In Silico Analysis of Glycoprotein Non-Metastatic Melanoma Protein B (GPNMB) Protein Detection in Congestive Heart Failure (CHF)

SAHASRA VEDAGIRI
Jordan High School, Fulshear, TX
Published November 2025

Congestive Heart Failure (CHF) occurs when the heart is unable to pump enough blood and oxygen to meet the body's needs. Glycoprotein Non-metastatic B (GPNMB) is a protein encoded by the GPNMB gene, playing a crucial role in cell and tissue survival and regulating inflammation. Recently, this protein has been shown to play a role in heart damage, contributing to tissue recovery and remodeling. In some cases, it has been linked to promoting fibrosis, disrupting electrical contractions, and the behavior of macrophages. Aptamers are single-stranded chains of nucleotides, either DNA or RNA, which are synthetically produced to bind to specific receptors in proteins and molecules. We hypothesize that aptamers binding to the surface of BPNMB could be used to aid in diagnosing heart failure. Based on the number of interactions, aptamer ss9 was selected as an appropriate candidate. Molecular docking simulations were employed to identify the most effective and optimal binding site between the aptamer and protein. This research aims to develop a solution that helps determine the extent of heart damage in patients, providing both the patient and physician with a more accurate understanding of their condition.

1. INTRODUCTION

Heart failure occurs when the heart cannot function effectively and can't pump enough blood to meet the body's needs.[1] Various factors, including sedentary lifestyles, genetics, smoking, alcohol, and illnesses, could cause heart failure. People who are most at risk are those with consistently poor lifestyles and those who don't get frequent exercise and eat a lot of excess triglycerides. The different types of heart failure are left-sided heart failure, right-sided heart failure, and congestive heart failure. Some common symptoms of heart failure are shortness of breath, fatigue, swelling (edema), irregular heartbeat, coughing, nocturia, and nausea. Heart failure is treated and managed by lifestyle changes, medications, devices, and surgical procedures. Some current detection methods include screening techniques such as cardiac MRI, ECGs, or EKGs, as well as clinical evaluation. The concern lies in many subtle inaccuracies. One of these is symptom overlap, when two or more illnesses share common symptoms. This could mislead diagnosis or manipulate the illness's seriousness, preventing early detection. Another problem is biomarker variability, which refers to the natural signs the body produces by releasing specific proteins in response to an illness. Additionally, imaging limitations pose a problem, as they may lead to the overlook of subtle abnormalities that carry significant implications for the person's wellness. It also wouldn't be able to capture factors such as diet, how the patient feels, and their stress levels. Adding on, transient conditions pose a threat as these are temporary symptoms or stages. This could mislead physicians during diagnosis and recovery. The temporary nature could cause assumptions to be made that could hinder detection with accuracy. Stress Test Challenges - These tests measure the heart's response to stress. They are mainly used to identify coronary artery disease. In the early stages of coronary artery disease, minor blockages or the production of one may not be detected, preventing early detection before it gets too serious. Another problem is invasive procedures, which can lead to more complications for the patient, both physically and mentally. These procedures often require the use of a needle or surgical incisions.

GPNMB (Glycoprotein Non-Metastatic B) protein is a transmembrane protein that is active in inflammation, wound healing, tissue remodeling, and immune modulation, all of which contribute to homeostasis. The protein is found in osteoblasts, osteoclasts, COPD, IPF, and neurons, which can be linked to Alzheimer's or Parkinson's disease, as well as in fibroblasts, macrophages, TAMs, epithelial cells, and other cell types. GPNMB is present in cardiomyocytes and is associated with vascular remodeling, atherosclerosis, and various other diseases due to its regulation of inflammation and tissue repair.

An aptamer is a synthetically produced single-stranded DNA or RNA molecule that can bind to proteins, molecules, ions, and cells by interacting with specific receptors.[2] This is deter-

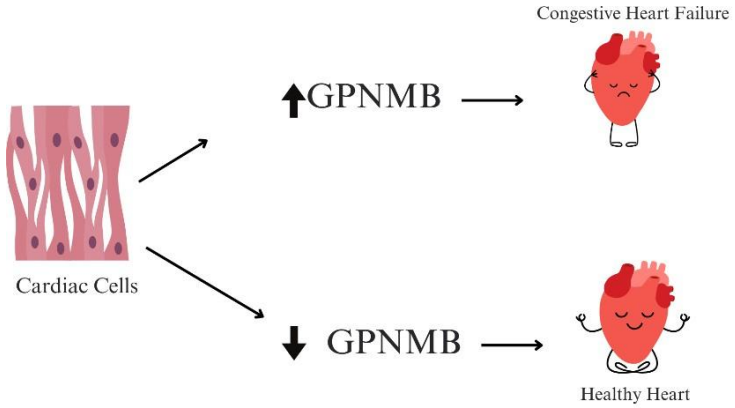

Fig. 1. GPNMB Levels as an Indicator of Cardiac Stress. Schematic showing how cardiac cells modulate GPNMB expression in response to physiological state. During cardiac distress or damage—such as in congestive heart failure—GPNMB is released from the cell surface into the bloodstream, leading to elevated circulating levels (↑GPNMB). In contrast, healthy cardiac tissue maintains low extracellular GPNMB levels (↓GPNMB). This differential expression highlights GPNMB as a potential biomarker for identifying cardiac dysfunction. (Images created using Canva.)

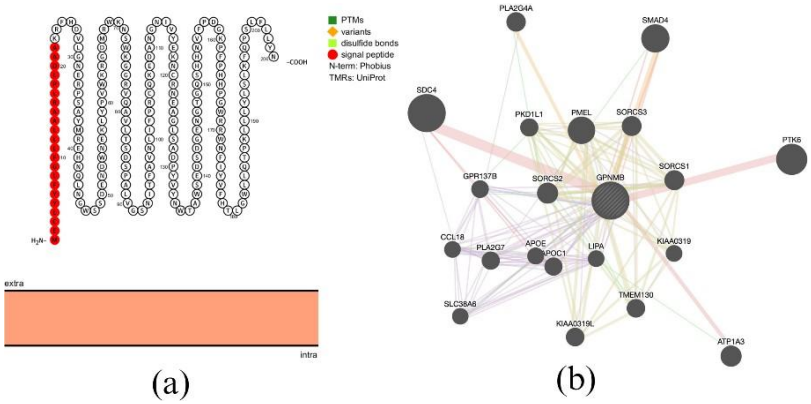

Fig. 2. Structural Features and Interaction Network of GPNMB. (a) Topological diagram of the GPNMB protein showing its transmembrane organization, extracellular and intracellular domains, and key annotated features such as post-translational modifications (PTMs), signal peptides, and sequence variants. The transmembrane region is highlighted, illustrating the orientation of the N- and C-termini relative to the membrane. (b) Protein–protein interaction (PPI) network centered on GPNMB, illustrating its functional associations with neighboring proteins involved in immune signaling, cell adhesion, lipid metabolism, and intracellular trafficking. Node size corresponds to interaction strength or network connectivity, and colored edges represent different types of predicted functional relationships. This network highlights GPNMB's central role in multiple cellular pathways relevant to disease biology.

mined by the sequence of base pairs and the three-dimensional structure that results when it folds.[3] Compared to antibodies, aptamers are more easily produced and modified and sustain harsher conditions. They are much tinier than antibodies, which helps them penetrate through tissue more efficiently, and their binding strength is equal. Aptamers are created and selected through the SELEX method, which stands for Systematic Evolution of Ligands by Exponential Enrichment.[4] First, in a controlled laboratory, many nucleic acid sequences are exposed to the target, GPNMB, and those that bind the strongest to it are selected. The selected ones undergo a polymerase Chain Reaction (PCR), which amplifies the aptamer sequences. The process is repeated until the most effective and strongest ones are picked.[3]

2. METHOD

Protein Modeling: The first step was to design a receptor protein, CD276. To obtain the protein's 3D structure, we used AlphaFold 3, a machine learning-based tool that predicts the 3D structures of proteins, DNA, and RNA.[5] The protein's amino acid sequence (Q5ZPR3) was retrieved from the UniProt database, which provides sequences for various proteins along with detailed information about each protein.[6]

Aptamer Modeling: The next step involved modeling the 3D structures of the aptamers ss2 and ss9. The sequences for these aptamers were obtained from previous research. While the base pair sequences were initially in 1D format, we needed to convert them into 3D structures for molecular docking. To achieve this, we used UNAfold to generate a ct file, which was then uploaded to the RNAstructure website by Rochester, yielding a dot and bracket sequence.[7,8] This conversion process transformed the 1D sequence into a 2D structure. To further convert the 2D structure into a 3D model, we utilized the FarFar2 tool within the ROSIE web server, ultimately obtaining a 3D model suitable for molecular docking simulations.[9]

Molecular Docking Simulations and Analysis: We performed protein-aptamer docking using the HDOCK Web Server. We analyzed the binding energy using PRED-PDA to validate the docking results and assessed the interaction sites through the Protein-Ligand Interaction Profiler (PLIP).[10] The binding energy provides insight into the strength of interactions between the protein and aptamer, significantly influencing their binding affinity. Key binding residues were identified using the PLIP Web Server, and binding site accessibility was analyzed through ScanNet. Finally, to visualize the docked complex, we employed ChimeraX.[11] To further validate the docking poses, we used P2Rank for hotspot prediction.[12]

3. RESULTS:

This study employed molecular docking simulations to identify aptamer ss9 as a promising candidate for detecting GPNMB, a biomarker for congestive heart failure, enabling early and non-invasive diagnosis.

Molecular docking: To formulate and build our innovation, we utilized computational techniques such as UnaFold and FARFAR2 for aptamer modeling and the hdock web server to understand aptamer and protein interactions. Throughout the different stages, our innovation was being explored further as we visually saw the aptamers, protein, and them binding together, which instilled more apparent connections for us to draw. First, the aptamer's 3D structure was validated using an inbuilt validation technique in FARFAR 2. In the next step, the aptamer binding interactions were selected based on the ScanNet web server, which is a deep learning-based method. After we sent it through Chimera X and received a visual representation of the protein and aptamer as seen in Figure 3.

Figure 5 illustrates different aspects of protein structure and dynamics analysis. Panel (a) displays the three-dimensional structure of a protein, where the chain is colored from blue to red based on the residue index. The arrows indicate the direction of major conformational movements, possibly derived from principal component analysis (PCA) or normal mode analysis, highlighting regions undergoing significant structural rearrangements. Panel (b) presents a dynamic cross-correlation matrix (DCCM), where each point represents the correlation in motion between a pair of residues during a molecular dynamics simulation. Red regions indicate highly positively correlated motions (residues moving in the same direction), blue regions show anti-correlated motions (residues moving in opposite directions), and white regions represent uncorrelated motions. This analysis helps identify coordinated domain movements and flexible regions within the protein. Panel (c) shows a contact map, where each gray dot denotes a pair of atoms or residues within a predefined distance threshold, typically used to represent stable interactions. The diagonal line indicates self-contacts or close neighbors along the sequence. This contact pattern provides insights into the overall fold and tertiary structure of the protein.

4. RNA SECONDARY STRUCTURES

RNA Secondary Structures: The table presents RNA sequences along with their corresponding secondary structures. Each sequence is identified by a unique Sequence ID and its nucleotide sequence. The secondary structure representationutilizes dot-bracket notation, where parentheses (') indicate base pairs and dots (') indicate unpaired nucleotides. This table facilitates understanding of the structural folding patterns of RNA sequences.

5. DISCUSSION

The following are the applications of our research: (a) Early Congestive Heart Failure Detection: Aptamer-based detection of GPNMB will serve as a non-invasive method to identify biomarkers for early congestive heart failure diagnosis, which would, in turn, decrease the risk and enable more promising outcomes with increased effectiveness of treatments. (b) Monitoring Treatment Response and Effectiveness: Tracking GPNMB levels can help identify and characterize the progression of congestive heart failure, which reflects treatment effectiveness. Understanding how counterintuitive the treatment is would help capture and create personalized plans that suit the biological response and physical needs of the patients. This would ensure that the right action is taken at the right time, without worsening the patient's condition.

6. LIMITATIONS:

This project's limitations include the lack of experimental validation. Computational findings need to be confirmed through laboratory experiments. Computational simulations do not account for real-world issues such as limited aptamer availability.

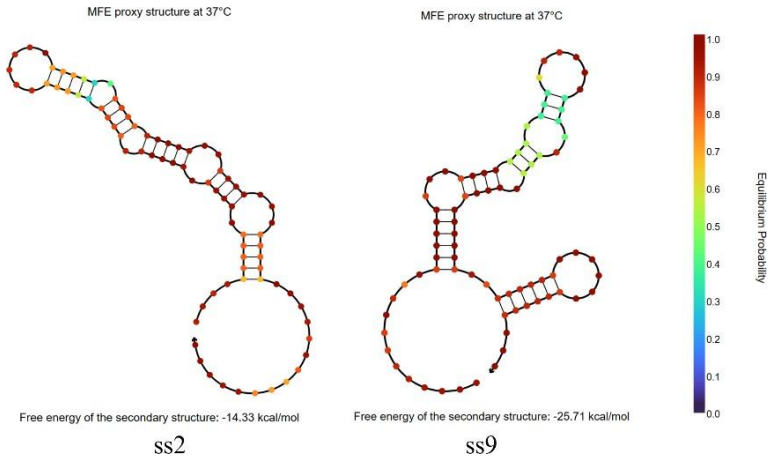

Fig. 3. Predicted minimum free energy (MFE) secondary structures of RNA sequences ss2 and ss9 at 37°C. The structures are color-coded based on base pair equilibrium probabilities, with red indicating high probability and blue indicating low probability. The free energy of the ss2 structure is -14.33 kcal/mol, while ss9 shows a more stable structure with a free energy of -25.71 kcal/mol.

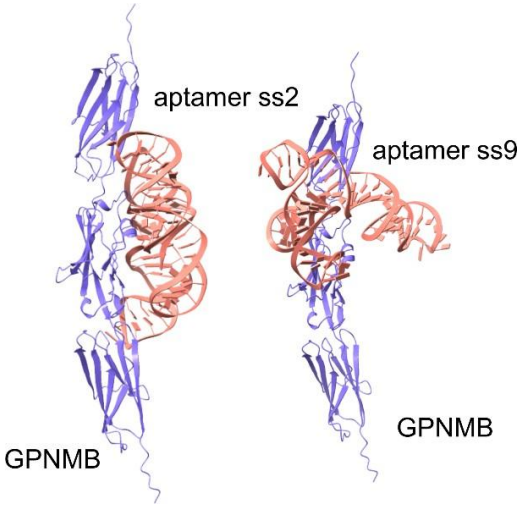

Fig. 4. Comparative Docking of Aptamers ss2 and ss9 to the GPNMB Receptor. Structural models illustrate the binding interactions between two candidate RNA aptamers—ss2 (left) and ss9 (right)—and the GPNMB protein (blue). The aptamer structures (shown in orange) adopt distinct three-dimensional conformations upon docking, engaging different surface regions of GPNMB. These variations in binding pose and orientation highlight potential differences in affinity and specificity, supporting the selection of optimal aptamers for targeting GPNMB in diagnostic or therapeutic applications.

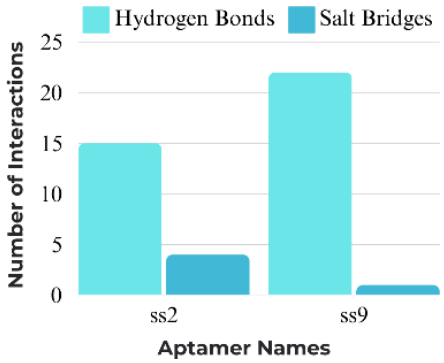

Fig. 5. Number of interactions formed between protein, GPNMB, and aptamers ss9 and ss2 computed using the PLIP web-server. The Bar Graph was generated using Canva. The number of interactions was further calculated by adding the number of slat bridges and hydrogen bonds created. The higher summation was concluded to be the more viable aptamer.

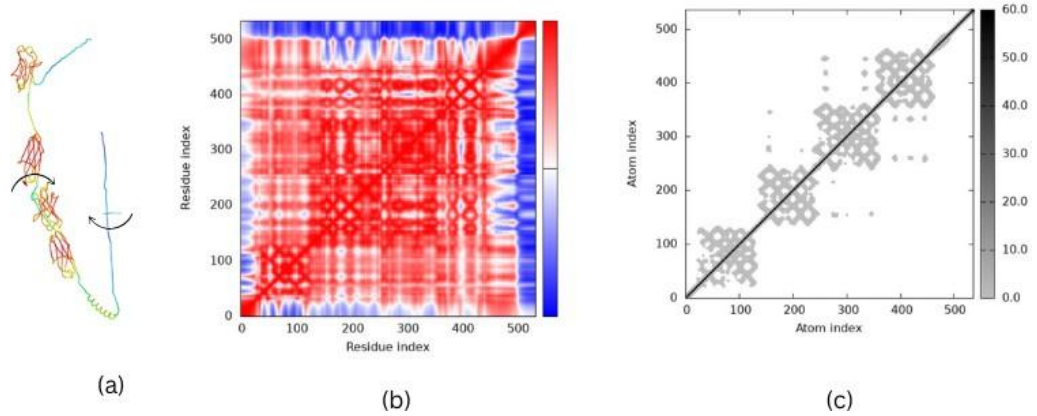

Fig. 6. Structural and contact map analysis of a protein or macromolecular complex. (a) 3D representation of the molecular structure showing the domain organization and folding pattern, with arrows indicating potential hinge or flexible regions. (b) Residue-residue dynamic cross-correlation matrix (DCCM), where red indicates high positive correlation, blue indicates negative correlation, and white denotes no correlation. This matrix provides insight into coordinated movements between residue pairs. (c) An atomic contact map represents distances between atom pairs, with darker shades indicating closer proximity. This map helps identify structural domains and interaction hotspots within the molecule.

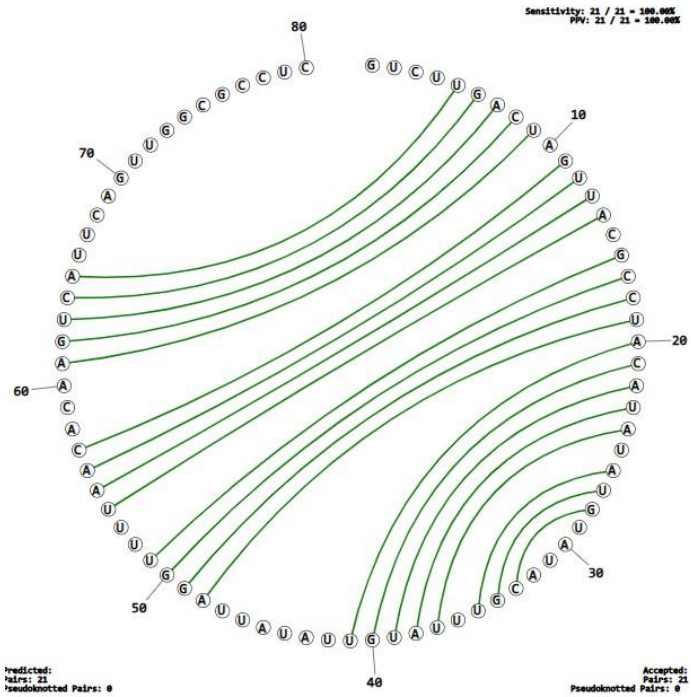

Fig. 7. Circular Representation of the Predicted RNA Secondary Structure. Circular plot showing the RNA aptamer's predicted secondary structure, with nucleotides arranged around the perimeter and Watson–Crick base-paired regions represented by green connecting arcs. The structure contains 21 predicted base pairs with no pseudoknots identified. Sensitivity and positive predictive value (PPV) both reached 100%, indicating that all predicted base pairs were accepted during validation. This visualization highlights the global folding pattern and long-range interactions within the aptamer.

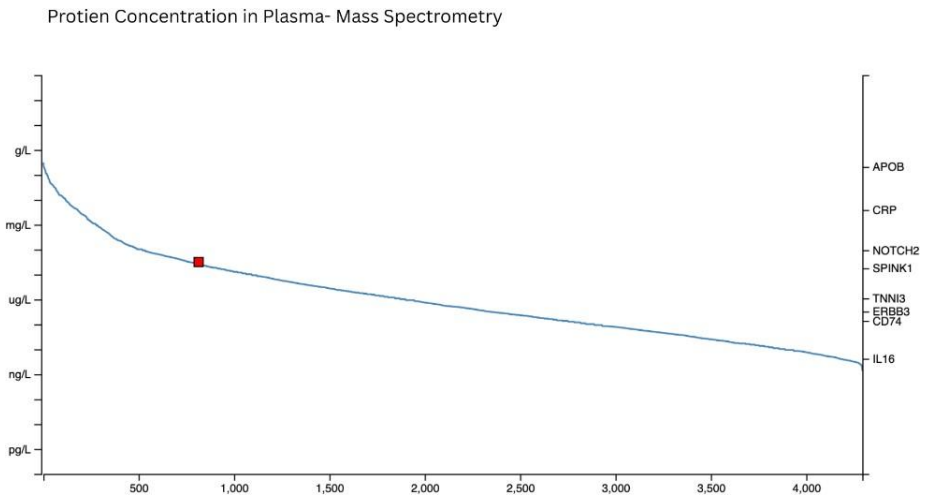

Fig. 8. This plot illustrates the wide dynamic range of protein concentrations detected in human plasma using mass spectrometry, spanning from grams per liter (g/L) for highly abundant proteins such as APOB to picograms per liter (pg/L) for low-abundance signaling molecules such as IL-16. Each labeled protein on the right represents a marker of clinical or biological significance, positioned according to its relative abundance. The red square denotes the approximate concentration range of the protein of interest within this continuum, emphasizing the challenge of detecting low-abundance biomarkers in complex plasma samples.

Table 1. RNA Aptamer Sequences and Predicted Secondary Structures.

Sequence ID	RNA Sequence	Secondary Structure (Dot–Bracket)
sS2	GUCUUGACUAGUUACGCCUACAUAUAUGUAUACGUUUAUG UUAUAUUAGGUUUUAACACAAGUCAUUCAGUUGGCGCCUC	((....))....(((..((((....))))..)))...(((((.....)))))... .....
sS9	GUCUUGACUAGUUACGCCGUAGCACCGGGGUAUCCGUUUA UGGCCGAGUGUGGCGUAUUGUGUCAUUCAGUUGGCGCCUC	((....))....((.(((.....))).))..(((....)))........(((((..... .....)))))

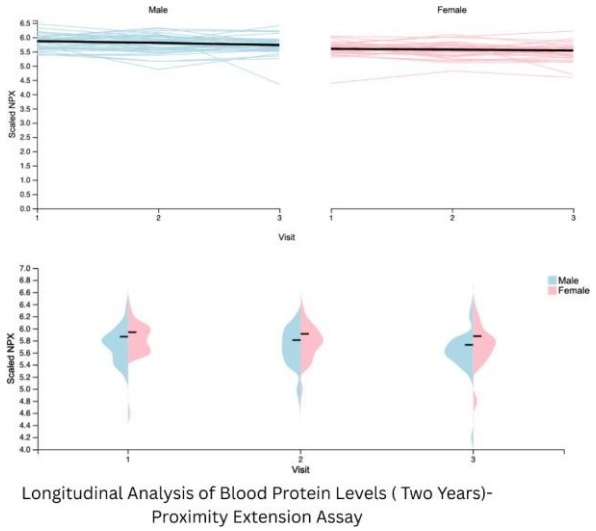

Fig. 9. Longitudinal Changes in Blood Protein Levels Over Three Clinical Visits Assessed by Proximity Extension Assay (PEA). Line plots (top) display individual trajectories of scaled NPX protein concentrations in male (blue) and female (pink) participants across three visits spanning two years, with bold lines indicating group means. Violin plots (bottom) illustrate the distribution and central tendency of NPX values at each visit for both sexes. Overall, protein levels remain broadly stable over time, with minimal sex-specific variation, suggesting consistent longitudinal expression patterns in the cohort.

A limited number of aptamers would affect their abilities, specificity, affinity, and sensitivity.

Practical Applications:

A test could be developed using early-detection technology to bridge the gap between application and theory. The aptamer and the blood sample, which have been centrifuged, are placed at the two ends. The centrifugation speed should be 15,000-20,000 RPM, which requires an ultracentrifuge. If the blood sample contains the biomarker or protein, it would be located where it is placed. With the help of the hyper-absorbent paper found in the detection case, they would be absorbed into the middle if present. The bonding would leave a visible mark due to the presence of the fluorescent dye. This process is best suited for a laboratory setting.

7. CONCLUSION

One of the three key conclusions of our research is the effective binding of aptamers. Molecular docking analysis shows that the aptamer ss9 exhibits strong binding interactions with GPNMB, making it a promising candidate for detecting congestive heart failure. Additionally, it facilitates the potential for early diagnosis. Aptamer-based detection of GPNMB could serve as a non-invasive method for identifying released biomarkers in congestive heart failure, thereby improving patient outcomes through timely intervention. Lastly, the third conclusion is the establishment of a computational framework. This study presents a computational pipeline for predicting and analyzing aptamer-protein interactions, facilitating future development and experimentation.

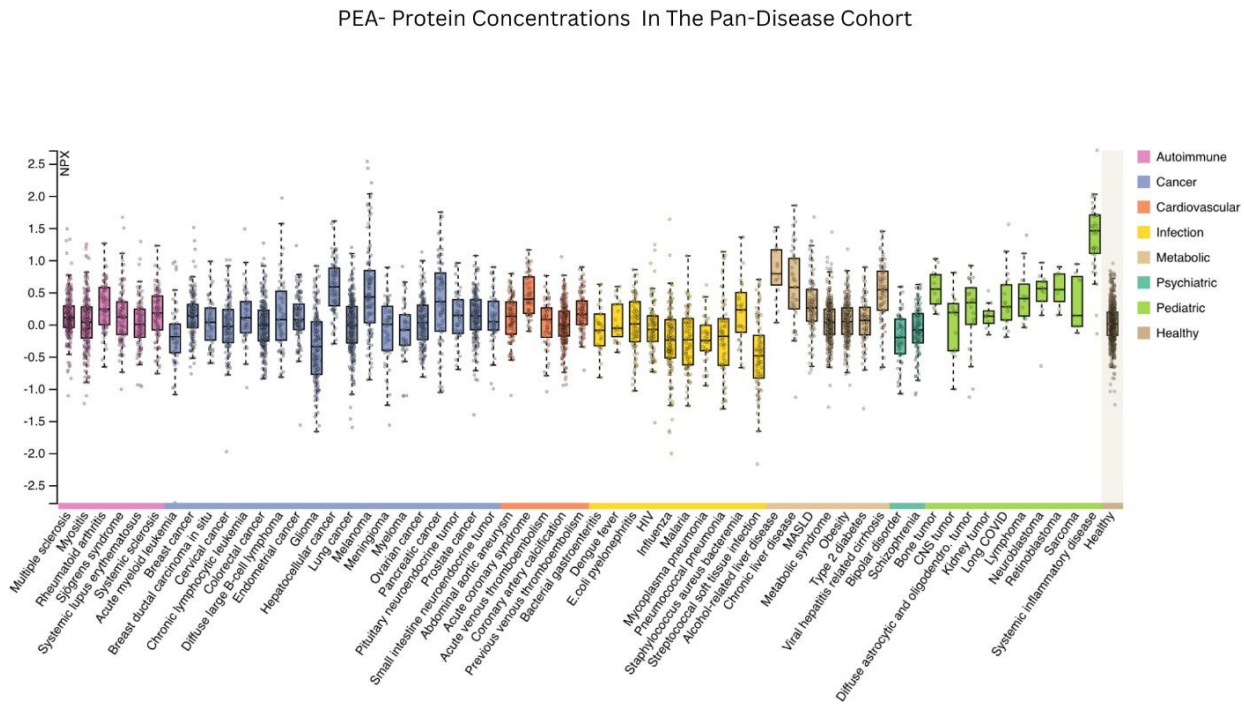

Fig. 10. Protein Expression Profiles Across the Pan-Disease Cohort. Box-and-whisker plots show the distribution of normalized protein expression (NPX) values measured by proximity extension assay (PEA) across a wide range of disease groups, including autoimmune, cancer, cardiovascular, infectious, metabolic, psychiatric, and pediatric disorders, as well as healthy controls. Each disease category is color-coded, allowing visualization of inter- and intra-group variability. The figure highlights distinct protein expression signatures across clinical phenotypes, demonstrating both disease-specific trends and overall heterogeneity within the cohort.

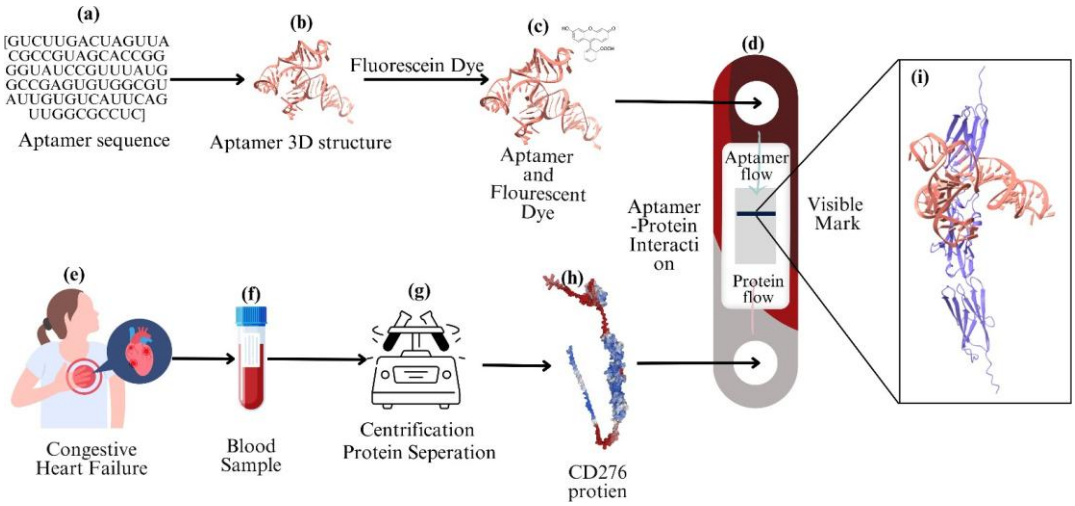

Fig. 11. A schematic representation of GPNMB protein-mediated detection of congestive heart failure. (a-b) The aptamer se- quence is converted into a 3D structure. (c) A fluorescent dye is bonded to the aptamer. (d) The conjugated aptamer is placed on the detection cassette. (e-f) Blood from a suspected congestive heart failure patient is retrieved. (g-h) If present, GPNMB is extracted using centrifugation and placed on the detection cassette. (I) GPNMB and the aptamer would meet in the middle of the detection cassette, and if both are present, they would bond, creating a visible mark.

8. REFERENCES

(1) Groenewegen, A.; Rutten, F. H.; Mosterd, A.; Hoes, A. W. Epidemiology of heart failure. European Journal of Heart Failure 2020, 22 (8), 1342-1356.
(2) Dunn, M. R.; Jimenez, R. M.; Chaput, J. C. Analysis of aptamer discovery and technology. Nature Reviews Chemistry 2017, 1 (10), 0076.
(3) Liu, Y.; Qian, X.; Ran, C.; Li, L.; Fu, T.; Su, D.; Xie, S.; Tan, W. Aptamer-Based Targeted Protein Degradation. ACS Nano 2023, 17 (7), 6150-6164.
(4) Darmostuk, M.; Rimpelova, S.; Gbelcova, H.; Ruml, T. Current approaches in SELEX: An update to aptamer selection technology. Biotechnology Advances 2015, 33 (6, Part 2), 1141-1161.
(5) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 2024, 630 (8016), 493-500.
(6) Consortium, T. U. UniProt: a hub for protein information. Nucleic Acids Research 2014, 43 (D1), D204-D212.
(7) Markham, N. R.; Zuker, M. UNAFold. In Bioinformatics: Structure, Function and Applications, Keith, J. M. Ed.; Humana Press, 2008; pp 3-31.
(8) Reuter, J. S.; Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 2010, 11 (1), 129.
(9) Watkins, A. M.; Das, R. RNA 3D Modeling with FARFAR2, Online. In RNA Structure Prediction, Kawaguchi, R. K., Iwakiri, J. Eds.; Springer US, 2023; pp 233-249.
(10) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: fully automated protein–ligand interaction profiler. Nucleic Acids Research 2015, 43 (W1), W443-W447.
(11) Meng, C. E.; Goddard, D. T.; Pettersen, F. E.; Couch, S. G.; Pearson, J. Z.; Morris, H. J.; Ferrin, E. T. <scp>UCSF ChimeraX</scp>: Tools for structure building and analysis. Protein Science 2023, 32 (11).
(12) Krivák, R.; Hoksza, D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. Journal of Cheminformatics 2018, 10 (1), 39.