International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Exploring Protein Biomarkers in Cancer: A Bioinformatics Approach

Pinki[1], Rita Singh[2]

[1]Research Scholar, Division of Molecular Endocrinology and Reproduction, University of Delhi, New Delhi - 110007
Corresponding Author: Email: *phogat.pinky[at]gmail.com*

[2]Division of Molecular Endocrinology and Reproduction. Department of Zoology, University of Delhi, New Delhi 110007, India

Abstract: ***Cancer is a leading cause of deaths worldwide. There are many clinically available biomarkers for its diagnostics and therapeutics. In our study, we demonstrate bioinformatics as an important tool in evaluating the potential of protein biomarkers. By utilizing Expasy ProtParam tool, we identify key molecular characteristics, including half - life, molecular weight, isoelectric point (pI), etc. and SOPMA is utilized for examining the secondary structure composition. Moreover, the protein - protein interaction of each selected biomarker is investigated using STRING analysis to mark functional relevance of selected biomarkers. Our findings underscore the multifaceted capabilities of bioinformatics tools in evaluating diverse aspects of protein biomarkers.***

Keywords: Biomarkers, Bioinformatics, Expasy Protparam tool, SOPMA, STRING

1. Introduction

In 2023, cancer caused around 9.6 to 10 million deaths globally, averaging approximately 26, 300 daily fatalities. It is one of the major concerns in public health domain. Despite progress in medical research, we still need better ways to detect and treat it. Biomarkers, which are like molecular signs in the body, can help in this effort. Biomarkers can be physical, chemical, or biological agents that are accessible in body matrices and can be measured in body fluids or cells. [1] The importance of biomarkers lies in their ability to provide early detection and diagnosis of diseases, monitor treatment efficacy, and predict disease outcomes. There are different categories of biomarkers, based on biomolecules (genetic, transcriptomic, proteomic, metabolomic and epigenetic), based on diseased state (diagnostic, prognostic, predictive, monitoring and pharmacodynamic/ response) and based on other criterion like imaging and pathology. [1 - 3] Of all these biomarkers, protein aka proteomic markers are particularly vital. They serve as insightful indicators revealing the alterations that build up in the tumor microenvironment formed at the onset of cancer. [4] By studying these protein markers, scientists and doctors can unravel the unique characteristics of cancer and develop targeted strategies to detect, treat, and monitor the disease.

In our study, we're diving into the world of these protein biomarkers using advanced computer tools called bioinformatics. We are focusing on known cancer biomarker proteins, checking out their physicochemical properties, secondary structure composition and their interacting partners. Our goal is not just to learn more about cancer biomarkers at a micro level but also to show how these computer tools can be helpful in studying all kinds of cancers. In the world of medical research, bioinformatics is useful in uncovering crucial information about identified proteins, providing a swift and insightful avenue for analysis. [5] By retrieving the protein sequence of target proteins, we can leverage a series of bioinformatics tools to gain information. For instance, the Expasy Protparam tool aids in deducing the physicochemical properties of proteins of interest, offering details about their molecular weight, Isoelectric point, hydropathy, half - life etc. [6] Tools like SOPMA and PSIPRED contribute by predicting the secondary structure of proteins, shedding light on the proportion of alpha helix, extended strand, beta turn, or random coil structures. [7] Further, the Net Phos, NetPhosK, PHOSITE, Predikin 1.0, DISPHOS, PredPhospho software etc enables the prediction of various phosphorylation sites, providing insights into the catalytic activity of the proteins. [8] The STRING software facilitates the study of protein - protein interactions, while other bioinformatics tools allow the prediction of 3D structures. [9] Overall, bioinformatics emerges as a helpful and efficient approach, swiftly revealing essential information about proteins, from their physical properties to intricate molecular interactions, enhancing our understanding of these critical biological entities.

In this study, we've focused on well - known cancer biomarkers, p53 (Tumor suppressor protein 53), BRCA1 (Breast Cancer Antigen 1) and carcinoembryonic antigen (CEA). [10 - 12] These biomarkers are clinically used for cancer diagnosis. Our interest lies in understanding structural and functional properties of these biomarkers. To do this, we've used various bioinformatics tools to predict and compare their properties with normal proteins found in the human body. We have chosen Beta actin as the control protein for the analysis. By comparing these biomarkers with normal protein, we aim to highlight the importance of bioinformatics in predicting their oncogenic potential.

2. Methods

Retrieval of protein sequences

Four protein sequences viz. P53, BRCA1, CEA and Beta Actin of *Homo Sapiens* were downloaded from UniProt (http: //www.uniprot. org) [13] in FASTA format for further analysis.

Analysis of Physicochemical properties

Physico - chemical properties of P53, BRCA1, CEA and Beta Actin were computed using ExPASy ProtParam tool (https: //web. expasy. org/protparam/). [14]

Secondary Structure Prediction

Secondary structures of the proteins were predicted with online SOPMA (Self Optimized Prediction Method and Alignment. This tool provides details about different conformations of proteins from the given sequences such as percentages of α - helices, β - sheets, turns, extended strands and random coils.
(https: //npsa - prabi. ibcp. fr/cgibin/npsa_automat. pl?page=/NPSA/npsa_sopma. html). [15]

Protein - Protein Interaction

Functional partners of the proteins P53, BRCA1, CEA and Beta Actin were predicted using STRING (Search tool for the retrieval of interacting genes/proteins) tool. (https: //string - db. org/). [16]

3. Results

Analysis of Physicochemical properties

We utilized the Expasy Protparam tool to examine important parameters such as half - life, instability index, aliphatic index, and GRAVY index for all the proteins under investigation. Interestingly, among the proteins studied, beta - actin stood out as its instability index was less than 40, suggesting it to be a stable protein. In contrast, the cancer biomarkers were predicted to have unstable structures. These findings underscore the potential of bioinformatics tools in predicting potential biomarkers by evaluating key properties, making them invaluable tools in the study of biomarkers. The detailed results for these parameters are summarized in Table 1, providing a comprehensive overview of the distinctive characteristics of each protein in our study.

Secondary Structure Prediction

In the next phase of our study, we delved into the secondary structure prediction of the proteins using the SOPMA software. This allowed us to explore the percentage distribution of key secondary structures such as alpha helix, beta turn, extended strand, and random coil in the selected proteins. Our analysis revealed varying percentages of these structures in all the proteins. Notably, beta - actin exhibited the minimum percentage of random coil structure at 41.33%, while carcinoembryonic antigen, P53, and BRCA1 showed percentages of 62.68%, 67.94%, and 82.77%, respectively. These various structural patterns are shown in the pie charts provided in Figure 1. Random coils are disordered regions of the protein chain that do not form any specific secondary structure, such as alpha helices or beta sheets. This flexibility can be important for the protein's function, as it may allow it to undergo conformational changes or interact with multiple binding partners. This analysis provides important information about the distinct structural features of each protein, setting the foundation for a better understanding of their functional roles. Moreover, this highlights how bioinformatics tools, like SOPMA, contribute significantly to unraveling the intricate structural features of biomolecules.

Protein - Protein Interaction

We used the STRING analysis tool to explore the interacting partners of the proteins and understand their functions. The results indicated that, P53 protein is found interacting with most of the proteins which are involved in cell cycle regulation. BRCA1 protein was found to be involved in DNA repair mechanisms, CEA was involved with proteins required in cell - cell adhesion and epithelial to mesenchymal transition processes. All these proteins were directly and indirectly related with maintenance of cellular homeostasis while beta actin was found involved with proteins having role in the maintenance of cytoskeleton structures within the cells. These observations are important as by checking out the functions of these binding partners, we can figure out the specific roles of each protein. These interaction networks are like maps that guide us in understanding how cells work on a larger scale. The binding partners of all the proteins and their key functions are presented in figure 2. Each node represents a protein. The blue line represents empirically determined interactions, the pink line represents curated database interactions, and the green line represents text mining interactions.

4. Discussion

Bioinformatics is a pivotal branch of science which has the capability to elucidate the complexity of various biomolecules. In the case of proteins, it begins by exploring the fundamental characteristics, progressing to intricate details like structural features, functionalities, and partnerships. Understanding bioinformatics makes complicated problems simpler by making sense out of the biological data. In this study, the comparison of biomarkers with a normal protein, beta - actin, revealed significant distinctions, particularly the instability index, a critical parameter in determining protein stability. The observation that beta - actin exhibited stability below the threshold of 40, while other proteins surpassed this limit, suggests a potential correlation between instability and biomarker candidacy. Unstable proteins are relatively prone to degradation and thus could be present in bodily fluids, making them promising biomarker candidates. The prevalence of random coil structures in unstable proteins further underscores their flexibility, indicating increased availability for various interactions and degradation mechanisms. Additionally, predicting the functional importance of proteins becomes feasible by looking at their interacting partners. In essence, bioinformatics emerges as a powerful tool, not only in deciphering the complex nature of biomolecules but also in predicting their biomarker potential, offering a transformative approach in biological research.

5. Conclusions

Bioinformatics is versatile as it can reveal extensive insights into genes and proteins. Starting with decoding simple properties like physicochemical characteristics, it progresses to unveil secondary and tertiary structures, functional aspects, and binding partners. This study, comparing known biomarkers with a regular protein, identified distinctive properties. Bioinformatics extends its utility to molecular docking for checking target molecule binding with diverse proteins, playing a crucial role in drug design and delivery. In conclusion, bioinformatics emerges as a dynamic tool, not

only providing vital information about identified proteins but also predicting their biomarker potential, contributing significantly to advancements in various scientific domains.

References

[1] Ahmad A, Imran M, Ahsan H. Biomarkers as Biomedical Bioindicators: Approaches and Techniques for the Detection, Analysis, and Validation of Novel Biomarkers of Diseases. *Pharmaceutics*.2023; 15 (6): 1630 - 1666.
[2] Sarhadi VK and Armengol G. Molecular Biomarkers in Cancer. *Biomolecules*.2022; 12 (8): 1021 - 1060.
[3] Califf RM. Biomarker definitions and their applications. *Exp Biol Med (Maywood).*2018; 243 (3): 213 - 221.
[4] Alharbi RA. Proteomics approach and techniques in identification of reliable biomarkers for diseases. *Saudi J Biol Sci.*2020; 27 (3): 968 - 974.
[5] María, Hernández - Domínguez E, Sofía Castillo - Ortega L, García - Esquivel Y, Mandujano - González V, Díaz - Godínez G, Álvarez - Cervantes J. Bioinformatics as a Tool for the Structural and Evolutionary Analysis of Proteins. *Computational Biology and Chemistry*.2020; 1 - 17.
[6] Chakma V, Barman DN, Das SC, Hossain A, Momin MB et al. In silico analysis of a novel hypothetical protein (YP_498675.1) from Staphylococcus aureus unravels the protein of tryptophan synthase beta superfamily (Try - synth - beta_ II). *J Genet Eng Biotechnol.*2023; 21 (1): 135 - 149.
[7] Kaur A, Pati PK, Pati AM, Nagpal AK. Physico - chemical characterization and topological analysis of pathogenesis - related proteins from Arabidopsis thaliana and Oryza sativa using in - silico approaches. *PLoS One*.2020; 15 (9): 1 - 15.
[8] Paul P, Muthu M, Chilukuri Y, Haga SW, Chun S, Oh JW. In Silico Tools and Phosphoproteomic Software Exclusives. *Processes* 2019; 7: 869 - 886.
[9] Chen SJ, Liao DL, Chen CH, Wang TY, Chen KC. Construction and Analysis of Protein - Protein Interaction Network of Heroin Use Disorder. *Sci Rep*.2019; 9 (1): 4980 - 4989.
[10] Qiu J, Keyser B, Lin ZT, Wu T. Autoantibodies as Potential Biomarkers in Breast Cancer. *Biosensors (Basel).*2018; 8 (3): 67 - 84.
[11] Afzal S, Hassan M, Ullah S, Abbas H, Tawakkal F, Khan MA. Breast Cancer; Discovery of Novel Diagnostic Biomarkers, Drug Resistance, and Therapeutic Implications. *Front Mol Biosci*.2022; 9: 1 - 10.
[12] Pal M, Muinao T, Boruah HPD, Mahindroo N. Current advances in prognostic and diagnostic biomarkers for solid cancers: Detection techniques and future challenges. *Biomed Pharmacother.*2022; 146: 1 - 19.
[13] Adapted from http: //www.uniprot. org
[14] Adapted from https: //web. expasy. org/protparam/
[15] Combet C, Blanchet C, Geourjon C, and Deléage G, NPS[at]: Network Protein Sequence Analysis TIBS 2000; 25 (3): 147 - 150.
[16] https: //npsa - prabi. ibcp. fr/cgibin/npsa_automat. pl?page=/NPSA/npsa_sopma. html
[17] Adapted from https: //string - db. org

Table 1: Physicochemical Properties of known cancer biomarker proteins and Beta actin proteins as predicted by Expasy Protparam tool

S. No.	Characteristics	P53	BRCA1	CEA	Beta Actin
1	Half life	30 hours (in mammalian reticulocytes	30 hours (in mammalian reticulocytes	30 hours (in mammalian reticulocytes	30 hours (in mammalian reticulocytes
2	Sequence Length	393 amino acids	1863 amino acids	702 amino acids	375 amino acids
3	Molecular Weight	43686.21	207720.25	76786.75	41736.73
4	Extinction Coefficient	36035, abs 0.825	103940, abs 0.500	113970, abs 1.484	44725, abs 1.072
5	Isoelectric point	6.47	5.29	5.51	5.29
6	Instability Index	71.93	54.68	41.78	35.29
7	Aliphatic Index	59.34	69.01	84.29	81.95
8	GRAVY Index	-0.755	-0.785	-0.312	-0.2
9	- R	50	283	54	49
10	+R	47	213	42	37

Figure Legends:
Figure 1. Comparison of secondary structure features of A) P53, B) BRCA1, C) CEA and D) Beta Actin proteins.
Figure 2. A) Protein - protein interaction network of P53 protein, the major functions of the interacting proteins are labelled in the figure.
Figure 2. B) Protein - protein interaction network of BRCA1 protein, the major functions of the interacting proteins are labelled in the figure.
Figure 2. C) Protein - protein interaction network of CEA protein, the major functions of the interacting proteins are labelled in the figure.
Figure 2. D) Protein - protein interaction network of Beta Actin protein, the major functions of the interacting proteins are labelled in the figure.

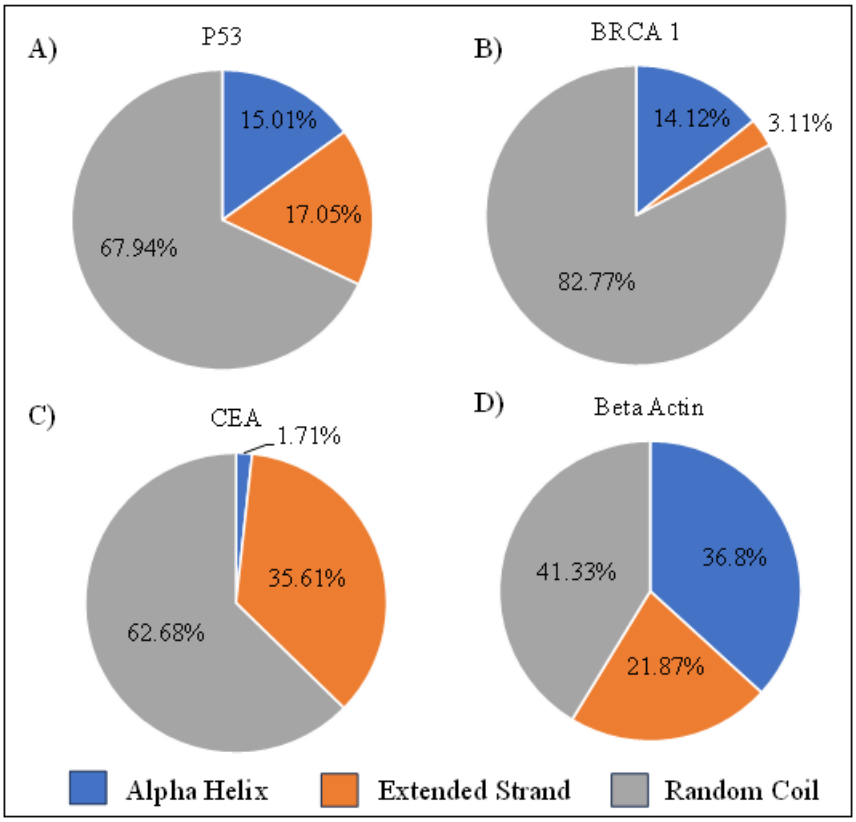

Figure 1: Comparison of secondary structure features of A) P53, B) BRCA1, C) CEA and D) Beta Actin proteins.

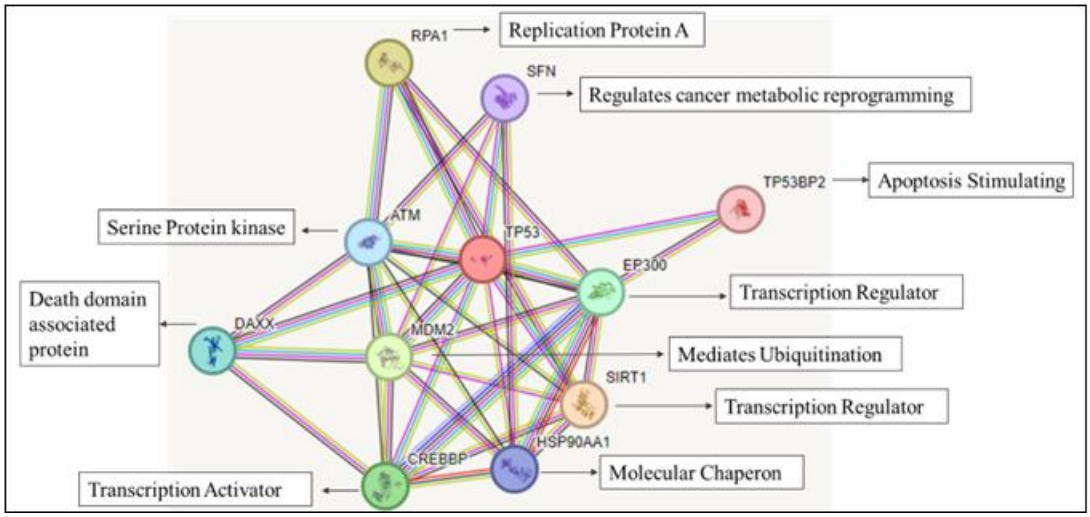

Figure 2: A) Protein - protein interaction network of P53 protein, the major functions of the interacting proteins are labelled in the figure.

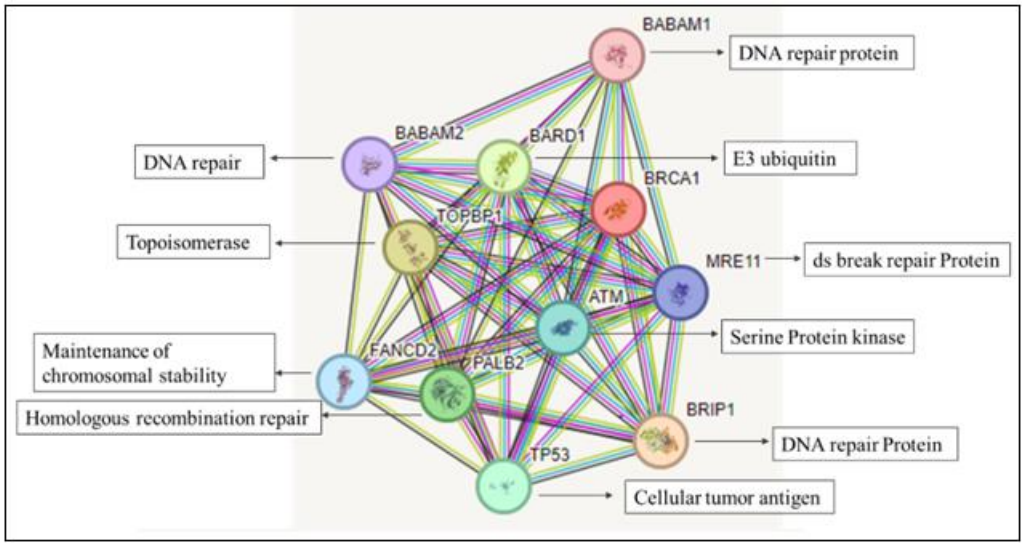

Figure 2: B) Protein - protein interaction network of BRCA1 protein, the major functions of the interacting proteins are labelled in the figure.

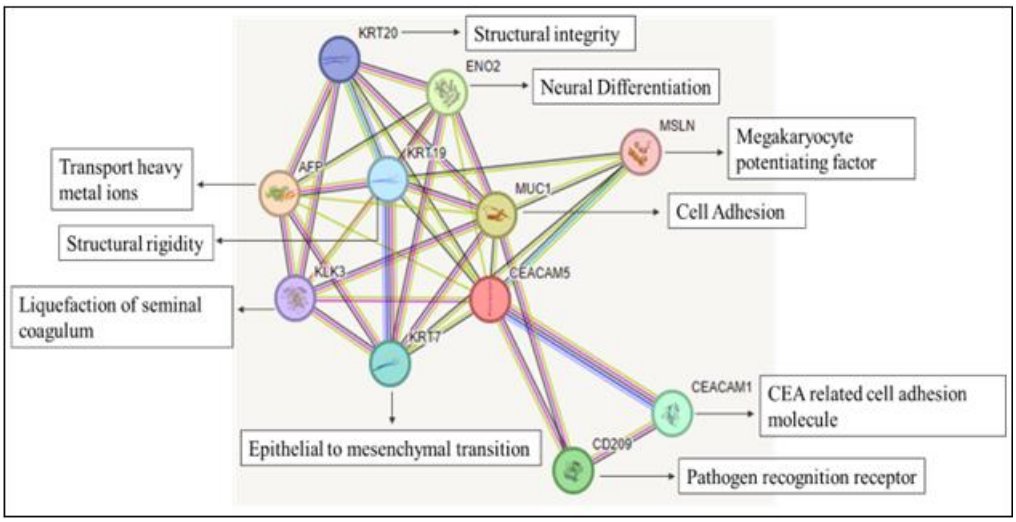

Figure 2: C) Protein - protein interaction network of CEA protein, the major functions of the interacting proteins are labelled in the figure

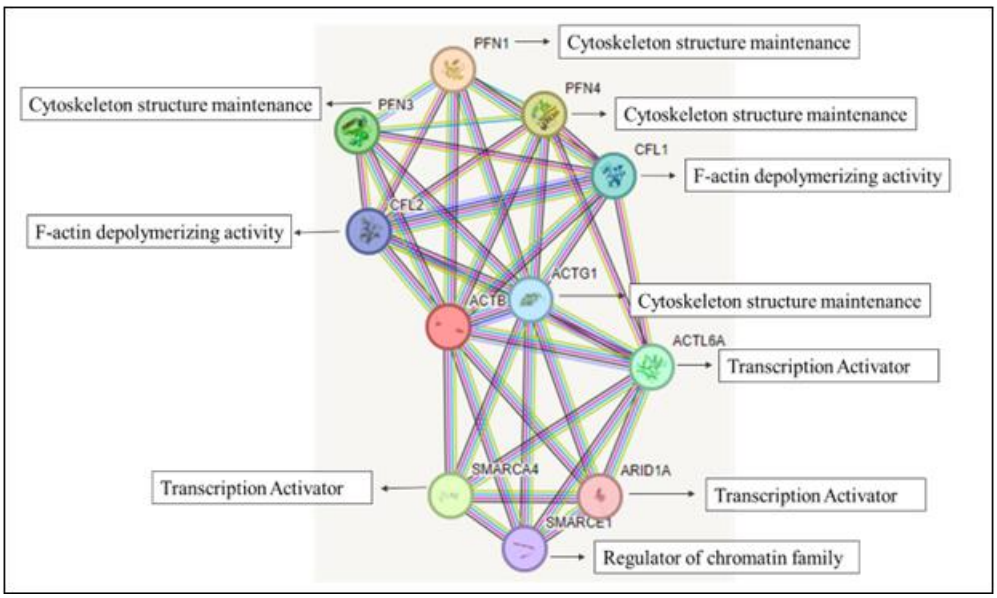

Figure 2: D) Protein - protein interaction network of Beta Actin protein, the major functions of the interacting proteins are labelled in the figure