IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Bioinformatics Breakthroughs in Thalassemia: Identifying DDX3 and Potential Drug Leads

Samaya Vaidya[1], Aneesha Vora[1], Nirupma Singh[2]

1) Cathedral and John Connon School, Mumbai, Maharashtra, India; adhithi.arun1304@gmail.com
2) Department of Biological Sciences and Engineering, Netaji Subhas University of Technology (Formerly NSIT), Dwarka, Delhi, India; nirupmajadaun21@gmail.com

ABSTRACT: Thalassemia is a genetic blood disorder that reduces hemoglobin production, leading to oxygen deprivation and severe health complications. Despite the high prevalence of this disease, especially in regions like India, research on potential therapeutic targets has largely focused on the HBB gene, limiting the discovery of effective treatment options. The lack of awareness about this disease remains a major impediment to its early diagnosis and prevention. At present, there are few studies done to discover possible proteins for therapeutic drugs to target, but they have not considered a wide range of genes. We hypothesized that the use of network analysis, pathway analysis, and virtual screening could reveal significantly differentially expressed genes (DEGs) and identify novel therapeutic targets for thalassemia treatment. Five differentially expressed genes were found to gain the most significant positions through network analysis: PRKY, EIF1AY, DDX3Y, CDY2B, and BPY2. Pathway analysis results showed that DDX3 was the top gene present in the top significant pathways that were overrepresented. Hence, DDX3 was found to be an emerging therapeutic target. The binding affinity of different ligands with this DDX3 was computed through virtual screening of large libraries of compounds. Neotetrazolium was found to have the highest binding affinity. Thus, this study contributes to the growing research on therapeutic targets and their counterdrugs for thalassemia and forms the foundation for further experimental validation and development of protein-ligand-based treatments.

KEYWORDS: Computational Biology and Bioinformatics, Proteomics, Thalassemia, Protein-protein interactions, Network and Pathway Analysis, Virtual Screening.

Introduction

Thalassemia is an inherited blood disorder that causes the body to have a lower amount of hemoglobin than what is considered normal due to defects in the synthesis of one or more of the hemoglobin chains.[1, 2] Haemoglobin is a protein found in red cells that carries oxygen from the lungs to all other organs in the body.[3] The body needs it to be able to function. Thalassemia is classified as a trait, minor, intermediate, or major, to describe the severity of the condition.[4] Thalassemia is most commonly found to affect individuals originating from the Mediterranean area, the Middle East, Transcaucasia, Central Asia, North Africa, the Indian subcontinent, and Southeast Asia.[5] India bears a huge burden of hemoglobinopathies, the most prevalent one being thalassemia. Approximately 100,000 Indians are affected by this disease.[6] Thalassemia requires lifelong care, and it was established that children affected with thalassemia need regular blood transfusions, at least twice a month, throughout their lives. The lack of awareness about this disease remains a major impediment to its early diagnosis and prevention.[7]

Thalassemia-affected individuals lack healthy red blood cells, causing a range of problems. When left untreated, severe forms of this disease can cause a change in daily activities and often threaten lives.[8] Thalassemia makes the bone marrow expand, which causes the bones to widen. This can result in abnormal bone structure, especially in the face and skull. Bone marrow expansion also makes bones thin and brittle, increasing the chance of broken bones. Additionally, congestive heart failure and abnormal heart rhythms can be associated with severe thalassemia.[2] Thromboembolic events occur when a thrombus forms and subsequently dislodges. In thalassemia, a common issue called splenectomy can further elevate this risk by increasing circulating platelets, leading to a higher incidence of thrombosis in diseased individuals. Thus, there is a dire need for additional research and early preventative measures to prevent thromboembolic events in thalassemia patients.

Some previous research studies have delved into gene mutations in relation to symptoms of beta-thalassemia. The effects of insertion mutations in the HBB exons have been investigated using the *in silico* approach by using the HbVar database to select sequences with uncharacterized insertion mutations and studying their effects on the structure and function of β-globin protein.[9] In one of the studies, PRMT5 was studied and was found to be a promising target for β-thalassemia, followed by molecular docking to identify its inhibitors.[10] However, none of the previous studies have been found to test a range of genes and proteins.

This study hypothesizes that a broader analysis of gene interactions through differential expression analysis, network analysis, and pathway analysis can identify novel therapeutic targets, which can then be validated for druggability using virtual screening. Network and pathway analysis are sets of widely used tools for research in life sciences intended to give meaning to high-throughput biological data. It is important for

DOI: 10.36838/v7i8.9

diseases such as thalassemia, as the tools analyze data obtained from these high-throughput technologies and detect relevant groups of related genes that are altered in the case of samples in comparison to a control, also known as differentially expressed genes.[11]

We hypothesized that a wider and more in-depth analysis of gene interactions would help discover unique targets that can effectively be used to develop therapeutic drugs for thalassemia treatment. The hypothesis is tested through a three-level validation framework, including DEG analysis, pathway enrichment, and molecular docking. This project introduces unconventional aspects for the study of therapeutic targets for thalassemia. In most cases, studies focus particularly on the HBB gene, whereas this study diverges from the traditional focus on one specific gene. Therefore, the implementation of the novel and innovative framework, which integrates gene expression data with molecular docking, offered fresh insights into therapeutic development for thalassemia. From the study, DDX3 was discovered as the most suitable druggable target for thalassemia, and Neotetrazolium was the best inhibitor, with a binding affinity of -10.2. (ChEMBL ID: 1183691). DDX3Y is located on the Y chromosome and therefore specifically expressed in males; however, DDX3X, its homolog on the X chromosome, may have overlapping functions relevant to both sexes in thalassemia. Thus, this study uses an innovative approach, a combination of differential expression analysis, network, and pathway analysis to identify the significant genes and pathways, and molecular docking to find the best inhibitors, broadening the scope of finding therapeutic drug targets with three-level validation.

Methods

This section presents the hypothesis testing framework of this study in the following sections:

Data Collection:

Gene Expression Omnibus (GEO)[12] was used to gather RNAseq data using different keywords related to thalassemia, such as "Thalassemia," "Beta-thalassemia," and "Haemoglobinopathy." The sample with GSE ID: GSE96060 was chosen from the GEO database to perform differential gene expression analysis of thalassemia and non-thalassemic individuals. The RNASeq dataset GSE96060 comprised samples from individuals diagnosed with β-thalassemia major and healthy controls. Differential gene expression analysis was performed comparing 4 thalassemia samples and 4 non-thalassemic controls. The GEO2R tool was used to identify differentially expressed genes (DEGs) after assigning the 'test' and 'control' groups to each category of samples. The top significantly differentially expressed genes were then downloaded. A library of 12206 drug compounds was downloaded from ChEMBL. The ChEMBL database was selected due to its extensive coverage of bioactive drug-like small molecules, and it has been widely cited in various drug discovery studies. Under 'Small Molecules,' Phase 1, Phase 2, Phase 3, Early Phase 1, and Approved compounds were downloaded. Another library of 21931 compounds was downloaded from the Therapeutic Target Database on PubChem. Similarly, PubChem offers a comprehensive library of small molecules and their biological activities, making it a valuable resource for identifying therapeutic compounds. The RCSB PDB database[13] was used to download the 3D structure of the identified therapeutic target protein.

Network Analysis:

The significant DEGs were then put into STRING (Version 12.0), and protein-protein interaction networks were made by including 5, 10, 20, and 50 interactors subsequently, and 4 networks were downloaded. STRING is a database of known and predicted protein-protein interactions. The networks were visualized using Cytoscape Version 3.10.2 [14] and then analyzed using the AnalyzeNetwork app of Cytoscape. Measures like degree, betweenness centrality, and clustering coefficient were considered to identify the hub genes of the network. The top 25 genes from all 4 networks were identified. The top 10 genes common to all 4 networks and those that were significantly differentially expressed were shortlisted.

Pathway Analysis:

The gene list of the most exhaustive and well-connected network (with 50 added interactors) was put into the Reactome Pathway Database Version 89 [15] to identify the significantly over-represented pathways. The significant pathway results without adding any interactors were downloaded. The top genes of the network were searched for their presence in the significant pathways. The overrepresented and compromised pathways in the case of a patient with Thalassemia were identified from this analysis, which showcased the involvement of top network genes.

Virtual Screening:

In the next phase of the study, from the ChEMBL database, 'Small Molecules' (Phase 1, Phase 2, Phase 3, Early Phase 1, Approved compounds) were downloaded. From PubChem, the 'Therapeutic Target Database' library was downloaded. DDX3 was prepared for docking using Autodock.[16] The libraries of ligands were virtually screened against DDX3 as the target protein, using PyRx software. [17] The AutoDock Vina scoring function was used to compute binding affinities. The binding affinity of each ligand was computed, and the top compounds from both libraries were identified. BIOVIA Discovery Studio was used to visualize the interaction maps of the top 3 ligands, each from TTD and ChEMBL libraries, with DDX3.

Result and Discussion

This section presents the results obtained from the bioinformatics analysis of differentially expressed genes, their pathways, and network interactions, providing a wider and more in-depth analysis of gene interactions in Thalassemia and testing the hypothesis. The analysis helped discover unique targets with a focus on understanding the potential drugs associated with therapeutic targets of Thalassemia.

Differential gene expression analysis:

After identifying the RNASeq dataset GSE96060, retrieved from the GEO, 20,047 differentially expressed genes were retained for analysis. A p-value threshold of <0.05 and a fold-change of ±2 were applied in GEO2R. The expression of DEGs was visualized in the form of a volcano plot (Figure 1). In total, 23 genes were found to have significant differential gene expression with a p-value <0.05 and a fold change of ±2 and were considered statistically significant. The list of these genes was used for further analysis.

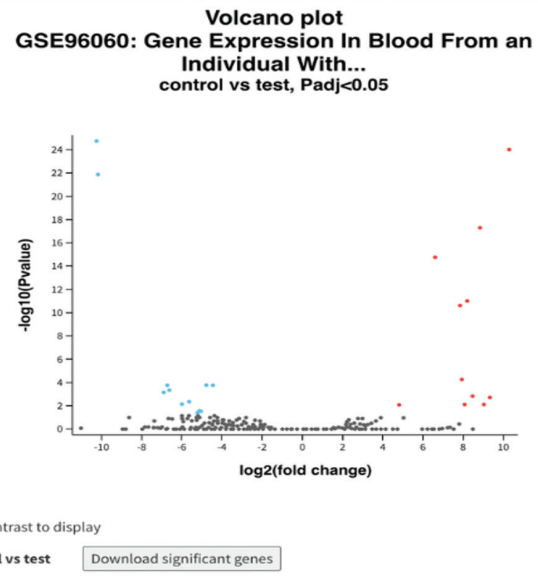

Figure 1: The volcano plot of differentially expressed genes in Thalassemia. This plot shows the gene expression in blood from control and test samples with Thalassemia, identified using the GEO2R tool. This helped in identifying the significant genes that were differentially expressed between normal individuals and thalassemia patients.

Network and Pathway Analysis:

Network study helped in visualizing the interaction maps of significantly differentially expressed genes of all 4 networks for 5, 10, 20, and 50 interactors (Figure 2). The network with 50 added interactors showed 4 clusters of connected components, while the network with 20 added interactors showed the most connected components in one network. The 50 interactors were the most exhaustive network in terms of connectivity and in showing the thalassemia disease network components; thus, they were used to perform pathway analysis. Networks were analyzed based on degree, betweenness centrality, and clustering coefficient. These are the topological measures of a network that help identify the top central genes, which are well-connected and are the hub genes of the network. The genes in all 4 networks that had the highest of all three centrality measures and those that were significantly differentially expressed were PRKY, EIF1AY, DDX3Y, CDY2B, and BPY2. These genes secured the most significant position in the disease network in thalassemia.

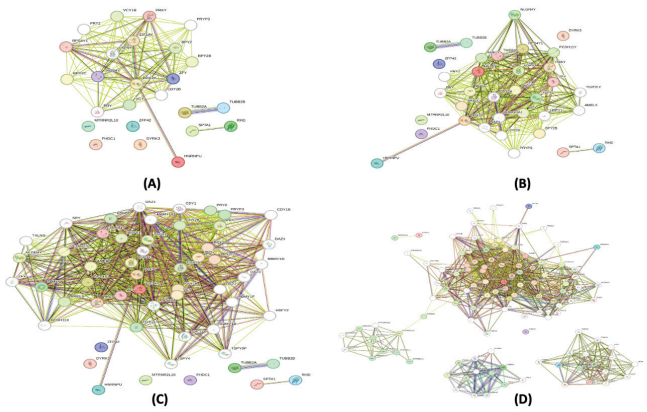

Figure 2: The interaction network of differentially expressed genes. (A) This shows an interaction network of differentially expressed genes with 5 added interactors in STRING. (B) This shows an interaction network of differentially expressed genes with 10 added interactors in STRING. (C) This shows an interaction network of differentially expressed genes with 20 added interactors in STRING. (D) This shows an interaction network of differentially expressed genes with 50 added interactors in STRING. Networks helped in identifying the well-connected hub genes of the thalassemia disease network, which were further cross-checked to be differentially expressed.

Pathway analysis using the Reactome Pathway database showed 47 over-represented pathways in the thalassemia disease network; those were found to be statistically significant with a p-value <0.05. The top enriched pathways in this analysis were key biological processes, such as protein synthesis, RBC membrane integrity, and immune responses, all of which play roles in the pathology of thalassemia. The blood group biosynthesis and translation initiation pathways show the disruptions in hemoglobin production and indicate the need for transfusion in thalassemia patients. These findings suggest a complex interplay between genetic mutations in hemoglobin genes, alterations in protein synthesis pathways, and the body's response to defective RBCs, all contributing to the clinical manifestations of thalassemia. The top genes of the network were searched in the significantly over-represented pathways to check their involvement in the top thalassemia-related pathways. DDX3Y and EIF1AY were found to be present in the top 10 significant pathways, thus indicating their prime role in thalassemia disease.

These two genes were screened through the literature for the existing evidence of their role in Thalassemia and their druggability role. It was identified that both DDX3Y and EIF1AY are suitable as drug targets, but DDX3Y was identified to have a stronger association and prime role in thalassemia than EIF1AY (12). Thus, amongst the top genes, DDX3Y was found to be the most suitable therapeutic target protein as it passed three levels of validation, which are holding the top position in the network, are involved in top significant pathways, and are backed by the literature for the involvement in thalassemia disease complications. DDX3 also showed significant differential gene expression in thalassemia patients as compared to non-thalassaemic individuals (Figure 3). These findings support the hypothesis that DDX3Y is a critical player in thalassemia's molecular landscape. DDX3 was thus

DOI: 10.36838/v7i8.9

selected for further analysis as a potential therapeutic target to identify its inhibitors.

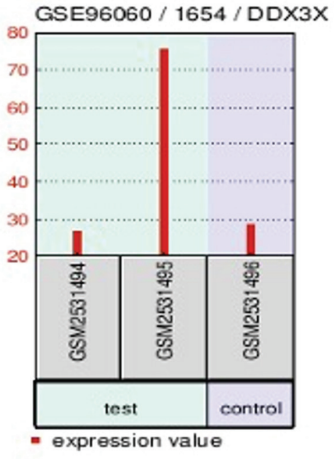

Figure 3: Differential expression of DDX3 in Thalassemia patients. The red bar in the bar graph shows the differential expression (over-expression in homozygous sample) of DDX3 in test samples of patients with Thalassemia as compared to the normal control samples. DDX3 was found to be significantly differentially expressed in thalassemic patients and was also found to be the top hub gene in the disease network.

Virtual Screening analysis:

After performing virtual screening of the libraries of compounds from PubChem and ChEMBL, the top 6 ligands from PubChem (13) TTD and ChEMBL (14) databases were identified based on binding affinity towards DDX3 (Table 1). Neotetrazolium with ChEMBL id CHEMBL1183691 was identified to have the highest binding affinity of -10.2. The lower the binding affinity, the higher the strength of the interaction. Thus, Neotetrazolium holds the most significant position in exhibiting the best binding affinity to DDX3. The surface interaction of Neotetrazolium was visualized in a docked position with DDX3 (Figure 4), along with the interaction of the compound with DDX3 in the ribbon cartoon structure (Figure 5). The amino acid residue level interaction of Neotetrazolium with DDX3 was identified using Biovia Discovery Studio. The interaction of residues of A and B chains of DDX3 with the Neotetrazolium compound was visualized (Figure 6). The compound was seen to form 1 hydrogen bond, 2 pi-cation bonds, 1 pi-pi stacking interaction, 2 pi-alkyl interactions, and 1 pi-sigma interaction with the A and B chains of DDX3.

Table 1: The binding affinities of the top 6 compounds from ChEMBL and PubChem libraries. RMSD/ub and RMSD/lb refer to upper and lower bound root mean square deviation, respectively.

	Compound ID and name	Binding Affinity	rmsd/ub	rmsd/lb
	CHEMBL1183691 (NEOTETRAZOLIUM)	-10.2	0	0
ChEMBL	CHEMBL1079593 (VS-5584)	-10	0	0
	CHEMBL1076263 (SETROBUVIR)	-9.2	0	0
	10393120 (manzamine Y|D05GIC)	-9.9	0	0
PubChem	10029385 (LY2090314|D0Z1DH)	-9.8	0	0
	10218379 (NDT9520492|D0J3FH)	-9.8	0	0

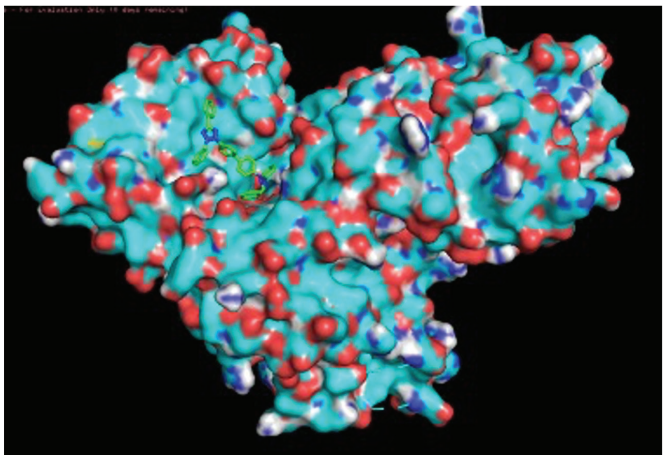

Figure 4: Neotetrazolium docked with DDX3 in surface structure. This shows the top-ranked ligand, Neotetrazolium ligand in green color, docked with DDX3 (surface shown in blue and red color), the most significantly differentially expressed gene in thalassemia. This helped in visualizing the perfect binding of Neotetrazolium in the surface pockets of DDX3.

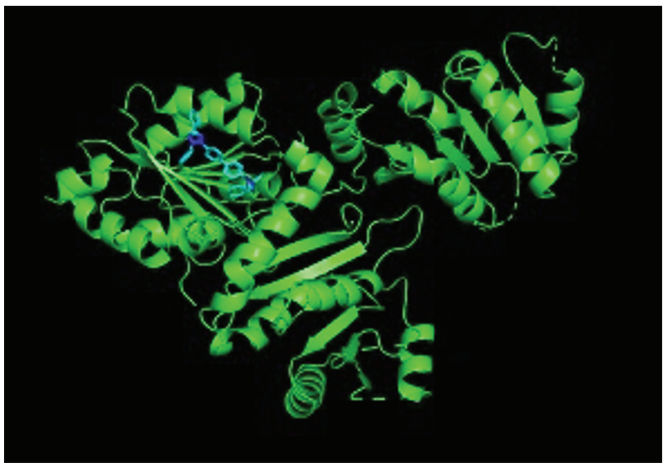

Figure 5: Neotetrazolium docked with DDX3 in ribbon form. This shows the top-ranked ligand, Neotetrazolium ligand in blue color, docked with DDX3 in green color, the most significantly differentially expressed gene.

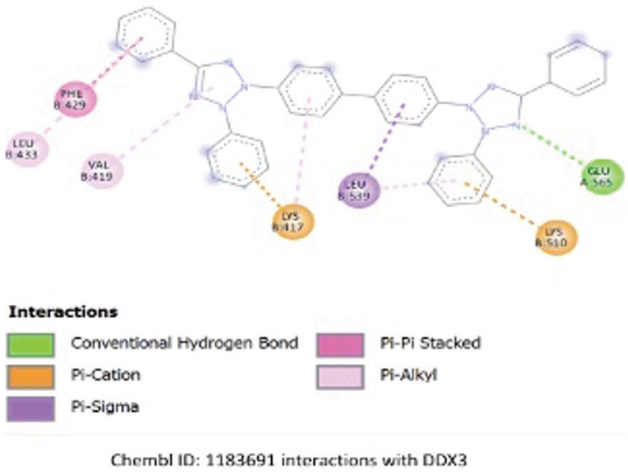

Figure 6: Interactions of Neotetrazolium with DDX3. This shows the amino acid residue interactions of the top-ranked ligand, Neotetrazolium, with DDX3. The hydrogen bond and π-interactions identified contribute to the binding strength and stability of the DDX3–Neotetrazolium complex. The rectangular color bars show the color coding of each type of interaction. This helped in identifying the chemical interactions between the protein and the top-ranked ligand.

Prior research on thalassemia has primarily focused on HBB. Therefore, the goal of this study was to go beyond traditional genes like HBB and discover new therapeutic targets and lead compounds through network and pathway analysis and virtual screening. Thus, it was hypothesized, through our key question, that a wider and more in-depth analysis of gene interactions would help discover unique targets that can effectively be used to develop well-analyzed and researched therapeutic drugs that will uncover new insights into Thalassemia treatment. The results validate the hypothesis that a systems-level approach can identify novel therapeutic targets for thalassemia. DDX3 was discovered as a therapeutic druggable target for thalassemia, and Neotetrazolium is the best ligand that can effectively bind to the target. DDX3 is also a proposed therapeutic target for lung cancer, which indicates its suitability as a drug target.[18]

Through network analysis using STRING,[19] the genes holding the top position in all 4 networks and those that were significantly differentially expressed were PRKY, EIF1AY, DDX3Y, CDY2B, and BPY2. Amongst the top genes, DDX3 was selected for further investigation as it was found to be the most suitable druggable protein in the literature. It was found to be involved in the top over-represented significant thalassemia-related pathways, was present in significantly differentially expressed genes, and held the top position in the network, i.e., it had a high degree, high betweenness centrality, and high clustering coefficient. The results indicated that DDX3 was significantly upregulated in individuals with Thalassemia, as seen in other diseases, such as Medulloblastoma.[20] Gene enrichment analysis revealed that the Rhesus Blood Group Biosynthesis Pathway is heavily involved in the progression of Thalassemia, providing new insights into potential therapeutic targets and negating HBB as the primary gene studied in such cases.[21] Next, the top 6 ligands from PubChem, TTD, and ChEMBL databases were identified based on binding affinity towards DDX3, where Neotetrazolium had the highest binding affinity of -10.2. (ChEMBL ID: 1183691). These findings also highlight the potential for re-purposing existing DDX3-targeted therapies or developing new agents for treating hematologic disorders such as thalassemia. The schematic of the whole workflow is shown in Figure 7.

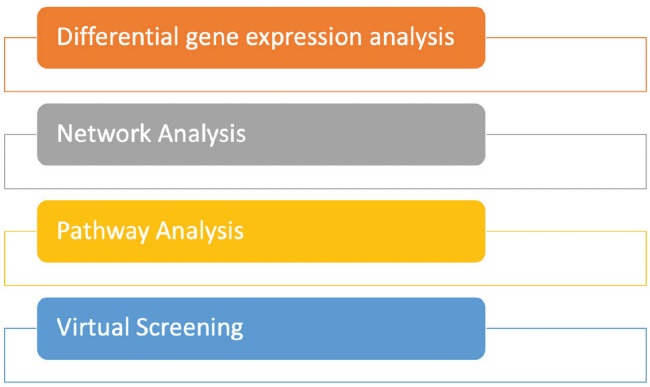

Figure 7: Schematic of the study workflow showing four phases. 1. Differential Gene Expression Analysis, 2. Network Analysis 3. Pathway Analysis and 4. Virtual Screening.

The research conducted, therefore, not only contributes to existing knowledge of potential therapeutic drug targets to inhibit the expression of thalassemia but also adds to it through the novel discovery of the potential use of DDX3. The gene DDX3 was originally identified through this gene expression and possibly would not have been studied if solely focusing on the most well-known genes, such as HBB, that manifest in disease pathways. The identification of DDX3 as a potential diagnostic marker suggests its utility in developing early detection methods for thalassemia, possibly even before symptoms manifest. A major strength of this study is the comprehensive use of multiple bioinformatics tools, including a multitude of datasets from PubChem and ChEMBL, which allowed for an in-depth analysis of the genomic data, leading to the eventual discovery of DDX3 as the most promising target.[22] Targeted therapies against DDX3 could potentially complement existing thalassemia management approaches, including transfusion protocols, iron chelation therapy, and emerging gene therapies. The integration of pathway and network analysis provided a holistic view of the underlying biological mechanisms. Furthermore, virtual screening and molecular docking served as further validation of an already sound workflow. Thus, the study depicts the potential of advanced bioinformatics tools in the identification of novel therapeutic targets for Thalassemia. The interaction between the DDX3 and identified potential ligands, such as Neotetrazolium, can be further experimentally validated for their suitability in the clinical setting. The hypothesis-driven approach used in this study provides a framework for future research, including experimental validation of identified targets and inhibitors.

Conclusion

Overall, the findings of this study contribute to establishing a foundation for further research, as they underscore the sheer importance of bioinformatics in exploring disease mechanisms as well as discovering new drug targets. The importance of this study extends far beyond just Thalassemia. It serves as an example of how these various approaches can uncover targets that have been previously overlooked. As bioinformatics continues to evolve, studies like this demonstrate the potential of combining computational methods with experimental research to speed up medicinal discoveries. With more research and experimental validation, DDX3 could become a fundamental principle in Thalassemia treatment strategies. The application of bioinformatics techniques continues to open new avenues for research, with the potential to improve personalized medicine and therapeutic interventions.

Acknowledgments

SV and AV would like to acknowledge Aashna Saraf, Founder of CreatED, for providing valuable feedback and guidance throughout the project.

References

1. Muncie, H. L., Jr.; Campbell, J., Alpha and beta thalassemia. *Am Fam Physician* **2009**, *80* (4), 339-44.

DOI: 10.36838/v7i8.9

2. Mayo Clinic Thalassemia. https://www.mayoclinic.org/diseases-conditions/thalassemia/symptoms-causes/syc-20354995 (accessed 27 August).
3. Marengo-Rowe, A. J., Structure-function relations of human hemoglobins. *Proc (Bayl Univ Med Cent)* **2006**, *19* (3), 239-45.
4. He, L. N.; Chen, W.; Yang, Y.; Xie, Y. J.; Xiong, Z. Y.; Chen, D. Y.; Lu, D.; Liu, N. Q.; Yang, Y. H.; Sun, X. F., Elevated Prevalence of Abnormal Glucose Metabolism and Other Endocrine Disorders in Patients with β-Thalassemia Major: A Meta-Analysis. *Biomed Res Int* **2019**, *2019*, 6573497.
5. Cao, A.; Kan, Y. W., The prevention of thalassemia. *Cold Spring Harb Perspect Med* **2013**, *3* (2), a011775.
6. Colah, R.; Italia, K.; Gorakshakar, A., Burden of thalassemia in India: The road map for control. *Pediatric Hematology Oncology Journal* **2017**, *2* (4), 79-84.
7. Ali, S.; Mumtaz, S.; Shakir, H. A.; Khan, M.; Tahir, H. M.; Mumtaz, S.; Mughal, T. A.; Hassan, A.; Kazmi, S. A. R.; Sadia; Irfan, M.; Khan, M. A., Current status of beta-thalassemia and its treatment strategies. *Mol Genet Genomic Med* **2021**, *9* (12), e1788.
8. Shi, P. A., Transfusion Management in Patients with Hemoglobinopathies. In *Transfusion Medicine and Hemostasis (Second Edition)*, al, S. e., Ed. Elsevier: San Diego, 2013; pp 327-336.
9. Qadah, T.; Jamal, M. S., Computational Analysis of Protein Structure Changes as a Result of Nondeletion Insertion Mutations in Human β-Globin Gene Suggests Possible Cause of β-Thalassemia. *Biomed Res Int* **2019**, *2019*, 9210841.
10. Pokharel, B.; Ravikumar, Y.; Rathinavel, L.; Chewonarin, T.; Pongpom, M.; Tipsuwan, W.; Koonyosying, P.; Srichairatanakool, S., The Discovery of Selective Protein Arginine Methyltransferase 5 Inhibitors in the Management of β-Thalassemia through Computational Methods. *Molecules* **2024**, *29* (11), 2662.
11. García-Campos, M. A.; Espinal-Enríquez, J.; Hernández-Lemus, E., Pathway Analysis: State of the Art. *Front Physiol* **2015**, *6*, 383.
12. Barrett, T.; Edgar, R., Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods Enzymol* **2006**, *411*, 352-69.
13. Zardecki, C.; Dutta, S.; Goodsell, D. S.; Voigt, M.; Burley, S. K., RCSB Protein Data Bank: A Resource for Chemical, Biochemical, and Structural Explorations of Large and Small Biomolecules. *Journal of Chemical Education* **2016**, *93* (3), 569-575.
14. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003**, *13* (11), 2498-504.
15. Fabregat, A.; Sidiropoulos, K.; Viteri, G.; Forner, O.; Marin-Garcia, P.; Arnau, V.; D'Eustachio, P.; Stein, L.; Hermjakob, H., Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics* **2017**, *18* (1), 142.
16. Santos-Martins, D.; Forli, S.; Ramos, M. J.; Olson, A. J., AutoDock4(Zn): an improved AutoDock force field for small-molecule docking to zinc metalloproteins. *J Chem Inf Model* **2014**, *54* (8), 2371-9.
17. Dallakyan, S.; Olson, A. J., Small-Molecule Library Screening by Docking with PyRx. In *Chemical Biology: Methods and Protocols*, Hempel, J. E.; Williams, C. H.; Hong, C. C., Eds. Springer New York: New York, NY, 2015; pp 243-250.
18. Bol, G. M.; Vesuna, F.; Xie, M.; Zeng, J.; Aziz, K.; Gandhi, N.; Levine, A.; Irving, A.; Korz, D.; Tantravedi, S.; Heerma van Voss, M. R.; Gabrielson, K.; Bordt, E. A.; Polster, B. M.; Cope, L.; van der Groep, P.; Kondaskar, A.; Rudek, M. A.; Hosmane, R. S.; van der Wall, E.; van Diest, P. J.; Tran, P. T.; Raman, V., Targeting DDX3 with a small molecule inhibitor for lung cancer therapy. *EMBO Molecular Medicine* **2015**, *7* (5), 648-669-669.
19. Mering, C. v.; Huynen, M.; Jaeggi, D.; Schmidt, S.; Bork, P.; Snel, B., STRING: a database of predicted functional associations between proteins. *Nucleic Acids Research* **2003**, *31* (1), 258-261.
20. Swarup, A.; Bolger, T. A., The Role of the RNA Helicase DDX3X in Medulloblastoma Progression. *Biomolecules* **2024**, *14* (7), 803.
21. Thein, S., Molecular basis of β thalassemia and potential therapeutic targets. *Blood Cells, Molecules, and Diseases* **2017**, *70*.
22. Rosenkranz, A. A.; Slastnikova, T. A., Prospects of Using Protein Engineering for Selective Drug Delivery into a Specific Compartment of Target Cells. *Pharmaceutics* **2023**, *15* (3)

Authors

Samaya is an 11th grader studying the International Baccalaureate Diploma Program at The Cathedral and John Connon School, Mumbai. She wrote this paper at the end of the 10th grade, leading into the 11th grade. Now, she works towards giving young teenagers in her school the opportunity to engage in research in STEM to garner a holistic view of what it means to be a person studying the sciences with the launch of her website, 'To The Core.' In the future, she hopes to explore the field of Medicine.

Aneesha is an 11th grader studying the International Baccalaureate Diploma Program at The Cathedral and John Connon School, Mumbai. Being extremely passionate about Mathematics and Biology, she undertook a Program in Bioinformatics in the Summer of 2024. This formed the basis of the research paper demonstrating her curiosity in this area. Going ahead, she is keen to pursue Biology for her undergraduate studies.

Nirupma Singh is a Bioinformatics Scientist with a doctorate in Biotechnology and Bioinformatics from the University of Delhi with six years of hands-on research and development experience. Her journey is marked by a robust foundation in Machine Learning and Python, with five years of expertise. Proficient in Linux and cloud servers like AWS. She excels in structural biology, systems biology, protein/gene network analysis, data mining, and computational genomics.