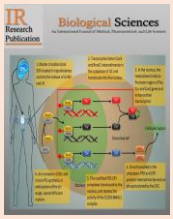
REVIEW ARTICLE

IR Research Publication

Biological Sciences

(An International Journal of Medical, Pharmaceutical, and Life Sciences)

Journal home: https://irrespub.com/biolsciences/index.php/1/index



Bioinformatics approaches for cancer biomarker discovery

Gaurav Mishra[1*#], Saloni Rahi[2#]

[1]*ISF College of Pharmacy, GT Road, Ghal Kalan, MOGA– 142001 (Punjab), India.*
[2]*Pharmacology Division, University Institute of Pharmaceutical Sciences, Punjab University, Chandigarh-160014, India.*

ARTICLE HISTORY

Received: 19-10-2022
Revised: 01-12-2022
Accepted: 05-12-2022
Online: 10-12-2022

KEYWORDS

Cancer
Biomarkers
Bioinformatics
CRISPR

ABSTRACT

Cancer biomarker discovery is essential for early diagnosis, prognosis, and personalized treatment strategies. Recent advancements in bioinformatics have significantly enhanced the identification of novel biomarkers, providing deeper insights into the molecular underpinnings of cancer. This review explores various bioinformatics approaches used in cancer biomarker discovery, including genomics, transcriptomics, proteomics, and metabolomics. We discuss how high-throughput sequencing, data integration, and machine learning tools enable the identification of diagnostic, prognostic, and predictive biomarkers across cancer types. Additionally, the review highlights the challenges associated with cancer biomarker discovery, such as data heterogeneity, validation issues, and ethical concerns. Furthermore, emerging technologies like single-cell RNA sequencing, CRISPR-based screening, and 3D tumor modeling are shaping the future of cancer biomarker research, enabling a more personalized approach to cancer treatment. Bioinformatics is pivotal in driving precision medicine, facilitating the development of targeted therapies based on tumor-specific biomarkers. As these technologies continue to evolve, bioinformatics will play a crucial role in advancing cancer diagnosis, improving treatment outcomes, and ultimately enhancing patient survival.

Introduction

Cancer is one of the leading causes of death worldwide, with an estimated 9.6 million deaths annually (World Health Organization, 2020) (1). The disease is characterized by uncontrolled cell growth and can arise from a variety of genetic and environmental factors. Over the years, the detection and management of cancer have significantly improved, largely due to advancements in cancer biomarkers. Cancer biomarkers are molecular signatures, including genes, proteins, and metabolites, which can provide critical insights into the presence, progression, and prognosis of cancer. These biomarkers serve as tools for early diagnosis, determining disease prognosis, and selecting the appropriate treatment regimens (2).

The identification and use of cancer biomarkers can be categorized into four major types: diagnostic, prognostic, predictive, and therapeutic. Diagnostic biomarkers, such as prostate-specific antigen (PSA) in prostate cancer, allow for the early detection of the disease, often before clinical symptoms

[#]*Contributed equally*
**Address for correspondence*
ISF College of Pharmacy, GT Road, Ghal Kalan, MOGA–142001 (Punjab), India.
Email: mgaurav3807@gmail.com
DOI: https://doi.org/10.55006/biolsciences.2022.2404
Published by IR Research Publication;

manifest. Prognostic biomarkers provide information about the likely progression of the disease, such as the HER2 receptor status in breast cancer, which helps predict tumor aggressiveness (3). Predictive biomarkers, such as mutations in the EGFR gene in lung cancer, help determine how well a patient will respond to specific treatments (4). Finally, therapeutic biomarkers can guide treatment decisions by monitoring the effectiveness of therapeutic interventions (5).

Despite the considerable promise of cancer biomarkers, their discovery and clinical implementation face several challenges. One of the primary obstacles is tumor heterogeneity, the genetic and phenotypic variability among tumor cells within the same tumor and across different patients. This variability complicates the identification of universal biomarkers that are applicable across different populations and cancer types (6). Additionally, the complexity of molecular pathways involved in cancer, such as mutations, epigenetic alterations, and interactions within the tumor microenvironment, makes it difficult to pinpoint reliable and reproducible biomarkers (7).

Bioinformatics has become a critical tool in overcoming these challenges. By leveraging high-throughput technologies and computational methods, bioinformatics enables the analysis of vast amounts of omics data (genomics, transcriptomics, proteomics, and metabolomics), facilitating the identification of potential biomarkers. Furthermore, bioinformatics approaches allow for the integration of multi-omics data, helping to unravel the complex interactions that drive cancer progression and identify biomarkers that might otherwise remain undetected (8). This computational approach has the potential to significantly accelerate cancer biomarker discovery, making it possible to develop more accurate and personalized diagnostic, prognostic, and therapeutic strategies.

Types of Cancer Biomarkers

Cancer biomarkers are categorized based on their utility in clinical practice, helping guide treatment, predict outcomes, or detect the disease at early stages. These biomarkers include diagnostic, prognostic, predictive, and therapeutic biomarkers, each serving a unique role in cancer management.

Diagnostic Biomarkers

These biomarkers are critical for the early detection and screening of cancer. Diagnostic biomarkers enable the identification of cancer in asymptomatic individuals or those at high risk, often before clinical signs manifest. For example, prostate-specific antigen (PSA) is a widely used diagnostic biomarker for prostate cancer, helping in the early detection and monitoring of the disease (9). Another example is CA-125, a biomarker for ovarian cancer, which is used in conjunction with imaging for early diagnosis (10). Early detection of cancer through these biomarkers significantly improves treatment outcomes by facilitating earlier intervention.

Prognostic Biomarkers

Prognostic biomarkers provide valuable information regarding the likely course and outcome of the disease, helping clinicians predict disease progression and survival rates. An example is HER2/neu in breast cancer, where overexpression of this protein correlates with a more aggressive form of the disease and a poorer prognosis (11). Similarly, in colorectal cancer, the KRAS mutation serves as a prognostic biomarker, with its presence being associated with a lower survival rate and reduced response to certain therapies (12). Prognostic biomarkers are essential for determining treatment intensity and follow-up strategies.

Predictive Biomarkers

These biomarkers are used to predict the response to specific treatments, helping clinicians select the most effective therapeutic options. For example, EGFR mutations in non-small cell lung cancer (NSCLC) serve as predictive biomarkers for the efficacy of EGFR inhibitors such as gefitinib or erlotinib (4). Similarly, the BRAF V600E mutation in melanoma is predictive of response to BRAF inhibitors like vemurafenib (13). Predictive biomarkers are crucial for personalizing cancer treatment and ensuring that patients receive the most appropriate therapy.

Therapeutic Biomarkers

These biomarkers are directly related to treatment efficacy and are used to monitor the effectiveness of ongoing treatment. For instance, circulating tumor DNA (ctDNA) has been explored as a therapeutic biomarker for assessing treatment response and monitoring minimal residual disease in various cancers, including breast and lung cancer (14). Another example is tumor PD-L1 expression, which is used as a therapeutic biomarker to predict response to immune checkpoint inhibitors like nivolumab and pembrolizumab in cancers such as melanoma, non-small cell lung cancer, and head and neck cancer (15).

These various types of biomarkers are essential in tailoring cancer management strategies, from early diagnosis to treatment monitoring, and continue to

evolve with advancements in bioinformatics and personalized medicine.

Bioinformatics Approaches for Cancer Biomarker Discovery

The discovery of cancer biomarkers through bioinformatics involves the analysis of vast datasets generated by various omics technologies. Bioinformatics tools help researchers mine, analyze, and integrate these datasets to identify potential biomarkers that can be used for diagnosis, prognosis, and treatment. Here, we explore key bioinformatics approaches, focusing on genomics, transcriptomics, proteomics, metabolomics, and the integration of multi-omics data.

Omics Technologies

Genomics: Genomics, particularly DNA sequencing, plays a central role in cancer biomarker discovery. Whole-genome sequencing (WGS) and targeted sequencing allow for the comprehensive analysis of genetic alterations, including mutations, amplifications, and deletions. These alterations are often associated with cancer progression and can serve as potential biomarkers. WGS provides a detailed view of the entire genome, enabling the detection of somatic mutations that may drive cancer (16). On the other hand, targeted sequencing focuses on specific genes or regions of interest, enabling a more cost-effective approach for biomarker discovery (17).

The identification of driver mutations, which are mutations that confer a selective growth advantage to tumor cells, is a key goal in cancer genomics. Bioinformatics tools such as Mutect2 (18) and GATK (19) are frequently used to identify somatic mutations and determine their potential role in cancer progression. Driver mutations like those in the EGFR gene in non-small cell lung cancer (NSCLC) are well-known examples of genetic alterations that have become targeted therapeutic biomarkers (4).

Transcriptomics: Transcriptomics, the study of the transcriptome (the complete set of RNA transcripts), is another critical omics technology in biomarker discovery. RNA sequencing (RNA-Seq) has revolutionized transcriptomics by enabling the comprehensive profiling of gene expression, alternative splicing, and non-coding RNA expression (20). By comparing gene expression profiles between normal and cancerous tissues, researchers can identify genes that are over- or under-expressed in cancer, which may serve as diagnostic, prognostic, or therapeutic biomarkers.

Tools like DESeq2 (21) and edgeR (22) are commonly used for differential gene expression analysis from RNA-Seq data. These tools help identify genes whose expression levels are significantly altered in cancer, leading to the identification of potential biomarkers. For example, overexpression of MYC and TP53 mutations are frequently observed in various cancers, making them important cancer biomarkers (7).

In addition to protein-coding genes, RNA-Seq can also uncover the role of non-coding RNAs, such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), which have been shown to regulate cancer-related pathways and can serve as biomarkers (23,24).

Proteomics: Proteomics is the study of the entire set of proteins in a cell, which reflects the functional output of the genome. Mass spectrometry (MS) is the primary technique used to profile the proteome, providing insights into the expression and post-translational modifications of proteins. Proteomics allows researchers to identify protein biomarkers that may be overlooked by genomics and transcriptomics. Techniques like liquid chromatography-tandem mass spectrometry (LC-MS/MS) are commonly used for high-throughput proteomics analysis.

Bioinformatics tools such as MaxQuant (25) and Perseus (26) help process and analyze proteomics data, allowing for the identification of cancer-specific proteins. Additionally, protein interaction networks, which map the interactions between proteins in a cell, provide valuable insights into cancer-associated pathways. Resources like STRING (27) enable researchers to explore these protein-protein interaction networks, helping to identify key regulatory nodes that could serve as therapeutic targets.

An example of a proteomic biomarker is HER2 in breast cancer, where elevated levels of the HER2 protein correlate with aggressive tumor growth and poor prognosis. This has led to the development of targeted therapies like trastuzumab (Herceptin) (28).

Metabolomics: Metabolomics focuses on the study of metabolites, small molecules that are the end products of cellular processes. Cancer cells often exhibit altered metabolism, which can be detected through metabolomics. This is known as the Warburg effect, where cancer cells primarily rely on glycolysis, even in the presence of oxygen, leading to the accumulation of specific metabolites (29).

Techniques such as gas chromatography-mass spectrometry (GC-MS) and NMR spectroscopy are

commonly used to profile metabolites in cancer. Bioinformatics tools like XCMS (30) and MetaboAnalyst (31) are used to analyze metabolomics data and identify potential cancer biomarkers. For example, altered levels of metabolites such as lactate, pyruvate, and 2-hydroxyglutarate have been observed in various cancers and could serve as diagnostic or prognostic biomarkers (32,33).

Data Integration and Multi-Omics: Cancer is a complex disease that involves alterations at multiple molecular levels. Integrating data from genomics, transcriptomics, proteomics, and metabolomics (multi-omics) allows researchers to gain a more comprehensive understanding of cancer biology and identify more robust biomarkers. Multi-omics approaches can highlight relationships between different molecular layers, such as how genetic mutations influence gene expression and protein levels.

Bioinformatics tools such as iCluster (34) and MOFA (35) facilitate the integration of multi-omics data and help identify biomarkers that are consistently altered across different omics layers. For example, integrating transcriptomic and proteomic data can uncover discrepancies between mRNA levels and protein expression, highlighting post-transcriptional regulatory mechanisms. This approach has led to the discovery of several potential biomarkers that may have been overlooked by individual omics analyses.

Machine Learning and AI: Machine learning (ML) and artificial intelligence (AI) techniques are increasingly used in bioinformatics for cancer biomarker discovery. These methods enable the analysis of large, complex datasets to identify patterns and relationships that might be hidden in the data. ML algorithms like Random Forests (36) and Support Vector Machines (SVM) are widely used for feature selection and classification, helping to identify the most relevant biomarkers from large datasets.

Deep learning, a subset of ML, is particularly effective in handling large-scale data such as genomic, transcriptomic, and imaging data. Deep neural networks (DNNs) have been used to predict cancer outcomes based on multi-omics data (37). For example, deep learning models have been employed to predict cancer subtypes and patient responses to therapies based on gene expression profiles (38).

Statistical and Computational Models: Statistical and computational models are essential for analyzing omics data and identifying cancer biomarkers. Principal Component Analysis (PCA) is commonly used to reduce the dimensionality of large datasets and identify key features that contribute to cancer progression (39). Clustering algorithms such as k-means and hierarchical clustering are used to group cancer samples based on similar molecular characteristics, helping to identify subtypes of cancer that may respond differently to treatment.

Network analysis tools like Cytoscape (40) are used to map molecular interactions and identify key nodes in cancer pathways. By identifying these nodes, researchers can pinpoint potential therapeutic targets. Computational models, such as gene set enrichment analysis (GSEA) (41), help validate biomarkers by identifying enriched pathways and gene sets that are dysregulated in cancer.

Public Databases and Tools: Public databases and resources play a crucial role in cancer biomarker discovery by providing access to large-scale cancer-related datasets. The Cancer Genome Atlas (TCGA) is one of the most widely used resources, offering comprehensive genomic, transcriptomic, and clinical data across multiple cancer types (42). Other valuable resources include Gene Expression Omnibus (GEO) for gene expression data and Cancer Cell Line Encyclopedia (CCLE) for cell line-based genomic data.

These databases allow researchers to mine cancer-related data and validate potential biomarkers across different cancer types. Bioinformatics tools like UCSC Genome Browser (43) and Ensembl (44) provide visualization and analysis platforms for exploring these datasets.

Challenges in Cancer Biomarker Discovery

Despite significant advancements in bioinformatics techniques, several challenges persist in cancer biomarker discovery, especially regarding data quality, validation, and ethical considerations.

Data Quality and Heterogeneity

One of the most significant challenges in cancer biomarker discovery is the heterogeneity of cancer itself, both within and between individual patients. Tumors are not uniform, and the genetic and molecular variations observed across different patients complicate the identification of consistent biomarkers. Cancer cells frequently exhibit genetic mutations, copy number alterations, and epigenetic modifications that are unique to each tumor, contributing to this heterogeneity (6). For example, breast cancer subtypes, such as HER2-positive,

estrogen receptor-positive, and triple-negative breast cancer, each display distinct molecular profiles and treatment responses, making it difficult to find universal biomarkers for early detection or prognosis (45).

Moreover, integrating data from diverse patient populations with varying genetic backgrounds, environmental exposures, and tumor types adds another layer of complexity. Data quality can vary significantly between different sequencing platforms, tissue types, and patient cohorts, making it challenging to derive reliable conclusions (46). These variations underscore the need for standardized protocols in data collection and analysis to ensure consistency and reproducibility in biomarker discovery.

Validation Issues

Another major obstacle is the challenge of translating bioinformatics findings into clinically validated biomarkers. Many biomarkers identified through bioinformatics approaches, such as next-generation sequencing (NGS) or RNA sequencing (RNA-Seq), are often based on in silico predictions or initial laboratory findings. However, these biomarkers must undergo rigorous validation in clinical settings to confirm their relevance and applicability in patient care. This validation process requires large, diverse patient cohorts, which are difficult to assemble, especially for rare cancer types (47).

Furthermore, biomarkers identified in preclinical or early clinical studies may not perform as well in larger, more heterogeneous patient populations. The ability to validate biomarkers in real-world clinical trials is often hindered by the difficulty of matching specific biomarkers to appropriate patient groups (48). Moreover, issues such as the cost and time required for these validation efforts contribute to the slow pace of biomarker translation.

Ethical Considerations

Ethical concerns are an integral part of cancer biomarker discovery, especially when dealing with patient data and biospecimens. The use of human samples for genomic research raises important issues related to patient consent, privacy, and the storage of sensitive data. Given that many bioinformatics approaches rely on large datasets of patient information, it is critical to ensure that these data are anonymized and stored securely to prevent misuse (49).

Patient consent for the use of tissue samples in research is another major ethical challenge. While many research studies require informed consent, the complexity of genomic data and the potential for findings with future implications can make it difficult to explain these risks to patients. Furthermore, there are concerns over the sharing of patient data, particularly in international collaborations, where privacy laws may differ (50). These ethical challenges must be addressed to maintain public trust in research and ensure that cancer biomarker discovery proceeds in a responsible and transparent manner.

Future Directions and Conclusion

The field of bioinformatics is rapidly evolving, with several emerging tools and technologies poised to revolutionize cancer biomarker discovery and precision medicine. One such advancement is single-cell RNA sequencing (scRNA-seq), which allows for the examination of gene expression at the individual cell level. This technology has been pivotal in uncovering tumor heterogeneity, identifying rare cancer cell populations, and understanding how these cells contribute to drug resistance and metastasis (51). By providing a more detailed picture of tumor biology, scRNA-seq holds immense potential for discovering novel biomarkers and refining cancer classification systems. Another promising area is the use of CRISPR-based screening for functional genomics in cancer. CRISPR-Cas9 technology enables the systematic disruption of genes to identify those essential for tumor growth and survival. Through genome-wide CRISPR screens, researchers can pinpoint novel cancer-related genes and pathways that could serve as therapeutic targets (52). This technology is increasingly integrated with bioinformatics tools to analyze large-scale data and identify genetic vulnerabilities in cancer cells, leading to the discovery of new biomarkers for both diagnosis and treatment. 3D tumor modeling represents another cutting-edge approach in bioinformatics. Tumor spheroids and organoids, when combined with genomic and transcriptomic data, provide more accurate representations of the tumor microenvironment compared to traditional 2D cell cultures. These models enable researchers to study cancer biology in a more physiologically relevant context, offering insights into tumor evolution, drug resistance mechanisms, and potential biomarkers (53). The integration of 3D models with bioinformatics pipelines will likely drive the next generation of cancer biomarker discovery. Looking forward, bioinformatics will continue to play a critical role in personalized cancer treatment. By analyzing the molecular profiles of tumors, bioinformatics tools can help identify the most effective treatments for individual patients, considering both genetic alterations and tumor microenvironment factors. For example, bioinformatics approaches are already

being used to match patients with targeted therapies based on specific mutations or biomarkers, such as EGFR mutations in non-small cell lung cancer (54-56). This alignment of biomarkers with treatment strategies is a cornerstone of precision medicine, ensuring that patients receive the most suitable and effective therapy. In conclusion, bioinformatics is a driving force behind the ongoing transformation of cancer biomarker discovery. Through the integration of advanced technologies such as scRNA-seq, CRISPR screening, and 3D tumor modeling, bioinformatics is enhancing our understanding of cancer biology and enabling the identification of novel biomarkers. These advances not only improve early detection and prognosis but also support the development of personalized treatment strategies, ultimately improving patient outcomes and survival.

Contribution of authors

Both the authors contributed equally to literature survey, compiling information, and drafting and editing manuscripts.

Acknowledgments

Not Applicable

Conflict of interest

Declared none

References

1. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global cancer observatory: cancer today. Lyon Int agency Res cancer. 2020;20182020.
2. Sarhadi VK, Armengol G. Molecular biomarkers in cancer. Biomolecules. 2022;12(8):1021.
3. Nicolini A, Ferrari P, Duffy MJ. Prognostic and predictive biomarkers in breast cancer: Past, present and future. In: Seminars in cancer biology. Elsevier; 2018. p. 56–73.
4. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non–small-cell lung cancer to gefitinib. N Engl J Med. 2004;350(21):2129–39.
5. Patel JN, Fong MK, Jagosky M. Colorectal cancer biomarkers in the era of personalized medicine. J Pers Med. 2019;9(1):3.
6. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74.
7. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz Jr LA, Kinzler KW. Cancer genome landscapes. Science (80-). 2013;339(6127):1546–58.
8. Ritchie ME, Phipson B, Wu DI, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47–e47.
9. Hayes JH, Barry MJ. Screening for Prostate Cancer With the Prostate-Specific Antigen Test: A Review of Current Evidence. JAMA [Internet]. 2014 Mar 19;311(11):1143–9. Available from: https://doi.org/10.1001/jama.2014.2085
10. Razmi N, Hasanzadeh M. Current advancement on diagnosis of ovarian cancer using biosensing of CA 125 biomarker: Analytical approaches. TrAC Trends Anal Chem [Internet]. 2018;108:1–12. Available from: https://www.sciencedirect.com/science/article/pii/S0165993618303650
11. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human Breast Cancer: Correlation of Relapse and Survival with Amplification of the HER-2/neu Oncogene. Science (80-) [Internet]. 1987 Jan 9;235(4785):177–82. Available from: https://doi.org/10.1126/science.3798106
12. Arrington AK, Heinrich EL, Lee W, Duldulao M, Patel S, Sanchez J, et al. Prognostic and Predictive Roles of KRAS Mutation in Colorectal Cancer. Vol. 13, International Journal of Molecular Sciences. 2012. p. 12153–68.
13. Ravnan MC, Matalka MS. Vemurafenib in Patients With BRAF V600E Mutation–Positive Advanced Melanoma. Clin Ther [Internet]. 2012;34(7):1474–86. Available from: https://www.sciencedirect.com/science/article/pii/S0149291812003724
14. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. Sci Transl Med [Internet]. 2014 Feb 19;6(224):224ra24-224ra24. Available from: https://doi.org/10.1126/scitranslmed.3007094
15. Hu X, Lin Z, Wang Z, Zhou Q. Emerging role of PD-L1 modification in cancer immunotherapy. Am J Cancer Res. 2021;11(8):3832–40.
16. Ansorge WJ. Next-generation DNA sequencing techniques. N Biotechnol [Internet]. 2009;25(4):195–203. Available from: https://www.sciencedirect.com/science/article/pii/S1871678409000089

17. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol [Internet]. 2008;26(10):1135–45. Available from: https://doi.org/10.1038/nbt1486
18. Bartha Á, Győrffy B. Comprehensive Outline of Whole Exome Sequencing Data Analysis Tools Available in Clinical Oncology. Vol. 11, Cancers. 2019.
19. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010 Sep;20(9):1297–303.
20. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet [Internet]. 2009;10(1):57–63. Available from: https://doi.org/10.1038/nrg2484
21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol [Internet]. 2014;15(12):550. Available from: https://doi.org/10.1186/s13059-014-0550-8
22. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics [Internet]. 2010 Jan 1;26(1):139–40. Available from: https://doi.org/10.1093/bioinformatics/btp616
23. Schmitt AM, Chang HY. Long Noncoding RNAs in Cancer Pathways. Cancer Cell [Internet]. 2016 Apr 11;29(4):452–63. Available from: https://doi.org/10.1016/j.ccell.2016.03.010
24. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer [Internet]. 2006;6(11):857–66. Available from: https://doi.org/10.1038/nrc1997
25. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol [Internet]. 2008;26(12):1367–72. Available from: https://doi.org/10.1038/nbt.1511
26. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods [Internet]. 2016;13(9):731–40. Available from: https://doi.org/10.1038/nmeth.3901
27. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res [Internet]. 2015 Jan 28;43(D1):D447–52. Available from: https://doi.org/10.1093/nar/gku1003
28. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med. 2001 Mar;344(11):783–92.
29. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. Science (80-) [Internet]. 2009 May 22;324(5930):1029–33. Available from: https://doi.org/10.1126/science.1160809
30. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Anal Chem [Internet]. 2006 Feb 1;78(3):779–87. Available from: https://doi.org/10.1021/ac051437y
31. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res [Internet]. 2018 Jul 2;46(W1):W486–94. Available from: https://doi.org/10.1093/nar/gky310
32. Sourbier C, Ricketts CJ, Matsumoto S, Crooks DR, Liao P-J, Mannes PZ, et al. Targeting ABL1-Mediated Oxidative Stress Adaptation in Fumarate Hydratase-Deficient Cancer. Cancer Cell [Internet]. 2014 Dec 8;26(6):840–50. Available from: https://doi.org/10.1016/j.ccell.2014.10.005
33. Ježek P. 2-Hydroxyglutarate in Cancer Cells. Antioxid Redox Signal [Internet]. 2019 Dec 18;33(13):903–26. Available from: https://doi.org/10.1089/ars.2019.7902
34. Shen R, Mo Q, Schultz N, Seshan VE, Olshen AB, Huse J, et al. Integrative Subtype Discovery in Glioblastoma Using iCluster. PLoS One [Internet]. 2012 Apr 23;7(4):e35236. Available from: https://doi.org/10.1371/journal.pone.0035236
35. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol [Internet]. 2020;21(1):111. Available from: https://doi.org/10.1186/s13059-020-02015-1
36. Cutler A, Cutler DR, Stevens JR. Random Forests BT - Ensemble Machine Learning: Methods and Applications. In: Zhang C, Ma Y, editors. New York, NY: Springer New York; 2012. p. 157–75. Available from: https://doi.org/10.1007/978-1-4419-9326-7_5
37. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature [Internet]. 2017;542(7639):115–8. Available from: https://doi.org/10.1038/nature21056

38. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson J V, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med [Internet]. 2021;13(1):152. Available from: https://doi.org/10.1186/s13073-021-00968-x
39. Jolliffe IT, editor. Principal Component Analysis and Factor Analysis BT - Principal Component Analysis. In New York, NY: Springer New York; 2002. p. 150–66. Available from: https://doi.org/10.1007/0-387-22440-8_7
40. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498–504.
41. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci [Internet]. 2005 Oct 25;102(43):15545–50. Available from: https://doi.org/10.1073/pnas.0506580102
42. Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet [Internet]. 2013;45(10):1113–20. Available from: https://doi.org/10.1038/ng.2764
43. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002 Jun;12(6):996–1006.
44. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res. 2020 Jan;48(D1):D682–8.
45. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature [Internet]. 2000;406(6797):747–52. Available from: https://doi.org/10.1038/35021093
46. Jia Q, Chu H, Jin Z, Long H, Zhu B. High-throughput single-cell sequencing in cancer research. Signal Transduct Target Ther [Internet]. 2022;7(1):145. Available from: https://doi.org/10.1038/s41392-022-00990-4
47. Brooks JD. Translational genomics: the challenge of developing cancer biomarkers. Genome Res. 2012 Feb;22(2):183–7.
48. Bensalah K, Montorsi F, Shariat SF. Challenges of Cancer Biomarker Profiling. Eur Urol [Internet]. 2007;52(6):1601–9. Available from: https://www.sciencedirect.com/science/article/pii/S0302283807012109
49. Angell M. The ethics of clinical research in the Third World. Vol. 337, The New England journal of medicine. United States; 1997. p. 847–9.
50. Oliver JM, Slashinski MJ, Wang T, Kelly PA, Hilsenbeck SG, McGuire AL. Balancing the Risks and Benefits of Genomic Data Sharing: Genome Research Participants' Perspectives. Public Health Genomics [Internet]. 2011 Dec 30;15(2):106–14. Available from: https://doi.org/10.1159/000334718
51. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature [Internet]. 2011;472(7341):90–4. Available from: https://doi.org/10.1038/nature09807
52. Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR–Cas9. Nat Rev Genet [Internet]. 2015;16(5):299–311. Available from: https://doi.org/10.1038/nrg3899
53. Hannoun Z, Steichen C, Dianat N, Weber A, Dubart-Kupperschmitt A. The potential of induced pluripotent stem cell derived hepatocytes. J Hepatol [Internet]. 2016;65(1):182–99. Available from: https://www.sciencedirect.com/science/article/pii/S0168827816001586
54. Jänne PA, Engelman JA, Johnson BE. Epidermal Growth Factor Receptor Mutations in Non–Small-Cell Lung Cancer: Implications for Treatment and Tumor Biology. J Clin Oncol [Internet]. 2025 Feb 2;23(14):3227–34.
55. Mishra G, Rahi S. Need of toxicity studies for cosmetic products and their approaches. biolsciences [Internet]. 2022 Jan. 17 [cited 2025 Mar. 3];2(1):105-9.
56. Gaurav Mishra, Amisha Thakur, Sukhbir Kaur. Process Validation of Metformin Hydrochloride Tablet (500 mg) according to USFDA. biolsciences [Internet]. 2021 Apr. 7 [cited 2025 Mar. 3];1(1):7-18.