

Integrated Analysis : Gene Expression, Survival, Biomarkers and Drug Targets

Submitted in partial fulfillment of the requirements for the award of Master of
Bioinformatics and Data Science

By

Nikhitha Sebastian.T (43739011)



DEPARTMENT OF BIOINFORMATICS AND DATA SCIENCE

SCHOOL OF BIO AND CHEMICAL ENGINEERING

SATHYABAMA

INSTITUTE OF SCIENCE AND
TECHNOLOGY (DEEMED TO BE
UNIVERSITY)

Accredited with Grade “A” by NAAC | 12B Status by UGC | Approved by
AICTE

JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119

April - 2024



DEPARTMENT OF BIOINFORMATICS AND DATASCIENCE

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **NIKHITHA SEBASTIAN.T (43739011)** who carried out the Project entitled **“Integrated Analysis:Gene Expression, Survival, Biomarkers and Drug Targets”** under my supervision from November 2023 to April 2024.

Internal Guide

External Guide (if Applicable)

Head of the Department

Submitted for Viva voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

I **NIKHITHA SEBASTIAN.T (4 3 7 3 9 0 1 1)** here by declare that the Project Report entitled **“Integrated Analysis:Gene Expression, Survival, Biomarkers and Drug Targets”** done by me under the guidance of **Dr.Mahalakshmi** is submitted in partial fulfillment of the requirements for the award of Master of science in **Bioinformatics and Datascience.**

DATE

NIKHITHA SEBASTIAN.T

PLACE: Chennai

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA INSTITUTE OF SCIENCE & TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them

I convey my thanks to **Dr. REKHA CHAKRAVARTHI M.E., Ph.D., Dean**, of Arts And Science, **Dr. JEMMY CHRISTY.**, Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. MAHALAKSHMI** department of bioinformatics and data science for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Bioinformatics and Data science** who were helpful in many ways for the completion.

(NIKHITHA SEBASTIAN.T)

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	06
	PLAGIARISM REPORT	07
I	REVIEW ARTICLE	08
II	TRANSLATIONAL BIOINFORMATICS	15
1	GENE EXPRESSION ANALYSIS	
	1.1 Introduction	
	1.2 Aim	15
	1.3 materials and methods	
	1.4 procedure Objectives	16
	1.5 result	17
2	Cumulative survival probability calculation using KM Plotter	19
	2.1 Introduction	
	2.2 Aim	19
	2.3 materials&methods	
	2.4 procedure	20
	2.5 result	21
3	Biomarkers: identifying biological markers for disease Diagonis and therupatic monitoring	24
	3.1 Introduction	
	3.2 Aim	24
	3.3 Procedure	
	3.4 Result	25
	3.5 Interpretation	25
4	Drug-Target Interaction Prediction: Approaches for predicting interactions specific to patient treatment and drug design	26

	4.1 Introduction	
	4.2 Aim	27
	4.3 materials and methods	
	4.4 procedure	28
	4.5 Result	29
	4.6 conclusion	30
III	MATLAB CERTIFICATES	31

ABSTRACT

Gene expression analysis is crucial in understanding the intricate mechanisms of life, from development to disease. Gene expression analysis facilitates the identification of potential drug targets and the evaluation of drug efficacy and safety. By studying how drugs modulate gene expression profiles, researchers. Gene expression analysis facilitates drug development by providing insights into the efficacy and safety of potential drug candidates. Additionally, it can identify existing drugs that may be repurposed for new indications based on their effects on gene expression profiles.

The identification of robust biomarkers holds immense promise for personalized medicine, aiding in early detection, disease classification, and treatment stratification. This abstract encapsulates the essence of our research endeavors, aimed at advancing our understanding of complex biological systems and translating genomic insights into clinical applications.

Drug-target interactions play a crucial role in various aspects of pharmacology and drug development. Understanding the interactions between drugs and their molecular targets is essential for ensuring therapeutic efficacy. Effective binding of a drug to its intended target can lead to desired pharmacological effects, such as alleviation of symptoms or treatment of diseases. This abstract explores the multifaceted importance of drug-target interactions across various facets of drug discovery and development. We delve into their critical role in validating potential drug targets, elucidating mechanisms of action, and guiding the design of novel therapeutic agents with enhanced efficacy and reduced adverse effects. Through a comprehensive understanding of drug-target interactions, we aim to advance the field of pharmacology.

Nikhitha Sebastian T

ORIGINALITY REPORT

8%	6%	5%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	hdl.handle.net Internet Source	3%
2	Submitted to Sir George Monoux College Student Paper	1%
3	buddyxtheme.com Internet Source	1%
4	www.mdpi.com Internet Source	1%
5	www.ncbi.nlm.nih.gov Internet Source	1%
6	Milad Besharatifard, Fatemeh Vafaei. "A review on graph neural networks for predicting synergistic drug combinations", Artificial Intelligence Review, 2024 Publication	1%
7	Weikaixin Kong, Gianmarco Midena, Yingjia Chen, Paschalis Athanasiadis, Tianduanyi Wang, Juho Rousu, Liye He, Tero Aittokallio. "Systematic review of computational methods for drug combination prediction",	1%

7

Computational and Structural Biotechnology
Journal, 2022
Publication

Exclude quotes Off
Exclude bibliography On

Exclude matches Off

UNLOCKING THERAPEUTIC SYNERGY : SYSTEMS PHARMACOLOGY PERSPECTIVES

Therapeutic synergy refers to the combined effect of two or more therapeutic agents or interventions that produce a greater effect together than the sum of their individual effects. This synergy can occur at various levels, including molecular, cellular, physiological, or psychological.

Here are some instances of therapeutic synergy:

Drug Combinations: In pharmacology, combining different medications can enhance their efficacy while minimizing side effects. For instance, combining two drugs with different mechanisms of action can target multiple pathways involved in a disease process. An example is the combination of antibiotics in treating bacterial infections to prevent resistance and achieve better outcomes.

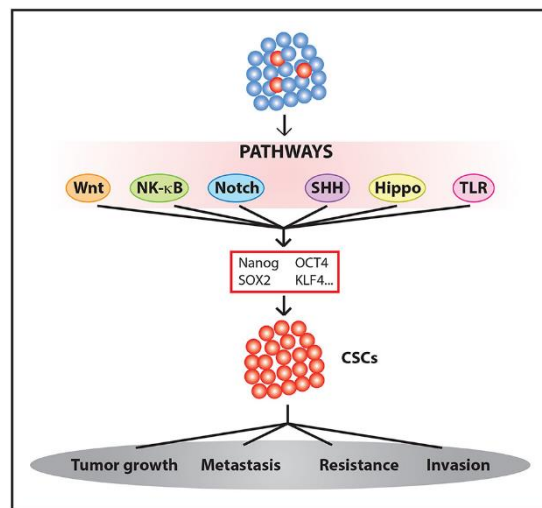
Complementary Therapies: Better treatment outcomes can result from combining various therapeutic methods, such as medicine, psychotherapy, and lifestyle modifications. For instance, when treating anxiety or depression, cognitive-behavioral therapy (CBT) plus medication is frequently more successful than either treatment alone.

Interactions among multiple drug targets:

Interactions among multiple drug targets refer to the complex relationships and effects that occur when therapeutic agents interact with more than one molecular target in the body. In many diseases, such as cancer or neurological disorders, the pathology involves multiple molecular pathways or targets. Targeting a single pathway or molecule may not be sufficient to achieve optimal therapeutic outcomes. Therefore, drugs that can interact with multiple targets simultaneously or sequentially can exhibit enhanced efficacy and therapeutic synergy.

For example, in cancer treatment, combination therapies often target multiple signaling pathways involved in tumor growth and survival. By simultaneously inhibiting different targets such as growth factor receptors, angiogenesis pathways, or cell cycle regulators, combination therapies can achieve greater tumor regression and reduce the likelihood of drug resistance.

Similarly, in psychiatric disorders like depression or schizophrenia, drugs may target multiple neurotransmitter systems such as serotonin, dopamine, and norepinephrine. By modulating the activity of multiple targets, these drugs can produce a more comprehensive effect on the complex neurochemical imbalances underlying these disorders.



NETWORK PHARMACOLOGY

Network pharmacology is an interdisciplinary field that combines ideas and methods from systems biology, pharmacology, bioinformatics, and network science to investigate complicated interactions that exist between pharmaceuticals, biological molecules, and disease pathways on a systemic level. Network pharmacology takes into account the larger network of interactions within biological systems, such as protein-protein interactions, metabolic processes, as opposed to concentrating only on specific drug-target interactions.

Techniques used in network pharmacology consist of:

Network Analysis: Drug interactions, targets, illnesses, and biological processes are represented and analyzed as networks or graphs in network pharmacology using network theory and analysis. These networks can be anything from straightforward networks of interactions to intricate, multi-layered networks that mimic the biological systems' hierarchical structure.

Identification of Drug Combination Therapies: By discovering synergistic medication combinations that target numerous nodes in illness networks, network pharmacology aids in the logical design of combination therapies. Through the consideration of inter-pathway connections and crosstalk, network pharmacology aims to maximize therapeutic benefits while minimizing negative effects.

Personalized Medicine: By examining individual differences in molecular networks and determining patient-specific medication responses, network pharmacology helps to

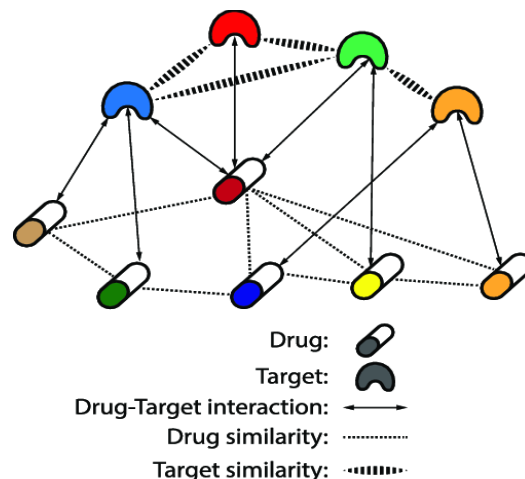
advance personalized medicine. Network pharmacology makes it possible to predict drug efficacy, toxicity, and the best dosing schedules for specific patients by combining genomic and clinical data.

Overall, network pharmacology offers an effective framework for comprehending the intricacy of biological systems, locating new therapeutic targets, and streamlining the medication development and discovery procedures.

Computational methods for drug combination prediction:

Computational methods for drug combination prediction leverage various approaches, including network-based analysis, machine learning algorithms, and systems biology modeling, to identify synergistic drug combinations with enhanced therapeutic effects. Here are some commonly used computational methods:

Network Analysis: Analyzing drug-target interaction networks, protein-protein interaction networks, or biological pathway networks to identify potential drug combinations that target different nodes within disease-related networks.



Deep Learning: Using deep learning methods to identify new drug combinations and extract intricate patterns from massive amounts of drug screening data, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs).

Multi-layer Network Integration: Integrating multiple types of biological networks (e.g., drug-target interactions, protein-protein interactions, gene expression networks) to capture the complex interplay between drugs, targets, and pathways.

Cell Signaling Pathway Models: Constructing computational models of cell signaling pathways to simulate the effects of drug combinations on intracellular signaling dynamics and identify synergistic interactions that modulate pathway activity.

Omics Data Integration: Combining various omics data, including transcriptomics, proteomics, and genomes, to find molecular signatures linked to beneficial drug combinations and rank potential drug combinations for confirmation in experiments.

Systems pharmacology analyses focus on the quantitative relationships and interactions between drugs and biological systems, considering the behavior of the system as a whole rather than its individual constituents.

Clinical applications of Systems Pharmacology:

Personalized medicine: Systems pharmacology allows the clinical data to be integrated with multi-omics data (genomics, transcriptomics, proteomics) to create patient-specific models that improve treatment plans based on individual features and predict drug response.

Drug Repurposing: Systems pharmacology can find new therapeutic indications for already-approved medications by examining drug-target interactions, biological pathways, and disease networks. This approach may accelerate drug development and lower expenses related to conventional drug discovery procedures.

Drug Combination Therapy: By offering a framework for comprehending the intricate relationships between medications, targets, and pathways, systems pharmacology helps doctors rationally design combination therapies that maximize therapeutic efficacy while reducing side effects and target multiple disease mechanisms.

Pharmacogenomics: By combining genomic information with pharmacological models, doctors can customize treatment plans for patients by anticipating individual differences in medication response, side effects, and metabolism.

Overall, systems pharmacology has the potential to completely transform drug research, discovery, and clinical practice by offering a thorough comprehension of drug action and disease mechanisms and by enabling personalized and precision medicine approaches customized to specific patients.

CHALLENGES OF SYSTEMS PHARMACOLOGY:

One of the main challenges in systems pharmacology is integrating heterogeneous data from many sources, including proteomics, metabolomics, genomes, and clinical data. Ensuring repeatability and interoperability across many platforms and research requires standardizing metadata, ontologies, and data formats.

Developing accurate and comprehensive computational models of biological systems requires accounting for the complexity and dynamics of molecular interactions, cellular processes, and physiological responses. Validating these models with experimental data

and clinical outcomes is challenging due to the inherent uncertainty and variability in biological systems.

In systems pharmacology research, handling sensitive patient data presents privacy, security, and ethical issues. Upholding public confidence in research endeavors necessitates adhering to data protection standards and putting strong security measures in place to protect patient information.

REFERENCES:

1. Jansen J, Assmus J, Rath-Wolfson L, Aarsæther E, Bache-Mathiesen L, Goksør LA, Dahle GO, Fosså SD, Ursin G, Engh JA. Impact of metformin on patient-reported outcomes in diabetes-related female breast cancer. *Endocrine-Related Cancer*. 2020;27(6):347-57
2. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nature chemical biology*. 2008 Nov;4(11):682-90
3. Bansal M, Yang J, Karan C, Menden MP, Costello JC, Tang H, Xiao G. A community computational challenge to predict the activity of pairs of compounds. *Nature Biotechnology*.
4. Zhu F, Ma XH, Qin C, Tao L, Liu X, Shi Z, Zhang CL, Tan CY, Chen YZ. Drug discovery prospect from untapped species: indications from approved natural product drugs. *PLoS One*. 2012 Apr 20;7(7):e40653
5. Sorger PK, Allerheiligen SRB, Abernethy DR, Altman RB, Brouwer KL, Califano A, et al. Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms. An NIH White Paper by the QSP Workshop Group—October. *CPT: Pharmacometrics & Systems Pharmacology*. 2011;1(1):e2.
6. Sedghamiz H, Morris MK, Crampin EJ. Challenges for the computational modeling of multiscale drug pharmacology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. 2018;10(3):e1401

TRANSLATIONAL BIOINFORMATICS

GENE EXPRESSION ANALYSIS

Introduction:

A biological network is a method of representation of biological systems as complex set of interactions between biological entities. In a biological network, nodes represent the biological molecules like genes or proteins and edges represent the interaction between them. Cytoscape is a open source software platforms for visualizing and analyzing complex biological networks and pathways. Cytohubba provide a user-friendly interface to explore important nodes in biological networks. It predicts and explores the important nodes and edges in a given network by several topological algorithm such as degree, bottleneck, MCC, DMNC, MNC, EPC, Eccentricity, Closeness, Stress, Clustering co-efficient and betweenness.

Aim:

To build a network to understand the gene interrelationship and regulatory mechanism.

Materials and methods:

GEO2R: It is tool used in bioinformatics for the analysis of gene expression data and helps to identify genes that are differentially expressed. It compares gene expression levels between sample groups like diseased vs. healthy.

Cytoscape: This is a software tool used for visualizing and analyzing molecular and genetic interaction networks.

Cytohubba: It is a plugin for Cytoscape and provides 11 topological algorithms such as Degree, EPC(Edge Percolated Component), MNC(Maximum Neighborhood Component), DMNC(Density of Maximum Neighborhood Component), MCC(Maximal Clique Centrality) and six centralities such as Stress, Bottleneck, Closeness, Betweenness, Radiality, EcCentricity based on shortest paths.

Cluego: This a Cytoscape app that facilitates the biological interpretation of large lists of genes and proteins by selecting representative Gene Ontology terms and pathways from multiple ontologies and visualizes them into functionally organized networks.

Procedure:

- For the database and disease Gastric cancer Geo2R is the source of the geo dataset (GSE261264).
- After choosing the suitable parameters, the Geo dataset is retrieved.
- Upregulated genes were copied and pasted into the cytoscape to build the network.
- Built a biological network using BioGrid in cytohubba to analyse the network and identify important nodes based on various topological algorithm to determine the top 10 genes.
- Genes must be copied and pasted in Cluego and continue cluego parameter such as species ontology and statistical settings.

Run cluego analysis to identify functional categories or pathways associated with the selected nodes.

Results:

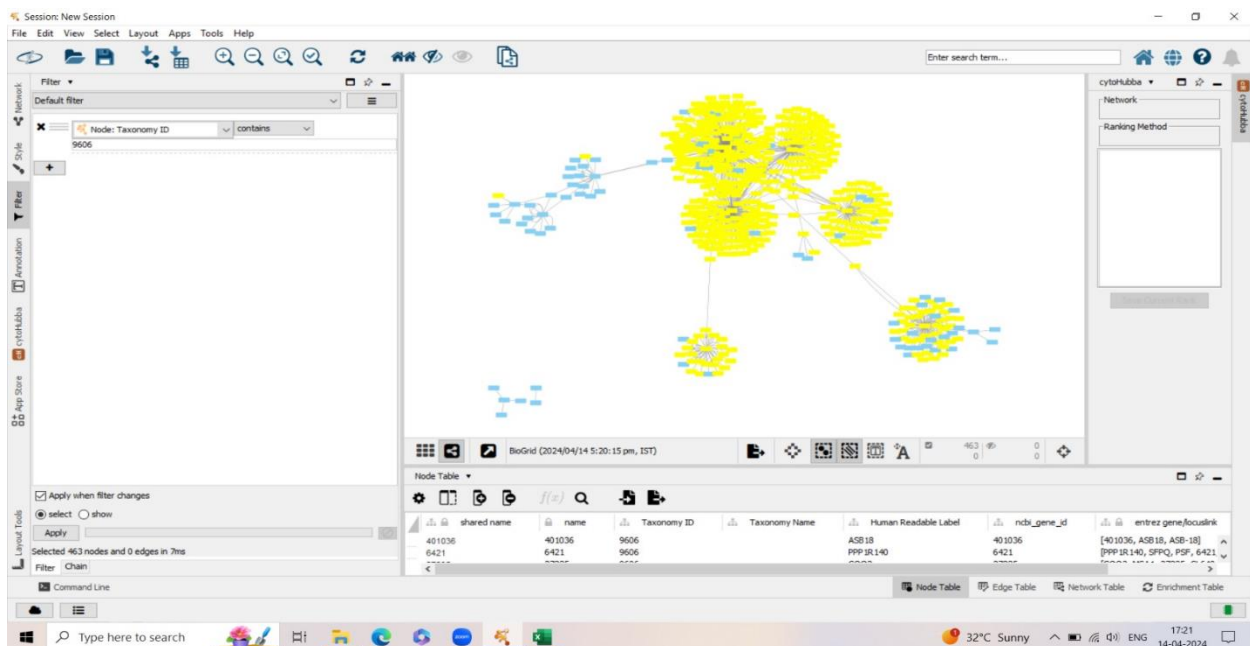


Fig 1.1

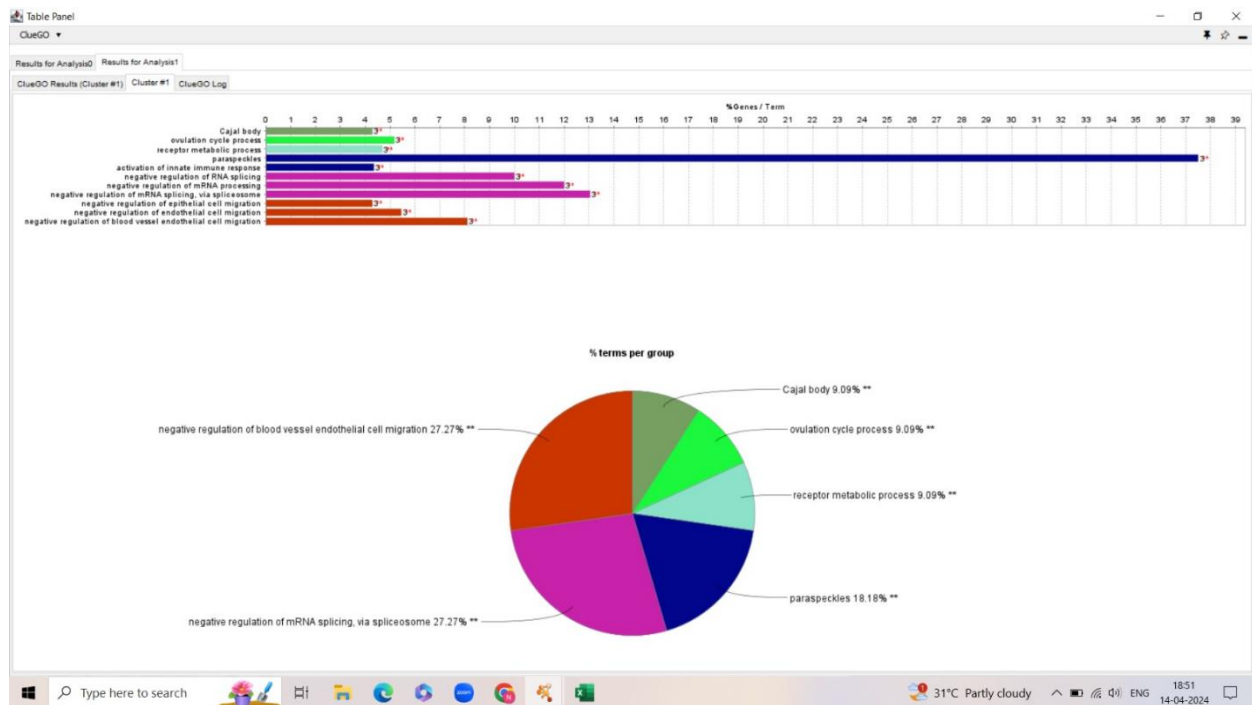


Fig 1.2

Interpretation:

The phrase "negative regulation of blood vessel endothelial cell migration" suggests a process that inhibits or slows down the movement of endothelial cells, which are crucial for forming new blood vessels (angiogenesis). This could potentially mean a reduction in the formation of new blood vessels, which is significant because tumors, including gastric cancer, often rely on angiogenesis to supply them with nutrients and oxygen.

The negative regulation of mRNA splicing via spliceosome in gastric cancer, indicated by a 27.27% decrease, suggests potential alterations in the splicing process of genes related to this type of cancer. The spliceosome is a complex molecular machine responsible for removing introns from pre-mRNA transcripts, thus forming mature mRNA. Any dysregulation in this process can lead to aberrant mRNA splicing, which in turn may contribute to the development and progression of cancer.

A Cajal body, also known as a coiled body or CB, is a subnuclear organelle found in the nucleus of cells. They are involved in various cellular processes including the assembly and modification of ribonucleoproteins. In the context of gastric cancer, it could indicate that there's some significant association between the presence or activity of Cajal bodies and gastric cancer.

Paraspeckles refers to a subnuclear body found in the nucleus of mammalian cells. They are dynamic structures composed of proteins and long non-coding RNAs. In the context

of gastric cancer, it might suggest that there's some association between paraspeckles and gastric cancer.

Receptor metabolic process likely refers to a biological process involving receptors, which are proteins on the surface of cells that can receive signals from the environment and trigger a response within the cell. In the context of cancer, including gastric cancer, abnormalities in receptor signaling pathways are common and can contribute to tumor growth and progression.

Cumulative survival probability calculation using KM Plotter

INTRODUCTION:

KM-plotter is the most sophisticated online survival analysis tool, performing all calculations in real time (not loading pre-calculated images). The background database is manually curated. Gene expression data and relapse free and overall survival information are downloaded from GEO, EGA and TCGA. The database is handled by a PostgreSQL server, which integrates gene expression and clinical data simultaneously. To analyze the prognostic value of a particular gene, the patient samples are split into two groups according to various quantile expressions of the proposed biomarker. The two patient cohorts are compared by a Kaplan-Meier survival plot, and the hazard ratio with 95% confidence intervals and log rank P value are calculated. Databases and clinical data are supervised and extended regularly.

To avoid missing correlations due to the use of a specific cutoff, all available cutoff values between the lower and upper quartiles of expression are used for the selected gene, and false discovery rate (FDR) using the Benjamini-Hochberg method is computed to correct for multiple hypothesis testing. The cutoff value with the highest significance (lowest FDR) is determined. In case of multiple cutoff values with identical significance, the cutoff with the highest hazard (HR) rate is selected for the final analysis.

AIM

Calculate Proposed gene JM21 cumulative survival probabilities.

MATERIALS AND METHODS:

KM-PLOTTER: KM-plotter is the most sophisticated online survival analysis, performing all calculations in real time (not loading pre-calculated images). The background database is manually curated. Gene expression data and relapse-free and overall survival information are downloaded from GEO, EGA and TCGA.

Procedure:

Kaplan Meier plotter home page was accessed using the following hyperlink <https://kmplot.com/analysis/index.php?p=background>

Proposed gene/protein (o/p of Ex No 1) ---- was submitted in the pan cancer portal

Below listed parameters were opted for analysis

Restriction subtype : All

Split patients by-Auto select best cut off

Survival: Overall survival(OS)

Survival : All

Restrict analysis based on cellular content: All

Interpretation

- KM-plot recognizes **proposed gene symbols JADE2** (including HUGO Gene Nomenclature Committee approved).
- Sources for the database include GEO, EGA, TCGA, Metabric, Impact, and PubMed repositories.

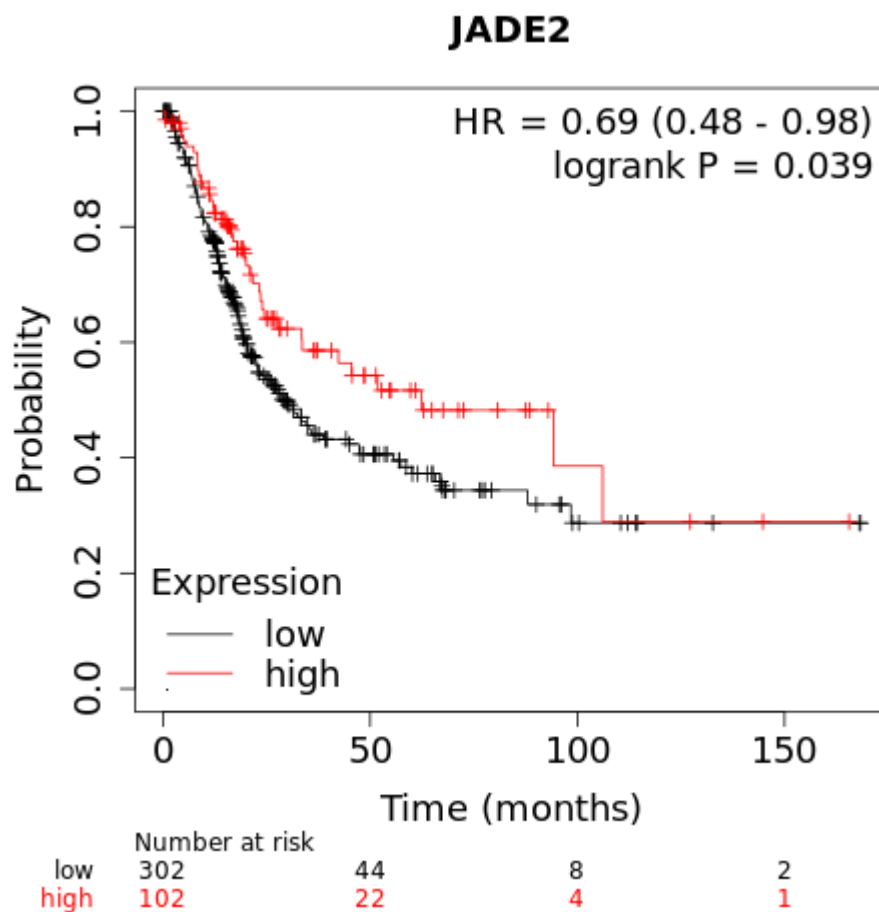


Fig 2.1

Hazard ratio for JADE2 is 0.69 and low and high expression cohort is 28.8 and 62.3 respectively.

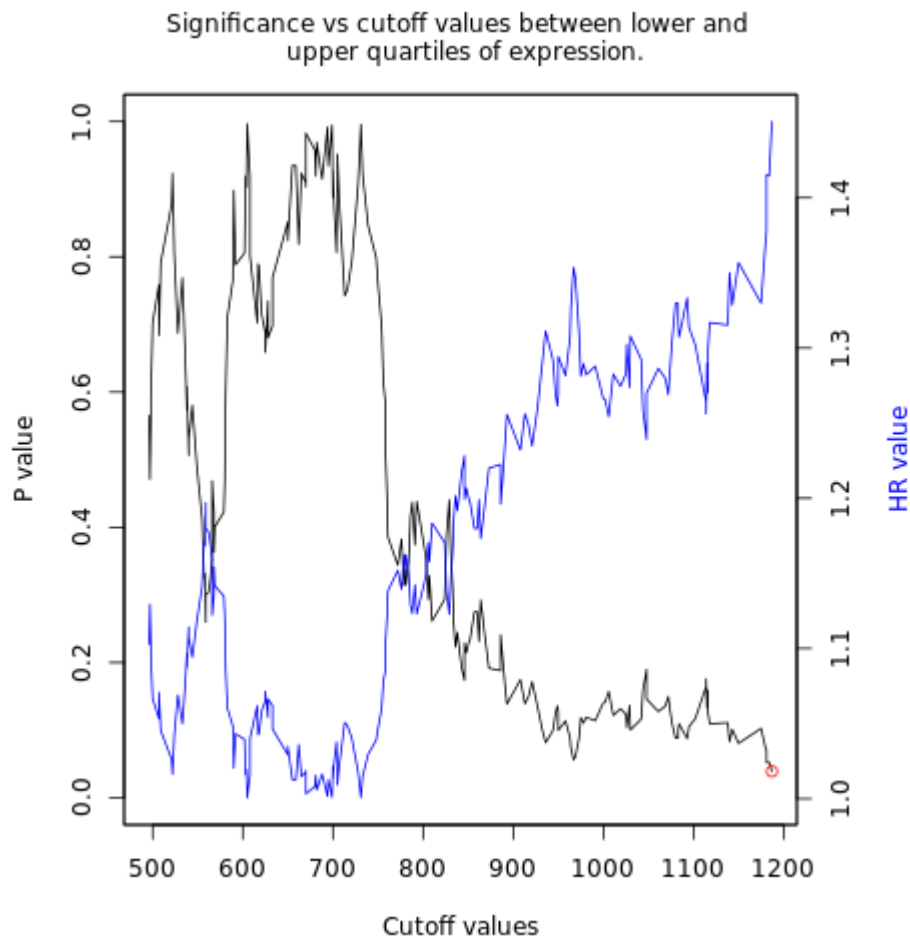


Fig 2.2

Low expression cohort (months)	High expression cohort (months)
28.8	62.3

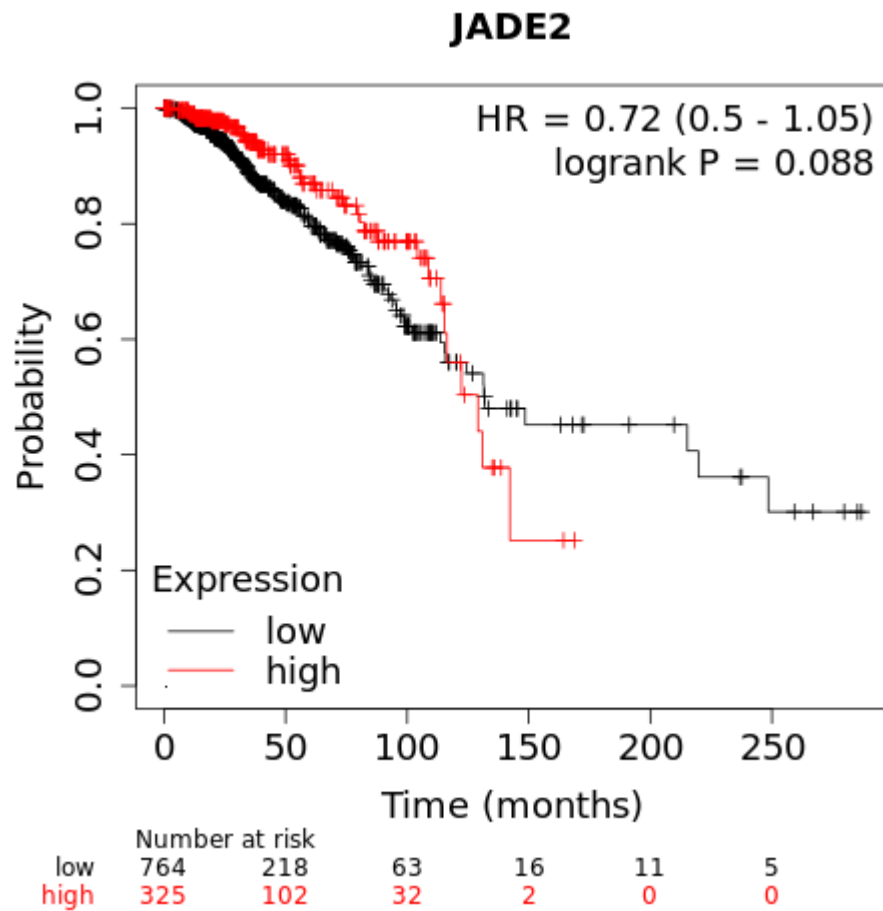


Fig 2.3

Hazard ratio of JADE2 is 0.72 and the low and high expression cohort is 131.97 and 129.1 respectively.

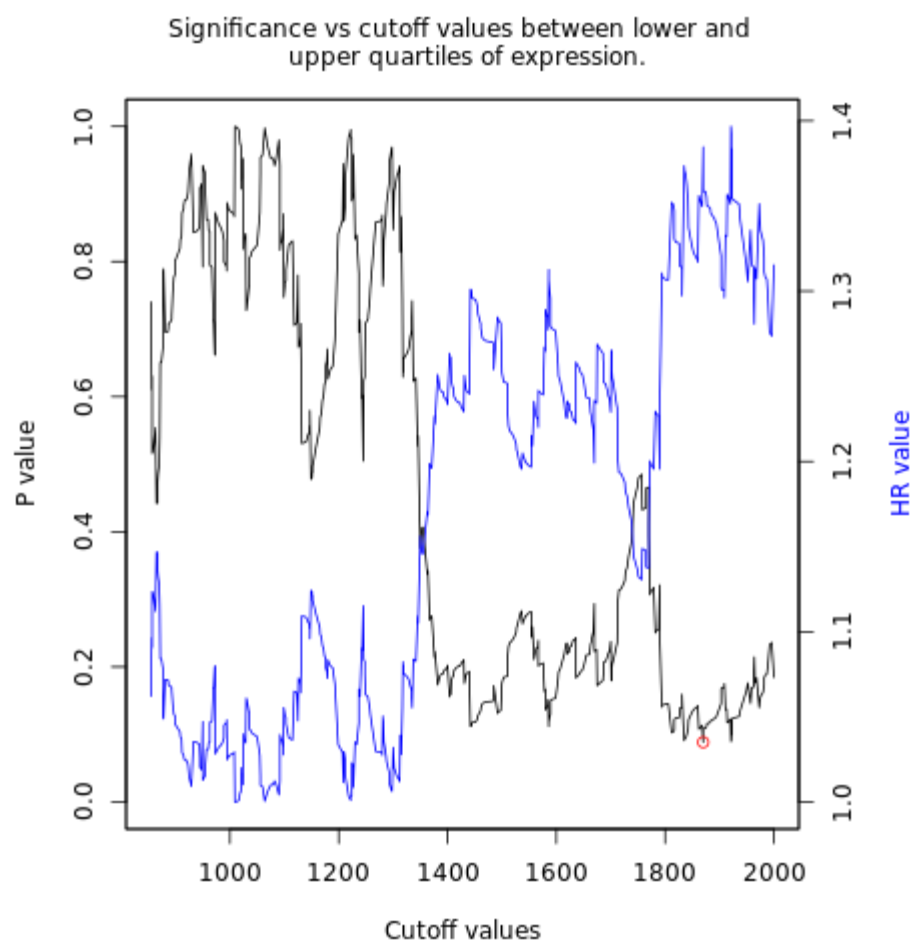


Fig 2.4

Low expression cohort (months)	High expression cohort (months)
131.97	129.1

Biomarkers: identifying biological markers for disease Diagnosis and therapeutic monitoring

INTRODUCTION:

TheMarker (a) systematically offers five types of therapeutic biomarker, which realized the collective consideration among different drug discovery stages, (b) comprehensively describes the therapeutic biomarker for the largest number of drugs among existing knowledge bases, and (c) extensively covers the largest classes of disease among available databases. Due to the rapid advances in AI technology, the comprehensive data collected by TheMarker were expected to have great implication and impact on the clinical development and clinical practice of drugs.

Aim:

Identify the disease markers related to Parkinsonism and categorize them into pharmacodynamic Biomarker, Monitoring biomarker, Safety biomarker.

Procedure:

- The marker database was retrieved using following link <http://themarker.idrblab.cn/about>
- Top listed mRNA/protein Biomarker relevant to the Parkinsonism data were retrieved and literature survey was done.

Result:

The biological marker is identified for disease diagnosis.

Interpretation:

General information about the marker:

BIOMARKER ID	BIOMARKER NAME	CLASS	GENE NAME	BIOMARKER TYPE	REPRESENTATIVE DRUG	REPRESENTATIVE DISEASE
BZ5K3P	Ferritin	Protein	FTMT	SAF PDY MOI	Deferiprone	Parkinsonism

Drug Name: Deferiprone
Car T cell therapy

Application of the biomarker in Diagnosis and Prognosis:

Diagnosis : Acute respiratory illness, Alzheimers disease.

Prognosis : Acute decompensated heart failure, Acute heart failure.

Tissue-Specific RNA Expression of this Biomarker (in total 32 tissues)

General Information of the Biomarker (Biomarker ID: BZ5K3P)							
Biomarker Name	Ferritin						
Biomarker Class	Protein						
Gene Name	FTMT						
Biomarker Type	PDY SAF MOI PRD SUR						
Sequence	> BZ5K3P Ferritin MLSCFRLLSR HISPSLASLR PVRCCFALPL RWAPGRPLDP RQIAPRRPLA AAASSRDPTG PAAGPSRVQR NFHPDSEAAI NRQINLELYA SYVYLSMAYY FSRDDVALNN FSRYFLHQSR Show More						
Function	Catalyzes the oxidation of ferrous iron to ferric iron and stores iron in a soluble, non-toxic, readily available form. Important for iron homeostasis. Iron is taken up in the ferrous form and deposited as ferric hydroxides after oxidation.						
External Links	<table> <tr> <td>Gene ID</td><td>94033 ↗</td></tr> <tr> <td>UniProt ID</td><td>FTMT_HUMAN ↗</td></tr> <tr> <td>HGNC ID</td><td>HGNC:17345 ↗</td></tr> </table> Show More	Gene ID	94033 ↗	UniProt ID	FTMT_HUMAN ↗	HGNC ID	HGNC:17345 ↗
Gene ID	94033 ↗						
UniProt ID	FTMT_HUMAN ↗						
HGNC ID	HGNC:17345 ↗						

Fig 3.1

Drug-Target Interaction Prediction: Approaches for predicting interactions specific to patient treatment and drug design

Introduction:

Drug targets are specific molecules or biological structures involved in disease pathways that can be modulated by drugs to produce a therapeutic effect. Selecting the right drug target is crucial for the development of effective and safe medications.

SwissTarget Prediction: SwissTargetPrediction is an online tool to predict the macromolecular targets (proteins from human, mouse and rat) of bioactive small molecules. This is useful to understand the molecular mechanisms underlying a given phenotype or bioactivity, to rationalize possible side-effects, to predict off-targets and to assess the possibility of repurposing therapeutically-relevant compounds.

SwissTargetPrediction is based on the so-called 'similarity principle', which generally states that two similar molecules are prone to have similar properties. In the context of SwissTargetPrediction, we have statistically quantified that two similar bioactive molecules are likely to share their protein targets. Therefore, for a query molecule, we identify the most similar molecules among a set of about 370'000 actives on known targets. The predicted targets are those having the actives displaying the highest similarity with the query molecule.

Aim:

Drug Deferiprone targets screening and binding assessment.

Materials and Methods:

Uniprot: UniProt is a freely accessible database of protein sequence and functional information, many entries being derived from genome sequencing projects. It contains a large amount of information about the biological function of proteins derived from the research literature. It is maintained by the UniProt consortium, which consists of several European bioinformatics organizations and a foundation from Washington, DC, USA.

PDB: The Protein Databank (PDB) is a file format describing the 3D structure of molecules held in protein data bank. The PDB format accordingly provides for description and annotations of protein and nucleic acid structures including atomic co-ordinates, secondary structure assignments as well atomic connectivity.

Drug Bank: Drug bank database is a comprehensive, freely accessible online database containing information on drugs and drug targets. This database is a unique bioinformatics and chemo-informatics resource that combines detail drug.

Swiss Target Prediction: `Swiss target prediction is a online tool to predict the targets of bioactive small molecules in human and other vertebrates. This is useful to understand the molecular mechanism underlying a given phenotype or bioactivity.

Results:

JM21 gene is predicting target for Deferiprone drug.

Conclusion:

ENTRY	ENTRYNAME	PROTEIN NAME	GENE NAME	ORGANISM
Q9UBN7	HDAC6_HUMAN	Histone deacetylase-6	HDAC6, KIAA0901, JM21	Homo sapiens (human)

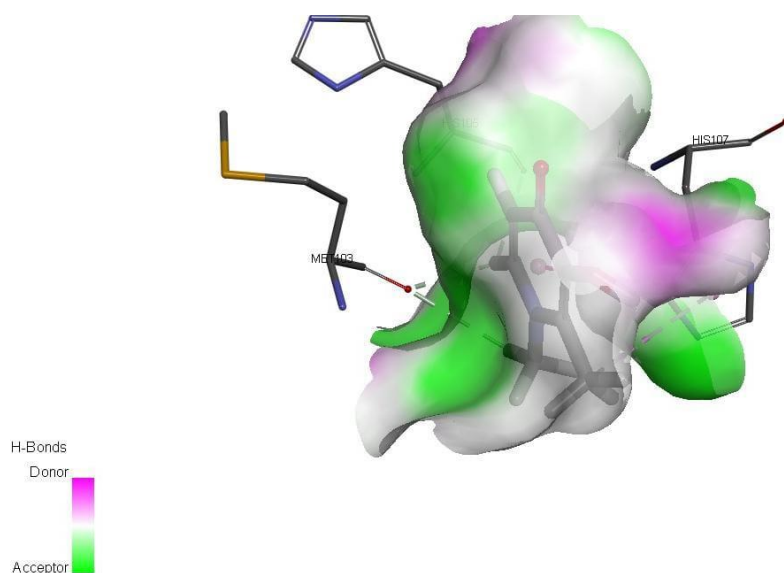


Fig 4.1

Name	Distance	Category	Types	From	To
2972:H11 - A:MET103:O	2.28243	Hydrogen Bond	Carbon Hydrogen Bond	2972:H11	A:MET103:O
2972:H11 - A:HIS105:O	2.60883	Hydrogen Bond	Carbon Hydrogen Bond	2972:H11	A:HIS105:O
2972:H16 - A:MET103:O	2.91902	Hydrogen Bond	Carbon Hydrogen Bond	2972:H16	A:MET103:O
A:HIS107 - 2972	4.52112	Hydrophobic	Pi-Pi T-shaped	A:HIS107	2972
A:HIS107 - 2972:C7	4.38782	Hydrophobic	Pi-Alkyl	A:HIS107	2972:C7

Non bond

name	-plp1	-plp2	jain	-pmf	dockscore
3C5K					
2972	31.24	21.64	-0.75	2.77	17.44
2972	31.24	21.64	-0.75	2.77	17.44
2972	31.24	21.64	-0.75	2.77	17.44
2972	31.17	21.31	-0.79	3.15	17.149
2972	31.17	21.31	-0.79	3.15	17.149
2972	30.94	20.99	-0.6	4.73	16.757
2972	30.93	20.9	-0.88	4.93	16.663
2972	30.93	21.65	-0.6	2.71	16.406
2972	30.93	21.64	-0.6	2.71	16.406
2972	24.76	16.93	-0.34	5.06	15.624

Molecule

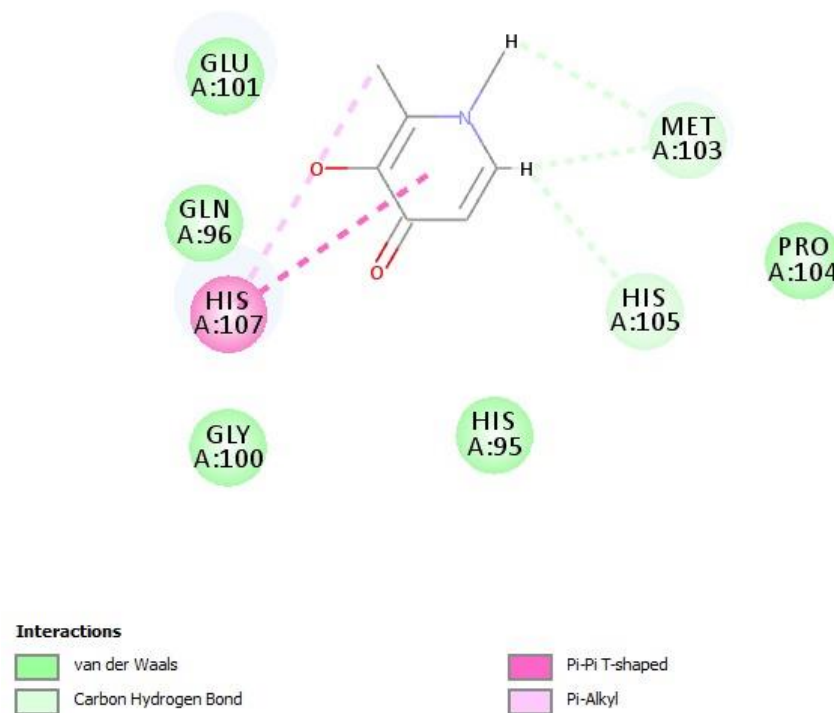


Fig 4.2

Discussion:

Deferiprone is a medication used to remove excess iron from the body, particularly in conditions like thalassemia. The interaction with HIS A:107 refers to the histidine at position 107 in a protein called HFE, which is involved in regulating iron absorption. This specific interaction isn't well-documented, but variations in the HFE gene (including at position 107) can affect how the body absorbs and regulates iron.

The interaction between deferiprone and GLN A:96 likely refers to the glutamine amino acid at position 96 in a protein. In the case of deferiprone, it primarily acts as an iron chelator, binding to excess iron in the body and allowing it to be excreted. However, it's possible that variations in proteins involved in iron metabolism, such as those containing glutamine at position 96, could influence how deferiprone is absorbed, distributed, metabolized, or eliminated in the body.

The interaction of deferiprone with GLU A:101 refers to a specific amino acid variant at position 101 of a protein. Certain amino acid variants in proteins involved in iron regulation could potentially affect how deferiprone functions. Glutamic acid residues can be involved in metal ion binding and coordination, so alterations in this amino acid position might influence the binding affinity of deferiprone to its target proteins or affect its overall effectiveness in chelating iron.

The interaction of deferiprone with MET A:103 refers to the methionine at position 103 in a protein. Methionine is a common amino acid in proteins and plays various roles in protein structure and function. The interaction of deferiprone with methionine at position 103 would likely depend on the specific protein and its function within the body.

The interaction of deferiprone with PRO A:104 refers to a specific variant of the protein known as proline at position 104. This variant is involved in the regulation or function of a protein related to iron metabolism or absorption. Deferiprone is a chelating agent, which means it binds to metals like iron and helps remove them from the body. Variations in proteins involved in iron metabolism can potentially influence how deferiprone is metabolized or its effectiveness in removing excess iron.

The interaction between deferiprone and histidine at position 105 (HIS A:105) likely pertains to the HFE gene, which encodes a protein involved in regulating iron absorption. Variations in the HFE gene, including at position 105, can influence iron metabolism and contribute to conditions like hemochromatosis. However, the specific interaction between deferiprone and histidine at position 105 would depend on how this variant affects the function of the HFE protein and how deferiprone interacts with this altered protein.

The interaction of deferiprone with histidine at position 95 (HIS A:95) likely involves its chelation properties. Deferiprone is a chelating agent, meaning it binds to metal ions like iron in the body, forming a stable complex that can then be excreted. Histidine residues are common binding sites for metal ions in proteins. In the context of iron overload disorders like thalassemia or hemochromatosis, deferiprone binds to excess iron in tissues and forms a complex that is then excreted from the body. Histidine residues, including HIS A:95, may play a role in facilitating this binding process.

The interaction of deferiprone with GLY A:100 refers to a specific variation at position 100 of the amino acid sequence, where glycine (Gly) is present. This could potentially relate to the binding or activity of deferiprone, as amino acid variations can sometimes alter the function or binding affinity of proteins.

MATLAB COURSES

Signal processing onramp:

Signal Processing with MATLAB course provides comprehensive training on utilizing MATLAB for analyzing, visualizing, and manipulating signals, equipping learners with essential skills for tackling real-world signal processing challenges.

Non linear equations:

Solving Nonlinear Equations with MATLAB course equips learners with the skills to tackle complex mathematical problems through MATLAB's powerful computational capabilities. We will be able to explore various numerical methods and optimization techniques to efficiently solve nonlinear equations, enhancing their problem-solving proficiency in MATLAB.

Image processing onramp:

The IMAGE PROCESSING ONRAMP course in MATLAB offers a comprehensive introduction to fundamental image processing techniques, equipping us with skills in image manipulation, enhancement, and analysis through hands-on exercises and interactive tutorials. This course is ideal for beginners and provides a solid foundation in MATLAB's image processing capabilities, empowering us to efficiently process and analyze images for various applications.

Matlab fundamentals:

The MATLAB Fundamentals course provides a comprehensive introduction to MATLAB, covering basic syntax, data types, plotting, and programming concepts. We gain proficiency in utilizing MATLAB for data analysis, visualization, and algorithm development through hands-on exercises and practical examples.

Matlab programming techniques:

Through this course, we will be able to learn advanced MATLAB programming techniques, including optimization, simulation, and data analysis, through hands-on projects and real-world applications.

Machine learning :

The Machine Learning with MATLAB course offers comprehensive instruction in utilizing MATLAB's powerful tools and algorithms for developing and deploying machine learning

models, empowering learners to tackle real-world data challenges effectively. We learn to apply different types of machine learning models for clustering, classification, and regression in MATLAB.

Data processing and visualization:

This MATLAB course focuses on data processing and visualization techniques, equipping us with essential skills to analyze and present data effectively. We will gain hands-on experience in utilizing MATLAB's powerful tools for statistical analysis, plotting, and interpreting data sets.

Deep learning onramp:

The Deep Learning with MATLAB course provides comprehensive instruction on leveraging MATLAB's tools for network development and training, enabling us to master the principles and applications of deep learning within a MATLAB environment. Through hands-on exercises and tutorials, we gain practical skills in implementing and optimizing deep learning algorithms for various tasks such as image classification, natural language processing, and more.

The Reinforcement Learning Onramp

This course in MATLAB provides a comprehensive introduction to the principles and applications of reinforcement learning, equipping learners with hands-on experience through practical exercises and examples. Through interactive modules, we gain proficiency in implementing RL algorithms and deploying them in real-world scenarios using MATLAB.

Computer vision onramp:

The Computer Vision Onramp course offers a comprehensive introduction to the principles and techniques of computer vision, covering topics such as image processing, feature extraction, object detection, and deep learning for vision tasks. We will learn about the role of features in computer vision, how to label data, train an object detector, and track wildlife in video.

Machine learning onramp:

The Machine Learning Onramp course offers a comprehensive introduction to the fundamentals of machine learning, covering key concepts, algorithms, and practical applications. We learn the basics of practical machine learning for classification problems in MATLAB and use a machine learning model that extracts information from real-world data to group our data into predefined categories.

Object oriented programming onramp:

The MATLAB Object-Oriented Programming Onramp course offers an introduction to principles and implementation of OOP within the MATLAB environment, focusing on creating and manipulating objects, classes. Through hands-on exercises, we learn to leverage classes, inheritance, and polymorphism to enhance code modularity and maintainability.

Matlab onramp:

The MATLAB Onramp course is a comprehensive introduction to MATLAB, covering essential topics like data analysis, visualization, and programming. It's designed for beginners to quickly grasp MATLAB's fundamentals. We can learn about the basic syntax, array types, array indexing in this course.

Image processing matlab:

The IMAGE Processing with MATLAB course provides a comprehensive introduction to image processing techniques using MATLAB, covering fundamentals like image enhancement, segmentation, and feature extraction, equipping us with practical skills for manipulating and analyzing images programmatically.

Statistics onramp:

Statistics Onramp is a MATLAB course designed to teach fundamental statistical concepts and how to implement them using MATLAB's powerful tools, catering to both beginners and those looking to enhance their statistical analysis skills in MATLAB.

Simulink onramp:

The Simulink Onramp course in MATLAB provides a comprehensive introduction to Simulink, MATLAB's graphical programming environment, enabling users to design, simulate, and analyze dynamic systems efficiently. Through interactive modules, we gain hands-on experience in modeling and simulating systems, paving the way for further exploration in engineering and scientific fields.

MATLAB CERTIFICATES:



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

MATLAB Onramp

A handwritten signature in black ink, reading "Craig L. Shenton".

DIRECTOR, TRAINING SERVICES

12 November 2023



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Object-Oriented Programming Onramp

A handwritten signature in black ink, reading "Craig L. Shenton".

DIRECTOR, TRAINING SERVICES

20 December 2023



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Machine Learning Onramp


DIRECTOR, TRAINING SERVICES

21 December 2023

1



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Computer Vision Onramp


DIRECTOR, TRAINING SERVICES

22 December 2023



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Reinforcement Learning Onramp



DIRECTOR, TRAINING SERVICES

17 February 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Deep Learning with MATLAB



DIRECTOR, TRAINING SERVICES

4 March 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

MATLAB for Data Processing and Visualization

A handwritten signature in black ink, reading "Craig T. Santos".

DIRECTOR, TRAINING SERVICES

5 March 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Machine Learning with MATLAB

A handwritten signature in black ink, reading "Craig T. Santos".

DIRECTOR, TRAINING SERVICES

5 March 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

MATLAB Programming Techniques


DIRECTOR, TRAINING SERVICES

7 March 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

MATLAB Fundamentals


DIRECTOR, TRAINING SERVICES

7 March 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Image Processing Onramp

A handwritten signature in black ink, reading "Craig L. Santos".

DIRECTOR, TRAINING SERVICES

7 March 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Solving Nonlinear Equations with MATLAB

A handwritten signature in black ink, reading "Craig L. Santos".

DIRECTOR, TRAINING SERVICES

7 March 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Signal Processing Onramp



Craig L. Santos

DIRECTOR, TRAINING SERVICES

13 March 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Simulink Onramp



Craig L. Santos

DIRECTOR, TRAINING SERVICES

30 March 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Signal Processing with MATLAB



DIRECTOR, TRAINING SERVICES

7 April 2024



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Image Processing with MATLAB



DIRECTOR, TRAINING SERVICES

29 April 2024

:



Course Completion Certificate

Nikhitha Sebastian T

has successfully completed **100%** of the self-paced training course

Statistics Onramp

A handwritten signature in black ink, reading "Craig L. Santos".
DIRECTOR, TRAINING SERVICES

2 March 2024