No. of entries			
Biomarker	*Biomolecule*	*Diagnostic*	*Prognostic*	*Predictive*	*No. of studies*
PCA3	mRNA/protein	22	2	0	17
Ki-67	Protein	4	7	1	9
Prostate Health index (phi)	mRNA/protein	11	0	0	9
%p2PSA	Protein	3	6	0	5
miR-141	miRNA	3	5	0	8
PSA	mRNA/protein	4	1	2	5
PTEN	mRNA/protein/DNA	0	7	0	4
miR-205	miRNA	3	4	0	3
VEGF	Protein	2	4	1	5
Methylation status of GSTP1	mRNA/methylation	1	6	0	5

PSA, prostate specific antigen; PTEN, phosphatase and tensin homolog; VEGF, vascular endothelial growth factor.

3.4. Querying the database

The ProCanBio has the facilities to retrieve the data using many different search options such as searching and browsing: including keyword search, complex search, type of biomarker, type of biomolecule, source, and basis of biomarker.

3.5. Search tools

These tools take an input from user and perform search queries on the database that are in agreement with the keyword(s) entered by the user.

3.6. Keyword search

It allows to search the database using a single keyword. It will retrieve all information related to the keyword search in the database. Fields can be selected according to name of biomarker, biomolecule, type of biomarker, subjects, PMID, and regulation conditions. The search is not case sensitive, so it takes queries in all cases of the user. Users can choose which fields to display by ticking the provided check boxes.

3.7. Complex search

It allows users to access the database when ProCanBio queries with more than one keyword. The complex query can be searched on biomolecules, source of biomarker, type of biomarkers, subjects, and year of publication. For numeric type operator (year), field option can be set to =, <,> to retrieve queries. The rest of the operators (excluding year) are string type operators and can be queried using ''like'' field option. Condition field contains all permissible values for the operator field. Thus, rows can be added and deleted to be more complex queries. The query field allows to join different individual queries by ''AND,'' ''NOT,'' or ''OR'' operations. For example, if one has to search all entries from database such that we have to find proteins with prognostic capabilities for PCa, we can do so by entering ''Biomolecule'' in the first row operator, with LIKE field, and ''protein'' condition then adds new row, and selects ''Type of Biomolecule'' operator, with ''Like'' field and ''prognostic'' in the condition field along with ''and'' in the query field (since we want our result to satisfy both these conditions).

3.8. Browse tools

These tools help to query the database on certain main keywords in an orderly manner. The user can query using four main tools—(1) type of biomarker, (2) biomolecule, (3) source, and (4) biomarker basis. On clicking on all these options, user is redirected to a page that will provide different categories for all the mentioned four tools. User can select any one of them to see all related entries for the keyword in the database.

While viewing query results for both search and browse, user can select how many entries to be shown on one page, and allows search within the queried data set. When a keyword is entered, if any of the rows contain that keyword in the row, then that row would be retained. The results can be downloaded in the form of Excel (.xls), comma separated values (.csv), or a PDF file. To view the detailed information regarding the biomarker, ID column should be selected that provides a biomarker card that contains all the information regarding that particular entry in the database.

In the general tab, user can view help dropdown that has detailed instructions on how to use the server. The statistics dropdown gives statistics about the various biomarkers in ProCanBio. Developers and contact dropdown show information about the developers of ProCanBio.

3.9. Working of ProCanBio database

3.9.1. Case Study I: prostate specific membrane antigen as biomarker. ProCanBio can facilitate the user for easy query and browse the database without having to go through hundreds of entries to find the relevant information. Here we illustrate the results for query against prostate specific membrane antigen (PSMA). After performing keyword search against ''PSMA,'' we find there are 15 unique entries in ProCanBio from six research publications as shown in Figure 3. This information tells us it has been used for diagnostic purposes (mostly against BPH). It has been extracted from a variety of samples including tissue, serum, and urine. It has been validated only once on independent data sets. We can observe that

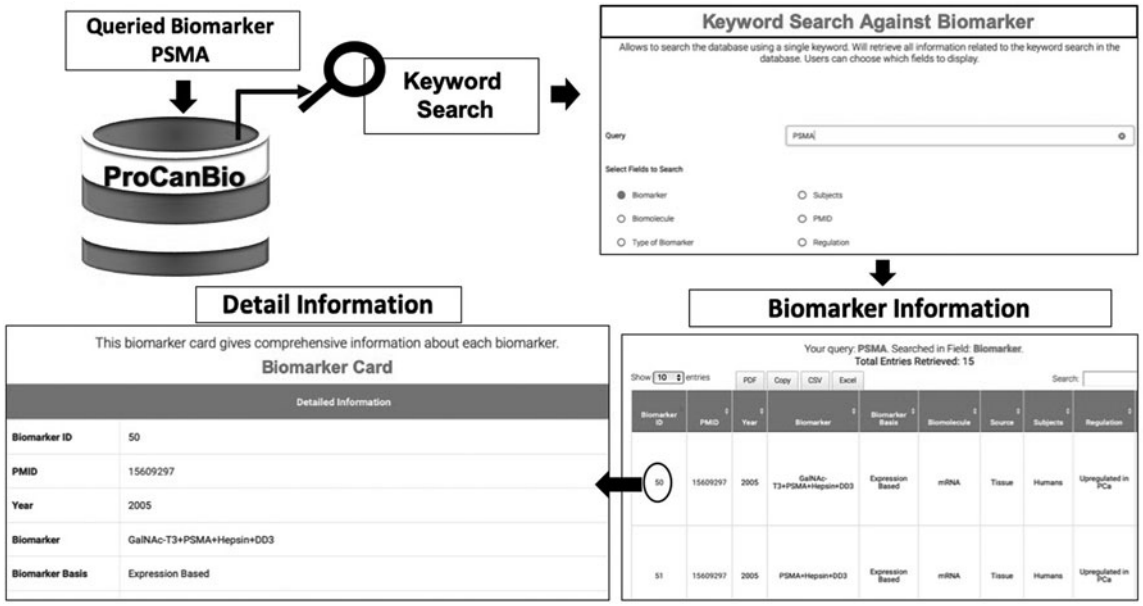

FIG. 3. Demonstration of querying and retrieval of information from ProCanBio using keyword search module.

PSMA mean expression is higher in PCa than in controls. Upon clicking the biomarker ID, we can see detailed information for PSMA including OR, performance, special remarks and degree of validity, and clinical trial number (none in this case, since PSMA is not studied in any clinical trials).

3.9.2. Case Study II: published prognostic protein biomarkers after 2015. To perform complex queries, we have a case study to show all prognostic protein biomarkers published after the year 2015. By performing this query, we retrieve 23 unique entries. Out of these 23 entries, 5 focus on biochemical recurrence and 7 focus on overall survival. These were extracted from either tissues or serum of patients. Some of the important biomarkers that are retrieved include Prohibitin, *WISP1*, *Ki-67*, *GLUT1*, and *AZGP1*. The query performed to find these results is as shown in Figure 4. Similar to these case studies, more queries can be performed to answer similar questions related to PCa biomarkers.

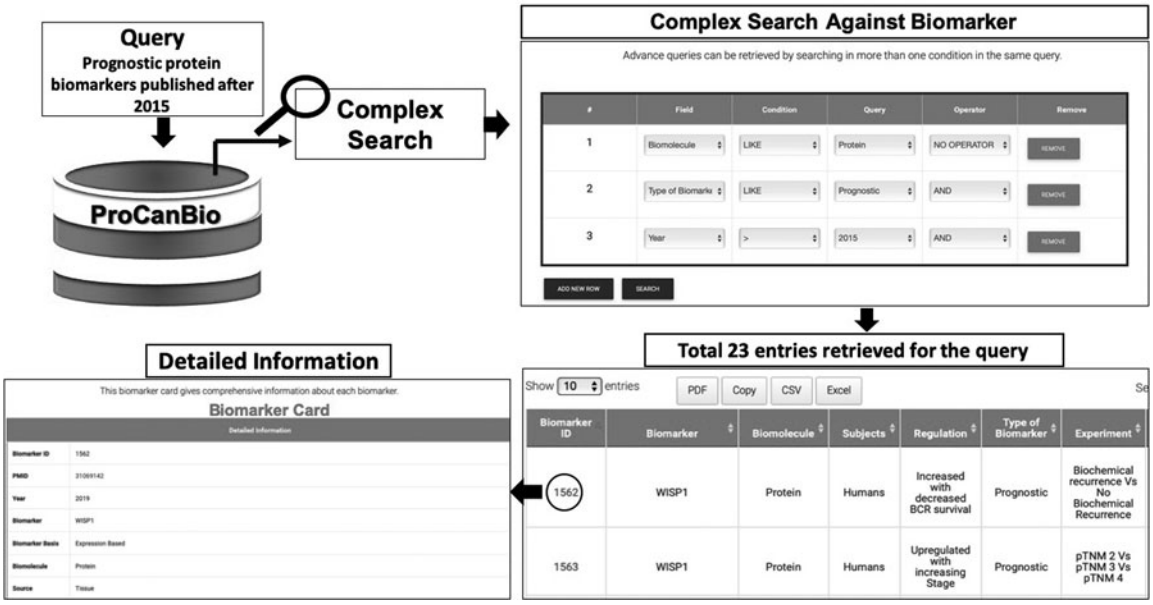

FIG. 4. Complex query search for prognostic protein biomarkers published after 2015 in ProCanBio.

4. DISCUSSION

ProCanBio provides cumulative information about biomarkers related to PCa. The information from each article is manually curated before adding it into the database. Extensive information was collected from each published article to provide detailed information to the user—providing not only the regulation status and fold change, but also more information such as patient cohort, degree of validity, methods used to analyze the biomarker, and performance metrics. Besides, this database is extremely exhaustive since it focuses on genomics, proteomics, and metabolomics instead of focusing on one particular aspect such as Cancer Proteomics Database (Arntzen et al., 2015). ProCanBio encompass a total of 2053 entries, with 1497 unique biomarkers. Tabular display of information and user-friendly interface makes it easy for people from all backgrounds to use this resource. It is freely available to the scientific community without having to login to any platform. The interface is extremely user friendly and one can easily browse and search through the biomarkers database.

4.1. Comparison with Other Resources

Currently, there are not many resources available for biomarkers related to PCa. One such resource is Cancer Proteomics Database. Although it contains proteomics biomarkers, it lacks genomic, metabolomic, and epigenomic biomarkers, which were provided by ProCanBio in addition to proteins-based biomarkers. Furthermore, Cancer Proteomics Database extracted information from 143 articles up to the year 2013, whereas ProCanBio covers 412 articles. Though Cancer Proteomics Database does give some information about the biomarkers, it lacks vital information such as the performance of the biomarker, significance level, and degree of validity.

In another resource, ERDN, National Cancer Institute (NCI) publishes a list of biomarkers for several types of cancers including lung, prostate, liver, breast, and lung. They provide information for each biomarker, its aliases, which cancers are they over expressed in, and whether there are any studies related to the biomarker published by their own organization. But no further details about the biomarkers is given.

ProCanBio can be used to find supporting evidence about a specific biomarker from the published literature to further use it for experimental validation. The motivation to develop ProCanBio was to provide a freely available comprehensive database regarding all related literature on a disease that affects >1.27 million lives every year. We anticipate this resource will be useful for the scientific community actively involved in the elucidation of biomarkers for the PCa.

5. APPLICATIONS AND FUTURE DEVELOPMENT

This database can be used for a variety of applications that can help the researcher and scientific community. Some of them are listed as follows:

1. To the best of the author's knowledge, no other database currently provides information about various signatures and biomarkers for PCa from wide fields such as genomics, epigenomics, and metabolomics.
2. Easy to comprehend user interface means people with minimal knowledge can use this database.
3. Pubmed, GeneCard, ClinicalTrials, and Enrichr are all linked to the platform, making it easy for the reader to connect all the different platforms and find the relevant information related to a particular biomarker.
4. ''Complex Search'' tool allows users to find answers to complex queries that can help narrow down research already done on a particular biomarker.
5. Browsing and searching tools are for better understanding and comprehension of the existing literature.
6. ProCanBio is particularly useful to retrieve detailed supporting evidence from published literature to select a particular biomarker for further research on PCa.

5.1. Database update

Since PCa is one of major concerns worldwide, the scientific community is continuously working in this field. With the availability of more articles and sufficient information for PCa, our aim will be to update the resource annually. Besides, we will also expand information from biomarkers to the drugs and therapeutic options.

AUTHORS' CONTRIBUTIONS

Conception and design of the study were carried out by D.S., H.K., and G.P.S.R.; development of methodology was taken care of D.S., H.K., and G.P.S.R.; acquisition of data was by D.S.; analysis and interpretation of data and results were done by D.S., H.K., A.D., and G.P.S.R.; webserver development was taken care of by D.S., H.K., and A.D.; writing, reviewing, and revision of the article were carried out by D.S., H.K., A.D., and G.P.S.R.; supervision and coordination of the project were taken care of by G.P.S.R.

ACKNOWLEDGMENTS

D.S., H.K., and A.D. are grateful to the Department of Biotechnology, India, the Council of Scientific and Industrial Research, India, and the Department of Science Technology, India, for providing fellowships, respectively.

AUTHOR DISCLOSURE STATEMENT

The authors declare they have no competing financial interests.

FUNDING INFORMATION

No funding was received for this article.

REFERENCES

Adhyam, M., and Gupta, A.K. 2012. A review on the clinical utility of PSA in cancer prostate. *Indian J. Surg. Oncol.* 3, 120–129.

Agarwal, S.M., Raghav, D., Singh, H., et al. 2011. CCDB: A curated database of genes involved in cervix cancer. *Nucleic Acids Res.* 39, D975–D979.

Arntzen, M.O., Boddie, P., Frick, R., et al. 2015. Consolidation of proteomics data in the Cancer Proteomics database. *Proteomics* 15, 3765–3771.

Atan, A., and Guzel, O. 2013. How should prostate specific antigen be interpreted? *Turk J. Urol.* 39, 188–193.

Bhalla, S., Chaudhary, K., Kumar, R., et al. 2017. Gene expression-based biomarkers for discriminating early and late stage of clear cell renal cancer. *Sci. Rep.* 7, 44997.

Bhalla, S., Kaur, H., Dhall, A., et al. 2019. Prediction and analysis of skin cancer progression using genomics profiles of patients. *Sci. Rep.* 9, 15790.

Brawer, M.K. 2005. Prostatic intraepithelial neoplasia: An overview. *Rev. Urol.* 7 Suppl 3, S11–S18.

Butti, M.D., Chanfreau, H., Martinez, D., et al. 2014. BioPlat: A software for human cancer biomarker discovery. *Bioinformatics* 30, 1782–1784.

Chu, Y.W., Chien, C.H., Sung, M.I., et al. 2020. dBMHCC: A comprehensive hepatocellular carcinoma (HCC) biomarker database provides a reliable prediction system for novel HCC phosphorylated biomarkers. *PLoS One* 15, e0234084.

Dhall, A., Patiyal, S., Kaur, H., et al. 2020. Computing skin cutaneous melanoma outcome from the HLA-alleles and clinical characteristics. *Front. Genet*. 11, 221.

Dingerdissen, H.M., Bastian, F., Vijay-Shanker, K., et al. 2020. OncoMX: A knowledgebase for exploring cancer biomarkers in the context of related cancer and healthy data. *JCO Clin. Cancer Inform.* 4, 210–220.

Droz, J.P., Balducci, L., Bolla, M., et al. 2010. Management of prostate cancer in older men: Recommendations of a working group of the International Society of Geriatric Oncology. *BJU Int.* 106, 462–469.

Hessels, D., Smit, F.P., Verhaegh, G.W., et al. 2007. Detection of TMPRSS2-ERG fusion transcripts and prostate cancer antigen 3 in urinary sediments may improve diagnosis of prostate cancer. *Clin. Cancer Res*. 13, 5103–5108.

Huggins, C., and Hodges, C.V. 2002. Studies on prostatic cancer: I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. 1941. *J. Urol.* 168,9–12.

Hutchinson, L.M., Chang, E.L., Becker, C.M., et al. 2005. Use of thymosin beta15 as a urinary biomarker in human prostate cancer. *Prostate* 64, 116–127.

Jemal, A., Siegel, R., Xu, J., et al. 2010. Cancer statistics, 2010. *CA Cancer J. Clin.* 60, 277–300.

Kaur, H., Bhalla, S., Kaur, D., et al. 2020. CancerLivER: A database of liver cancer gene expression resources and biomarkers. *Database (Oxford)* 2020.

Kaur, H., Bhalla, S., and Raghava, G.P.S. 2019a. Classification of early and late stage liver hepatocellular carcinoma patients from their genomics and epigenomics profiles. *PLoS One* 14, e0221476.

Kaur, H., Dhall, A., Kumar, R., et al. 2019b. Identification of platform-independent diagnostic biomarker panel for hepatocellular carcinoma using large-scale transcriptomics data. *Front. Genet*. 10, 1306.

Kaur, H., Kumar, R., Lathwal, A., et al. 2021. Computational resources for identification of cancer biomarkers from omics data. *Brief Funct. Genomics* 20, 213–222.

Kosari, F., Cheville, J.C., Ida, C.M., et al. 2012. Shared gene expression alterations in prostate cancer and histologically benign prostate from patients with prostate cancer. *Am J. Pathol.* 181, 34–42.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., et al. 2016. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97.

Mellinger, G.T., Gleason, D., and Bailar, J, III. 1967. The histology and prognosis of prostatic cancer. *J. Urol*. 97, 331–337.

Moul, J.W., Kibel, A.S., Roach, M., 3rd, et al. 2011. Indications and practice with androgen deprivation therapy. *Urology* 78, S478–S481.

Nadler, R.B., Humphrey, P.A., Smith, D.S., et al. 1995. Effect of inflammation and benign prostatic hyperplasia on elevated serum prostate specific antigen levels. *J. Urol.* 154, 407–413.

Nalejska, E., Maczynska, E., and Lewandowska, M.A. 2014. Prognostic and predictive biomarkers: Tools in personalized oncology. *Mol. Diagn. Ther.* 18, 273–284.

Perez-Granado, J., Pinero, J., and Furlong, L.I. 2019. ResMarkerDB: A database of biomarkers of response to antibody therapy in breast and colorectal cancer. *Database (Oxford)* 2019, baz060.

Pretorius, M.E., Waehre, H., Abeler, V.M., et al. 2009. Large scale genomic instability as an additive prognostic marker in early prostate cancer. *Cell Oncol.* 31, 251–259.

Rawla, P. 2019. Epidemiology of prostate cancer. *World J. Oncol*. 10, 63–89.

Sallam, R.M. 2015. Proteomics in cancer biomarkers discovery: Challenges and applications. *Dis. Markers* 2015, 321370.

Stelzer, G., Rosen, N., Plaschkes, I., et al. 2016. The GeneCards Suite: From gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics* 54, 1.30.1–1.30.33.

Sung, H., Ferlay, J., Siegel, R.L., et al. 2021. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin*. 71,209–249.

Terkelsen, T., Krogh, A., and Papaleo, E. 2020. CAncer bioMarker Prediction Pipeline (CAMPP)—A standardized framework for the analysis of quantitative biological data. *PLoS Comput. Biol.* 16, e1007665.

Verma, S., Rajesh, A., Morales, H., et al. 2011. Assessment of aggressiveness of prostate cancer: Correlation of apparent diffusion coefficient with histologic grade after radical prostatectomy. *AJR Am. J. Roentgenol.* 196, 374–381.

Zhang, X., Sun, X.F., Cao, Y., et al. 2018. CBD: A biomarker database for colorectal cancer. *Database (Oxford)* 2018, bay046.

Zuo, Z., Hu, H., Xu, Q., et al. 2020. BBCancer: An expression atlas of blood-based biomarkers in the early diagnosis of cancers. *Nucleic Acids Res.* 48, D789–D796.

Address correspondence to:
Dr. Gajendra Pal Singh Raghava
Department of Computational Biology
Indraprastha Institute of Information Technology, Delhi
Okhla Industrial Estate, Phase III
New Delhi 110020
India

E-mail: raghava@iiitd.ac.in