

LiqBioer: a manually curated database of cancer biomarkers in body fluid

Yiding Geng^{1,†}, Lu Jin^{2,†}, Guangjue Tang², Zhangxiang Zhao³, Yunyan Gu^{1b} and Dan Yang^{1,*}

¹Department of Biochemistry and Molecular Biology, Harbin Medical University, 157 Baojian Road, Nangang District, Harbin 150081, China

²Department of Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, 157 Baojian Road, Nangang District, Harbin 150081, China

³The Sino-Russian Medical Research Centre, The Institute of Chronic Disease, The First Affiliated Hospital, Jinan University, Guangzhou, Guangdong 510630, China

*Corresponding author: Tel: +86-0451-8660-1286; Fax: +0086045186601286; Email: danyanghmu@163.com

[†]These authors contributed equally to this work.

Citation details: Geng, Y., Jin, L., Tang, G. *et al.* LiqBioer: a manually curated database of cancer biomarkers in body fluid. *Database* (2022) Vol. 2022: article ID baac077; DOI: <https://doi.org/10.1093/database/baac077>

Abstract

Cancer biomarkers are measurable indicators that play vital roles in clinical applications. Biomarkers in body fluids have gained considerable attention since the development of liquid biopsy, and their data volume is rapidly increasing. Nevertheless, current research lacks the compilation of published cancer body fluid biomarkers into a centralized and sustainable repository for researchers and clinicians, despite a handful of small-scale and specific data resources. To fulfill this purpose, we developed liquid biomarker (LiqBioer) containing 6231 manually curated records from 3447 studies, covering 3056 biomarkers and 74 types of cancer in 22 tissues. LiqBioer allows users to browse and download comprehensive information on body liquid biomarkers, including cancer types, source studies and clinical usage. As a comprehensive resource for body fluid biomarkers of cancer, LiqBioer is a powerful tool for researchers and clinicians to query and retrieve biomarkers in liquid biopsy.

Database URL: <http://www.medsysbio.org:8080/LiqBioer>

Introduction

Biomarkers are biochemical, cellular, gene-related or molecular indicators of diseases that are measurable in physical media, such as blood, body fluids, tissues or cells (1). Biomarkers are widely used in clinical applications including early cancer diagnosis, improved cancer prognosis, early detection of cancer relapse, real-time monitoring of therapeutic efficacy and therapeutic resistance (2).

Cancer is a complex disease with a high mortality rate and is a major global public health concern (3, 4). Biomarkers have been widely used for cancer diagnosis and treatment since their development (5). In the era of precision oncology, molecular profiles of individual cancer patients can facilitate treatment decisions, monitor subsequent responses and alert patients to the emergence of therapeutic resistance and relapse. Gold standard molecular profiling is performed for resected or biopsied tumors (6). However, tumor biopsy is invasive, inconvenient and difficult to track in terms of temporal and spatial heterogeneity (7).

In the last two decades, advances in liquid biopsy have paved the way for tumor tracking (8). The management and prognosis of patients will also benefit from the development of liquid biopsy technology (9, 10). Compared with tissue-based biopsy approaches, liquid biopsy is less invasive and

samples are more easily obtainable during disease progression, thus complementing the development of dynamic molecular imaging assessments (11). Liquid biopsy tests include circulating tumor cells (CTCs), circulating nucleic acids, including cell-free DNA (cfDNA) and circulating tumor DNA (ctDNA) in the plasma, cell-free RNAs (mRNAs, long non-coding RNAs and microRNAs), extracellular vesicles, tumor-educated platelets, proteins and metabolites that can be found in a range of body fluids (12, 13).

Clinically, the Food and Drug Administration approved CTC enumeration using CellSearch technology to stratify patients with breast, colorectal and prostate cancers (14). ctDNA analysis has been optimized for routine diagnosis in disease monitoring and the identification of actionable mutations (15, 16). For example, the detection of EGFR mutations in ctDNA in non-small cell lung cancer indicates that patients are ready for targeted therapy (17). Recent studies have shown that exosomes are effective in liquid biopsy for early diagnosis, disease monitoring and prognosis prediction (18–20). CTC, cfDNA and extracellular vesicles are currently the most commonly used tests in clinical practice.

Although many online tumor biomarker databases are available, they are used for tumor biopsies. For example, Liver Cancer Expression Resource (21) and Colorectal Cancer

Received 8 June 2022; Revised 15 August 2022; Accepted 27 August 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Biomarker Database (22). There are a few liquid biopsy databases, such as the Protein Body Fluid Database: human body fluid proteome (HBFP) (23). Nevertheless, HBFP contains only information about proteins and is not specific to cancer. Thus, we developed a database of liquid biomarker (LiqBioer) that covers multiple types of biomarkers for different types of cancers.

Result

Data collection

We extracted body fluid biomarkers of cancer from the literature. We searched the published literature before 28 February 2022, from the PubMed database with keywords such as ‘body fluid’, ‘liquid biopsy’, ‘circulating tumor cells’, ‘circulating DNA’, ‘circulating RNA’, ‘carcinoembryonic antigen’, ‘extracellular vesicle’, ‘exosome’, ‘ectosome’, ‘shedding microvesicle’, ‘apoptotic body’ and ‘secretome’, removing reviews, meta-analysis papers and papers without full text. We used R programming language to filter out abstracts that did not contain the ‘cancer’ word. Subsequently, we manually searched the abstracts for literature containing biomarker names, cancer types and body fluid sources. We collected drug target, sensitivity and resistance data of biomarkers from the collected papers. Figure 1A shows the common types of body fluids and biomarkers. The first two parts of Figure 1B schematically illustrate the data-collection process.

Database and web interface design

LiqBioer is freely available without registration. LiqBioer was designed to operate optimally with Google Chrome and Firefox. The web interface was implemented using Java technology (Java Server Pages, Asynchronous JavaScript, and Ajax), Google Code API and XML under the integrated development environment Eclipse (<https://www.eclipse.org/>). Public access to the database was strictly read-only. The web application was hosted and deployed on the Apache Tomcat server (version 9). LiqBioer provides a user-friendly web interface that allows for intuitive browsing and searching. The website includes five primary functional pages: ‘Home’, ‘Browser’, ‘Search’, ‘Download’ and ‘Help’. The third part of Figure 1B shows the structure of LiqBioer. Figure 2 illustrates the pages and how to use the browse and search functions.

- The ‘Home’ page provides a speedy search and synopsis of the database LiqBioer. Users can input the name of a biomarker or the type of cancer to retrieve the corresponding information.
- The ‘Browse’ page provides a browse through modules of all records. The ‘Browse’ page contains five modules, ‘biomarker’, ‘cancer types’, ‘other species’, ‘drugs’ and ‘scores’. The ‘biomarker’ module includes the types, sources and clinical usage of biomarkers. The ‘cancer’-type module displays the primary sites of cancer. Users can click on the buttons of organs or icons on the human body to browse for cancer biomarkers of specific organs.
- The ‘Search’ page allows users to set four search conditions: species, source of body fluid, clinical usage and types of biomarkers. Subsequently, the users can search for biomarkers using the name of a biomarker and the type of cancer.

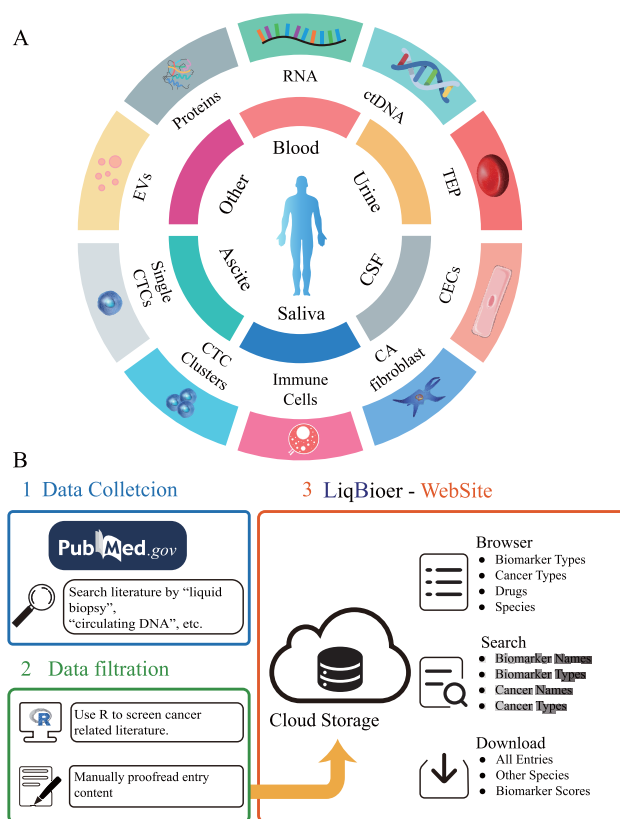


Figure 1. Construction of LiqBioer. (A) Biomarkers in body fluids, such as blood, urine, CSF, saliva and ascites. EVs, proteins, circulating cell-free RNA (non-coding and messenger RNA), ctDNA, TEPs, the tumor cell fraction with CTCs and the non-tumor cell fraction (immune cells, CECs or CA fibroblasts) can be detected in the different body fluids. CSF, cerebrospinal fluid; EV, extracellular vesicle; CA fibroblast, cancer-associated fibroblast; CEC, circulating endothelial cell; TEP, tumor-educated platelet. (B) Process of data collection, filtering and website structure. 1: Data collection. 2: Data filtration. 3: Structure of website.

- The ‘Download’ page provides the latest data on body fluid biomarkers of cancer. Users can download four files containing all the information for all markers. Other-Species.txt contains information on the biomarkers of species other than humans. Drugs.txt contains information about biomarkers that contain information about drugs. Scores.txt contains the clinical scores of the biomarkers.
- The ‘Help’ page guides users to use the website, learn the website’s statistics and contact us by e-mail. Users can obtain statistics through pie charts designed using High-Charts (a JavaScript framework for drawing statistical charts). Pie charts allow users to click on sectors to see the details of the charts. If users have any questions or suggestions, they can send us an e-mail by clicking on ‘contact us’.

Clinical scores of biomarkers

The LiqBioer database was used to calculate a statistical score to evaluate the clinical confidence of the body fluid biomarkers of cancer. The clinical score was assessed based on the following four aspects: presence/absence of a drug in that entry ($S1 = \{1, 0\}$), presence/absence

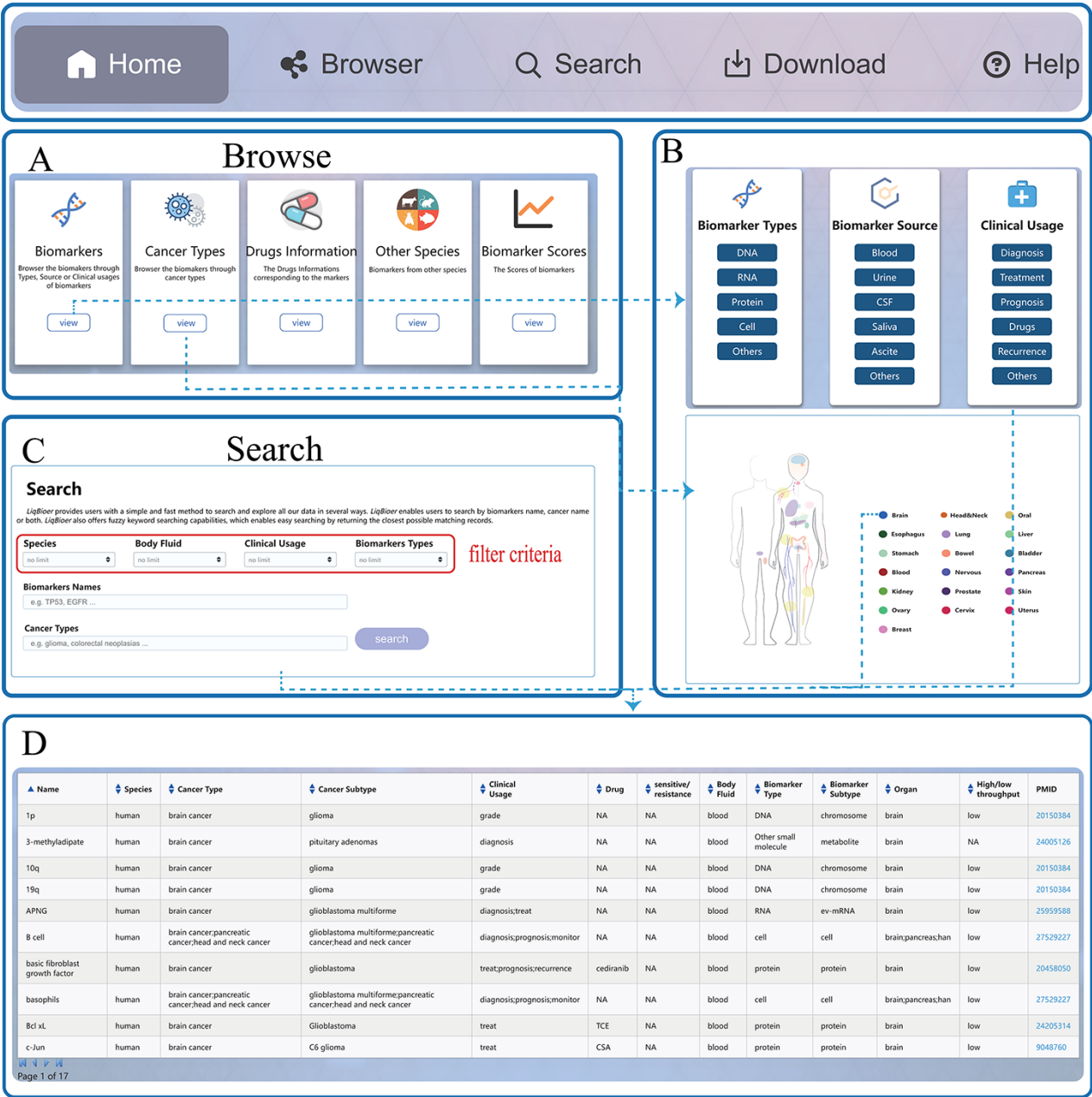


Figure 2. Usage instructions of LiqBioer. (A) Browse page. (B) Subpages of the browse page. (C) Search page. (D) Table of search and browse results.

of information on drug resistance ($S2 = \{1,0\}$), high-throughput/low-throughput/absence of detection techniques ($S3 = \{0.2,1,0\}$) and the number of papers containing the biomarker (n). In summary, a high clinical score indicates detailed medication information and low-throughput biological experiments. Clinical scores were calculated as follows:

$$\text{Score} = \sum_{n=1}^{n_i} (S1 + S2 + S3).$$

Table 1 shows the 10 biomarkers with the highest clinical scores.

Statistics of LiqBioer

LiqBioer contains 6231 entries of biomarkers among 3056 biomarkers for 74 types of cancer in 22 tissues from 3447 articles. The biomarkers were obtained from 14 body fluids,

Table 1. The 10 biomarkers with top clinical scores

Biomarkers	Score
miR-342-3p	5.854×10^3
C/EBP β	5.854×10^3
MUC5AC	5.854×10^3
circ_0007385	5.853×10^3
sTn	5.853×10^3
PHEX-AS1	5.852×10^3
RP11-77G23.5	5.851×10^3
GOLPH3	5.850×10^3
circTRPS1	5.850×10^3
miR141-3p	5.850×10^3

including blood, ascites, urine and saliva, encompassing DNA, RNA, protein, cell and other types. The LiqBioer database also contains species other than humans, such as

mouse, canine, feline and gallus (Figure 3). Figure 3A shows the distribution of biomarker types; more than half of them were RNA biomarkers. Figure 3B shows that the tissue sites and the breast tissue had the most biomarkers in body fluids. Figure 3C shows the distribution of body fluid types and

the blood accounted for the vast majority of cases. Figure 3D shows the species distribution; the human species are the majority of all species. Figure 3E shows the distribution of cancer types; breast cancer tops the list, followed by lung cancer.

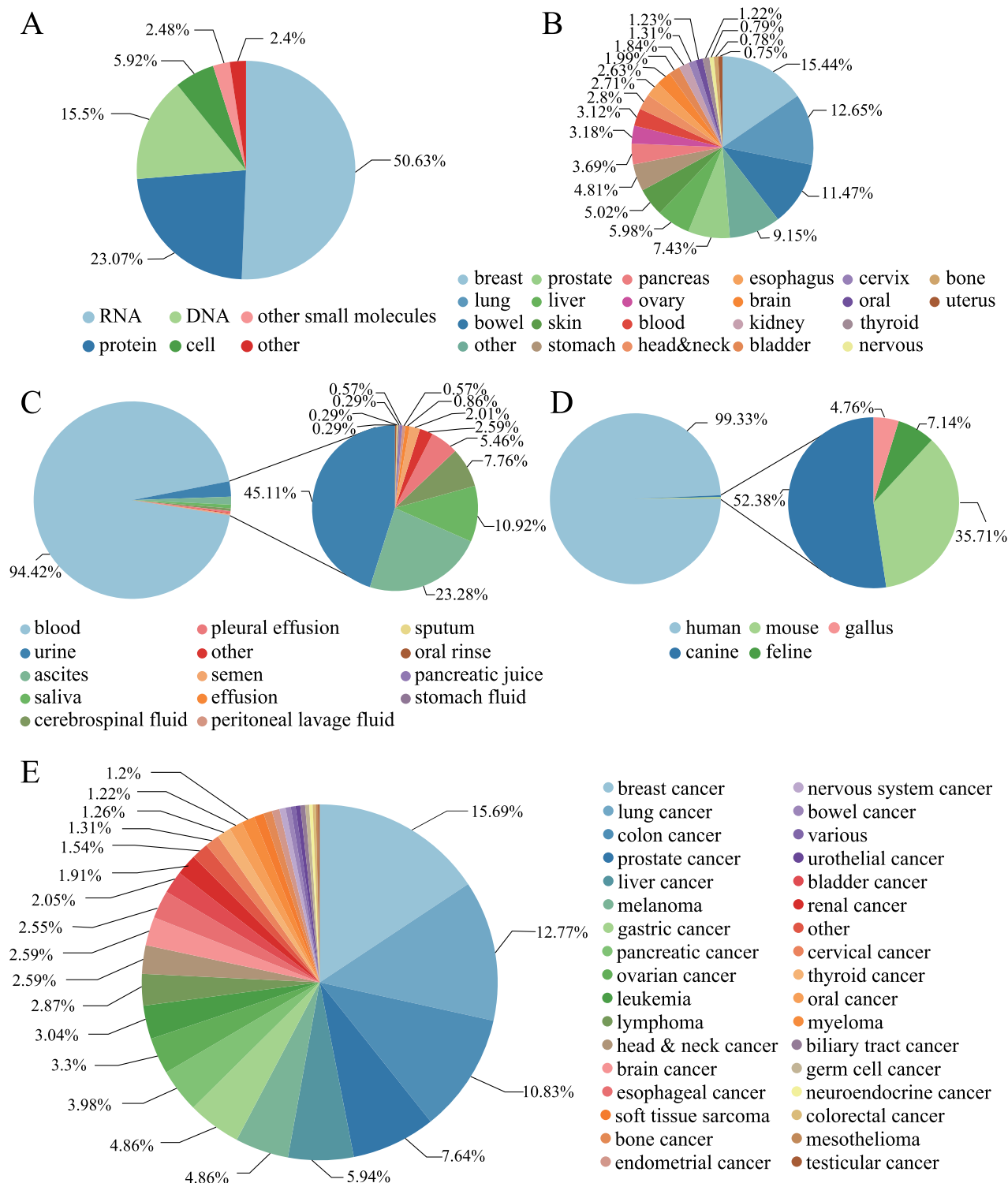


Figure 3. Statistics of biomarkers in LiqBioer. (A) Classification of biomarkers. (B) Tissues of cancer. (C) Source of body fluid. (D) Source species for biomarkers. (E) Cancer types for biomarkers.

Discussion and conclusion

LiqBioer is a manually curated database that contains multiple types of body fluid cancer biomarkers. LiqBioer can assist clinicians in diagnosis and scientific researchers in selecting the research objectives. LiqBioer will be updated regularly according to the newly available biomarkers and user-submitted data as an interactive website.

However, LiqBioer only collected data and did not design the analytics for further investigation. Analysis tools and other functional modules will be available for subsequent updates. A module on clinical guidelines and clinical trials will be added to enable clinicians in accessing information more effectively.

In conclusion, LiqBioer will be a valuable tool for researchers and clinicians to retrieve biomarkers for liquid biopsy of tumors and manage cancer treatment.

Abbreviations

cfDNA: tumor-derived fraction of cell-free DNA

ctDNA: circulating tumor DNA

CA fibroblast: cancer-associated fibroblast

CEC: circulating endothelial cell

CTC: circulating tumor cell

EV: extracellular vesicle

LiqBioer: Liquid Biomarker

TEP: tumor-educated platelet

Conflict of interest

The authors declare that they have no competing interests.

References

- Walker, L.E., Janigro, D., Heinemann, U. *et al.* (2016) WONOEP appraisal: molecular and cellular biomarkers for epilepsy. *Epilepsia*, **57**, 1354–1362. [10.1111/epi.13460](#).
- Alix-Panabieres, C. and Pantel, K. (2021) Liquid biopsy: from discovery to clinical application. *Cancer Discov.*, **11**, 858–873. [10.1158/2159-8290.CD-20-1311](#).
- 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. [10.3322/caac.21660](#).
- Mao, J.J., Pillai, G.G., Andrade, C.J. *et al.* (2022) Integrative oncology: addressing the global challenges of cancer prevention and treatment. *CA Cancer J. Clin.*, **72**, 144–164. [10.3322/caac.21706](#).
- Wira Wiguna, I.G.W., Indrani Remitha, N.P.S., Sadvika, I. *et al.* (2022) Pretreatment leukocyte count ratios as metastatic predictive factors in luminal type breast cancer. *Asian Pac. J. Cancer Prev.*, **23**, 1595–1601. [10.31557/APJCP.2022.23.5.1595](#).
- Kilgour, E., Rothwell, D.G., Brady, G. *et al.* (2020) Liquid biopsy-based biomarkers of treatment response and resistance. *Cancer Cell*, **37**, 485–495. [10.1016/j.ccell.2020.03.012](#).
- Fatima, S., Ma, Y., Safrachi, A. *et al.* (2022) Harnessing liquid biopsies to guide immune checkpoint inhibitor therapy. *Cancers (Basel)*, **14**, 1669–1669. [10.3390/cancers14071669](#).
- Patel, A., Patel, S., Patel, P. *et al.* (2022) Saliva based liquid biopsies in head and neck cancer: how far are we from the clinic? *Front. Oncol.*, **12**, 828434–828434. [10.3389/fonc.2022.828434](#).
- Pantel, K. and Alix-Panabieres, C. (2010) Circulating tumour cells in cancer patients: challenges and perspectives. *Trends Mol. Med.*, **16**, 398–406. [10.1016/j.molmed.2010.07.001](#).
- Wan, J.C.M., Massie, C., Garcia-Corbacho, J. *et al.* (2017) Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat. Rev. Cancer*, **17**, 223–238. [10.1038/nrc.2017.7](#).
- Cherry, S.R., Jones, T., Karp, J.S. *et al.* (2018) Total-body PET: maximizing sensitivity to create new opportunities for clinical research and patient care. *J. Nucl. Med.*, **59**, 3–12. [10.2967/jnumed.116.184028](#).
- Heitzer, E., Haque, I.S., Roberts, C.E.S. *et al.* (2019) Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.*, **20**, 71–88. [10.1038/s41576-018-0071-5](#).
- Ramalingam, N. and Jeffrey, S.S. (2018) Future of liquid biopsies with growing technological and bioinformatics studies: opportunities and challenges in discovering tumor heterogeneity with single-cell level analysis. *Cancer J.*, **24**, 104–108. [10.1097/PP0.0000000000000308](#).
- Riethdorf, S., Fritsche, H., Muller, V. *et al.* (2007) Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: a validation study of the cell search system. *Clin. Cancer Res.*, **13**, 920–928. [10.1158/1078-0432.CCR-06-1695](#).
- Dawson, S.J., Tsui, D.W., Murtaza, M. *et al.* (2013) Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.*, **368**, 1199–1209. [10.1056/NEJMoa1213261](#).
- Garcia-Murillas, I., Schiavon, G., Weigelt, B. *et al.* (2015) Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci. Transl. Med.*, **7**, 302ra133. [10.1126/scitranslmed.aab0021](#).
- Kwapisz, D. (2017) The first liquid biopsy test approved. Is it a new era of mutation testing for non-small cell lung cancer? *Ann. Transl. Med.*, **5**, 46. [10.21037/atm.2017.01.32](#).
- Sun, Z., Shi, K., Yang, S. *et al.* (2018) Effect of exosomal miRNA on cancer biology and clinical applications. *Mol. Cancer*, **17**, 147. [10.1186/s12943-018-0897-7](#).
- Lee, Y.R., Kim, G., Tak, W.Y. *et al.* (2019) Circulating exosomal noncoding RNAs as prognostic biomarkers in human hepatocellular carcinoma. *Int. J. Cancer*, **144**, 1444–1452. [10.1002/ijc.31931](#).
- Vasconcelos, M.H., Caires, H.R., Abols, A. *et al.* (2019) Extracellular vesicles as a novel source of biomarkers in liquid biopsies for monitoring cancer progression and drug resistance. *Drug Resist. Updat.*, **47**, 100647. [10.1016/j.drug.2019.100647](#).
- Kaur, H., Bhalla, S., Kaur, D. *et al.* (2020) CancerLiver: a database of liver cancer gene expression resources and biomarkers. *Database (Oxford)*, **2020**, 1–11. [10.1093/database/baaa012](#).
- Zhang, X., Sun, X.F., Cao, Y. *et al.* (2018) CBD: a biomarker database for colorectal cancer. *Database (Oxford)*, **2018**, 1–12. [10.1093/database/bay046](#).
- Shao, D., Huang, L., Wang, Y. *et al.* (2021) HBFP: a new repository for human body fluid proteome. *Database (Oxford)*, **2021**, 1–14. [10.1093/database/baab065](#).