

“Drug-Target Interaction Prediction”

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

By

GOWSALYA.S (Reg No. 43739015)



DEPARTMENT OF BIOINFORMATICS AND DATA SCIENCE

SCHOOL OF BIO AND CHEMICAL ENGINEERING

SATHYABAMA

**INSTITUTE OF SCIENCE AND
TECHNOLOGY**

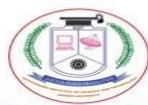
(DEEMED TO BE UNIVERSITY)

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

UGC | Approved by AICTE

JEPPIAAR NAGAR, RAJIV GANDHI

SALAI, CHENNAI - 600119



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Category - I University by UGC

Accredited "A++" by NAAC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS AND DATASCIENCE

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **GOWSALYA.S(43739015)** who carried out the Project entitled **“Drug-Target Interaction Prediction”** under my supervision from November 2023 to April 2024

DECLARATION

I.G O W S A L Y A . S (4 3 7 3 9 0 1 5) here by declare that the Project Report entitled **“Drug-Target Interaction Prediction”** 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:

GOWSALYA.S

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.

(GOWSALYA.S)

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 Diagnosis 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 plays a pivotal role in understanding the intricate mechanisms underlying various biological processes and diseases. This study focuses on employing advanced computational techniques to analyze gene expression data and identify potential biomarkers for diagnostic, prognostic, and therapeutic purposes. Through comprehensive bioinformatics pipelines, differential gene expression patterns are elucidated, highlighting key genes dysregulated in specific conditions. Integration of multiple omics data further enriches the analysis, offering a holistic view of molecular interactions. 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.

The identification of drug-target interactions (DTIs) is a crucial step in drug discovery and development processes. Traditional experimental methods for DTI prediction are time-consuming, costly, and often yield limited coverage. In recent years, computational approaches have emerged as promising alternatives to predict DTIs efficiently and accurately. This abstract provides an overview of computational methods for DTI prediction, highlighting various strategies such as machine learning algorithms, network-based methods, and molecular docking simulations. Additionally, challenges such as data sparsity, feature selection, and model interpretability are discussed, along with potential solutions. Moreover, the integration of multi-omics data and the utilization of deep learning techniques are explored as avenues for improving prediction performance. Finally, applications of DTI prediction in drug repurposing, side effect prediction, and personalized medicine are outlined, underscoring the significant impact of computational approaches in accelerating drug discovery pipelines.

Gowsalya S

ORIGINALITY REPORT

5%	1%	1%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Colorado Technical University Online Student Paper	1%
2	Hansaim Lim, Lei Xie. "Chapter 11 Omics Data Integration and Analysis for Systems Pharmacology", Springer Science and Business Media LLC, 2019 Publication	1%
3	Submitted to University of Wales Institute, Cardiff Student Paper	1%
4	Submitted to October University for Modern Sciences and Arts (MSA) Student Paper	1%
5	www.researchgate.net Internet Source	1%
6	www.techsciresearch.com Internet Source	1%

SYSTEM PHARMACOLOGY

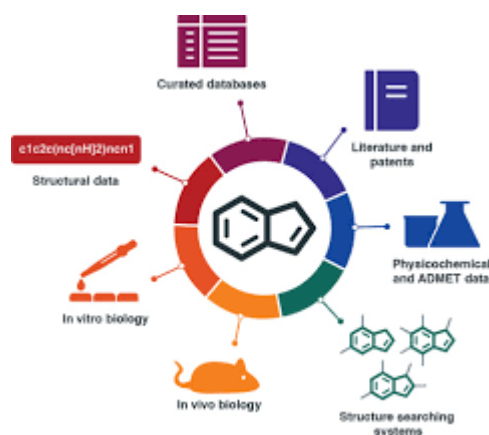
"HARNESSING BIG DATA FOR DRUG DEVELOPMENT: SYSTEMSPHARMACOLOGY PERSPECTIVES"

Introduction :

Systems pharmacology seeks to enhance clinical practice and drug development by providing a comprehensive understanding of drug action mechanisms. Systems pharmacology modeling, or SPM, is driven by data. Data from several scales are included, including hereditary, molecular, cell-based, organismal, and ecological it is rapidly growing. The goal of SPM is to develop interpretable and useful mechanistic or predictive multiscale models. The current eruptions in genomics and other omics data, coupled with the amazing progress in big data methods, have allowed scientists to generate novel hypotheses and learn novel concepts through computational simulations of genome-wide, diverse, and dynamic data sets.

The action of drugs is a complex process. The properties of a chemical interactions with biomolecules (proteins, DNA, or RNA), also known as its targets, are the first step in how it affects biological systems. A chemical can be synthetic, organic, or endogenous. For the most often studied protein targets, protein structural dynamics control a broad spectrum of relationships, such as biased signaling, inverse agonism, incomplete agonism, and allosteric modulation.

Modifications in the functional condition of the biomolecule which depend on the kinetics of ligand attachment and separating as well as the subsequent thermodynamics transitions are ultimately. The creation of individualized treatment plans based on each patient's unique genetic, molecular, and clinical characteristics is made possible by big data. By customizing medication regimens for certain patient groups, systems pharmacology enhances therapeutic results and reduces side effects. responsible for the biological effects. Furthermore, a chemical almost never attaches to a single target.



Data Integration and Analysis in Systems Pharmacology :

Data integration and analysis play a key role in systems pharmacology by providing insights into intricate biological systems and medication activities. Compiling various data types such as cellular, molecular, and clinical data allows for a comprehensive comprehension of disease causes and therapeutic effects. Integrating data from different sources like literature databases, omics datasets, and electronic health records—is known as data integration. In order to standardize data formats, address discrepancies, and connect diverse data parts, this frequently calls for the application of sophisticated bioinformatics and computational approaches. By use of integration, scientists can unearth obscure connections, pinpoint biomarkers, and clarify the mechanisms behind medication reaction and illness advancement.

Information of human disease, adverse reactions to medications, and mice phenotypes can be linked with data reporting on how pharmaceuticals alter the phenotypic of human cell lines and how drugs produce changes in gene and protein expression in human cell lines. The translation of data from the various resources into multipartite charts, gene-set libraries, or single-node-type networks might facilitate such data integrating efforts. This strategy can help us benchmark computational and experimental techniques and uncover further connections between genes, medications, and phenotypes.

Harmonizing data from various sources, including omics datasets, electronic medical records, and literary databases, is the process of data integration. In order establish uniform data formats, address discrepancies, and connect diverse data parts, this

frequently calls for the application of sophisticated bioinformatics and computational approaches. By use of integration, scientists can unearth obscure connections, pinpoint biomarkers, and clarify the mechanisms behind medication reaction and illness advancement. System pharmacology uses methods of dynamic simulation to mimic the impacts of drugs in biological systems. Pharmacokinetic-pharmacodynamic (PK-PD) and quantitative systems pharmacology (QSP) models are two examples of computational frameworks that make it possible to predict drug behavior over time and at various biological scales. It is possible for researchers to prioritize therapeutic candidates, optimize drug dosing regimens, and reduce side effects by combining data from experiments with computational models.

Pharmacovigilance and Real-World Evidence in Drug Safety

Real-world evidence (RWE) and pharmacovigilance are essential for guaranteeing medication safety at every stage of a pharmaceutical product's life cycle. The systematic observation, identification, evaluation, and avoidance of side effects or any other medication-related issues are all included in pharmacovigilance. RWE, on the other hand, refers to information obtained from actual sources that offer insights on medication safety and efficacy in actual clinical settings. RWE complements pharmacovigilance by providing longitudinal data on drug utilization, treatment patterns, and patient outcomes in routine clinical practice.

By leveraging large-scale databases and advanced analytics, RWE enables the identification of safety signals, the evaluation of drug effectiveness, and the assessment of risk-benefit profiles in real-world populations.



This real-world evidence helps regulators, healthcare providers, and pharmaceutical companies make informed decisions regarding drug labeling, post-marketing surveillance, and risk management strategies. Examples of these sources include electronic health records, claims databases, and patient registries. Pharmacovigilance is instrumental in communicating safety information to healthcare professionals, regulatory authorities, and the general public. Timely dissemination of safety alerts,

risk minimization measures, and product labeling updates ensures that healthcare providers and patients are well-informed about the latest safety information.

Regulatory Considerations in Data-Driven Drug Discovery :

Systems pharmacology's development and application of data-driven drug discovery methodologies are heavily influenced by regulatory concerns. Researchers must navigate a regulatory environment that guarantees the safety, efficacy, and moral application of developing therapies as they leverage the power of varied datasets and computational approaches to clarify intricate biological connections. A critical component of regulatory considerations is the integrity and quality of data. Strict guidelines for data gathering, archiving, and processing are required by regulatory bodies to guarantee the accuracy and repeatability of study results. To preserve data integrity and comply with regulatory standards, data-driven drug development initiatives must follow good laboratory procedures (GLP) and good clinical practices (GCP).

Regulatory agencies such as the FDA and EMA increasingly recognize the value of data-driven approaches in drug discovery and development. However, they also emphasize the importance of transparent and interpretable methodologies to validate computational models and ensure their reliability for decision-making. Therefore, researchers must employ robust validation strategies and provide clear documentation of their data sources, algorithms, and assumptions to facilitate regulatory scrutiny and approval.

regulatory considerations are integral to the successful application of data-driven drug discovery in systems pharmacology. Adherence to data quality standards, transparency in methodology, ethical conduct, and comprehensive evidence generation are essential for navigating the regulatory landscape and bringing innovative therapeutics to market.

Additionally, regulatory systems guiding data-driven drug discovery heavily weigh ethical issues. The safeguarding of patient data is required under privacy laws, including as the GDPR in Europe and HIPAA in the US, which also demand informed consent before using personal health data for research purposes. Adherence to these

standards is crucial in order to protect patient confidentiality and preserve public confidence in research ventures that leverage data.

Ethical and Privacy Concerns in Big Data Utilization for Pharmacological Research :

When using big data for pharmacological research in the field of systems pharmacology, ethical and privacy issues are major considerations. Utilizing enormous volumes of data from various sources, including wearable technology, genomic databases, and electronic health records, researchers must resolve difficult moral conundrums and guarantee the privacy of individuals. Informed consent and data ownership are two main areas of concern. Sensitive information from patient medical records or genetic profiles is frequently included in big data. Upholding ethical standards requires making sure people are fully informed about how their data will be used and getting their express agreement for research purposes. However, there are difficulties in maintaining the usefulness of data for research while anonymizing it to safeguard privacy, particularly in datasets with abundant clinical information. Data security and breach risks are brought up by the aggregation and integration of large data from many sources. Maintaining patient trust and adhering to privacy requirements depend on protecting data from illegal access, cyberattacks, and unintentional exposures. Strong data encryption, access controls, and data anonymization strategies can be put into place to assist reduce these dangers and safeguard people's right to privacy.

The possibility of discrimination and bias in data is another ethical issue. Research findings can be distorted and health inequities can be made worse by inherent biases in big data, such as the overrepresentation of particular populations or the underrepresentation of certain demographic groups. By using the right data preprocessing procedures and making sure that study groups are diverse and representative, researchers can mitigate biases with caution.

Concerns of data accountability and transparency are brought up by the use of big data in pharmaceutical research. To ensure scientific integrity and reproducibility, researchers should be open and honest about their data sources, analytical techniques,

and potential conflicts of interest. Furthermore, procedures for impartial supervision and peer review might support ensuring the accuracy and dependability of study results.

When using big data for pharmacological research in systems pharmacology, ethical and privacy issues are crucial. Promoting ethical standards and appropriate data-driven research procedures that put patient privacy and welfare first need adherence to informed consent, data security, fairness, transparency, and accountability principles.

REFERENCES:

Lei Xie, Eli J. Draizen, and Philip E. Bourne "Harnessing Big Data for Systems Pharmacology" The Graduate Center, The City University of New York, New York, NY 10016

Nussinov R, Tsai CJ, Csermely P. Allo-network drugs: harnessing allostery in cellular networks. *Trends Pharmacol Sci.* 2011;32:686–93.

Kenakin T, Christopoulos A. 2013. Signalling bias in new drug discovery: detection, quantification and therapeutic impact. *Nat. Rev. Drug Discov.* 12:205–16

Andreas Bender Dr., Josef Scheiber Dr., Meir Glick Dr., John W. Davies Dr., Kamal Azzaoui Dr., Jacques Hamon Dr., Laszlo Urban Dr., Steven Whitebread, Jeremy L. Jenkins Dr. Adam Lavertu, Bianca Vora, Kathleen M. Giacomini, Russ Altman, Stefano Rensi."Analysis of Pharmacology Data and the Prediction

Xie L, Ge X, Tan H, Xie L, Zhang YL, et al. Towards structural systems pharmacology to study complex disease and personalized medicine. *PLOS Comput Biol.* 2014;10:e1003554.

irst published: 31 May 2007 2013 Early definitions of Big Data in healthcare, emphasizing the importance of volume, velocity, variety, veracity, and value (Marx, 2013; Mateosian, 2013; McAfee & Brynjolfsson, 2012

TRANSLATIONAL BIOINFORMATICS

1.GENE EXPRESSION ANALYSIS

1.1 Introduction:

A Biological network represents systems as complex set of binary interactions or relation between various biological systems and can reveal informative patterns within the systems. Cytoscape is a open source software platforms for visualizing molecule interactions network and biological pathway and integrating the network with annotations, gene expression profile. Cytohubba predict and explore the important nodes and edges in given network by several topological algorithm such as degree, bottleneck, MCC, DMNC, MNC, EPC, Eccentricity, Closeness, Stress, Clustering co-efficient and betweenness.

1.2 Aim: Build Network to understand gene interrelationship and regulatory mechanism.

1.3 Materials and Methods:

GEO2R is an interactive web tool that allows users to compare two or more groups of samples in a GEO series in order to identify genes that are differentially expressed across experimental conditions.

CYTOSCAPE: Cytoscape is a open source software platforms for visualizing molecule interactions network and biological pathway and integrating the network with annotations, gene expression profile.

CYTOHUBBA: CytoHubba provides 11 topological analysis methods including Degree, Edge Percolated Component, Maximum Neighborhood Component, Density of Maximum Neighborhood Component, Maximal Clique Centrality and six centralities (Bottleneck, Eccentricity, Closeness, Radiality, Betweenness, and Stress) based on shortest paths.

CLUEGO: ClueGO is 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.

1.4 Procedure:

- For the database and disease Thyroid Carcinoma Geo2R is the source of the geo dataset (GSE53157)
- After choosing the suitable parameters, the Geodataset is retrieved.
- Genes that were upregulated genes were copied and paste into the cytoscape network.
- Built a biological network using a BioGrid in cytohubba to analyse the network and identify importance node based various topologies algorithm to determine the the top 10 gens
- Genes must be copied in cluego and contigue cluego param3eter such as species ontology and statistical settings.
- Run cluego analysis to identify functional categories or pathways associated with the selected nodes.

1.5 Results:

Thyroid Carcinoma genes are differentially expressed

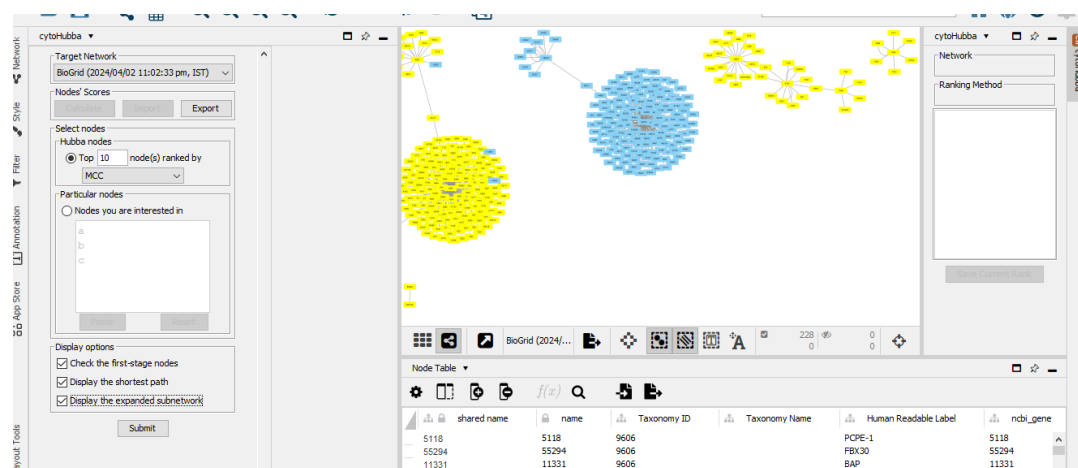


Fig:1.1



Fig:1.2

Interpretation:

Gene expression analysis is a process that involves examining the activity of genes within a cell or organism. This analysis can provide insights into how genes are regulated and how they contribute to biological processes. There are several methods for gene expression analysis, including microarray analysis and RNA sequencing.

"U2-type prespliceosome" refers to a complex of proteins and RNA molecules involved in the splicing of pre-messenger RNA (pre-mRNA) during gene expression. This complex is part of the spliceosome machinery, which removes introns (non-coding regions) from pre-mRNA transcripts to produce mature mRNA that can be translated into proteins.

The 14.29% enrichment value indicates the proportion of genes associated with the Notch receptor signaling pathway within your dataset. Essentially, it suggests that

approximately 14.29% of the genes in your dataset are involved in or related to the Notch receptor signaling pathway.

A 14.29% enrichment indicates that a subset of proteins associated with monoubiquitination is overrepresented compared to what would be expected by chance alone. This enrichment suggests that monoubiquitination may play a specific and potentially important role in the biological processes or pathways being analyzed.

Interpreting this result requires considering the context of the study and the specific proteins involved. For example, monoubiquitination of certain proteins may regulate their stability, activity, or subcellular localization, thereby influencing key cellular processes such as signal transduction, DNA repair, or protein degradation

2.Cumulative survival probability calculation using KM Plotter

2.1 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 logrank 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.

2.2 AIM:

Calculate Proposed gene ECT2 cumulative survival probabilities

2.3 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.

2.4 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

2.5 Interpretation

- KM-plot recognizes **proposed gene symbols ECT2**(including HUGO Gene Nomenclature Committee approved
- Sources for the database include GEO, EGA, TCGA, Metabric, Impact, and PubMed repositories. For more details, please check our publications.

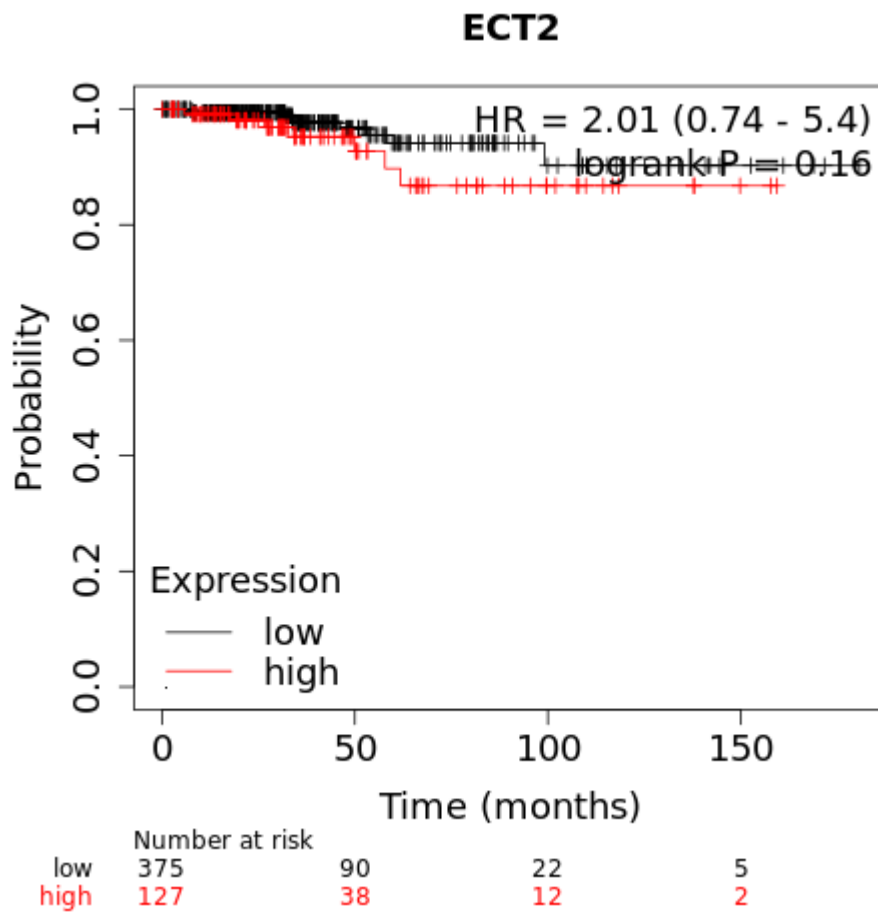


Fig:2.1

Hazard Ratio for ECT2 is 2.01 and low and high expression cohort is Not available

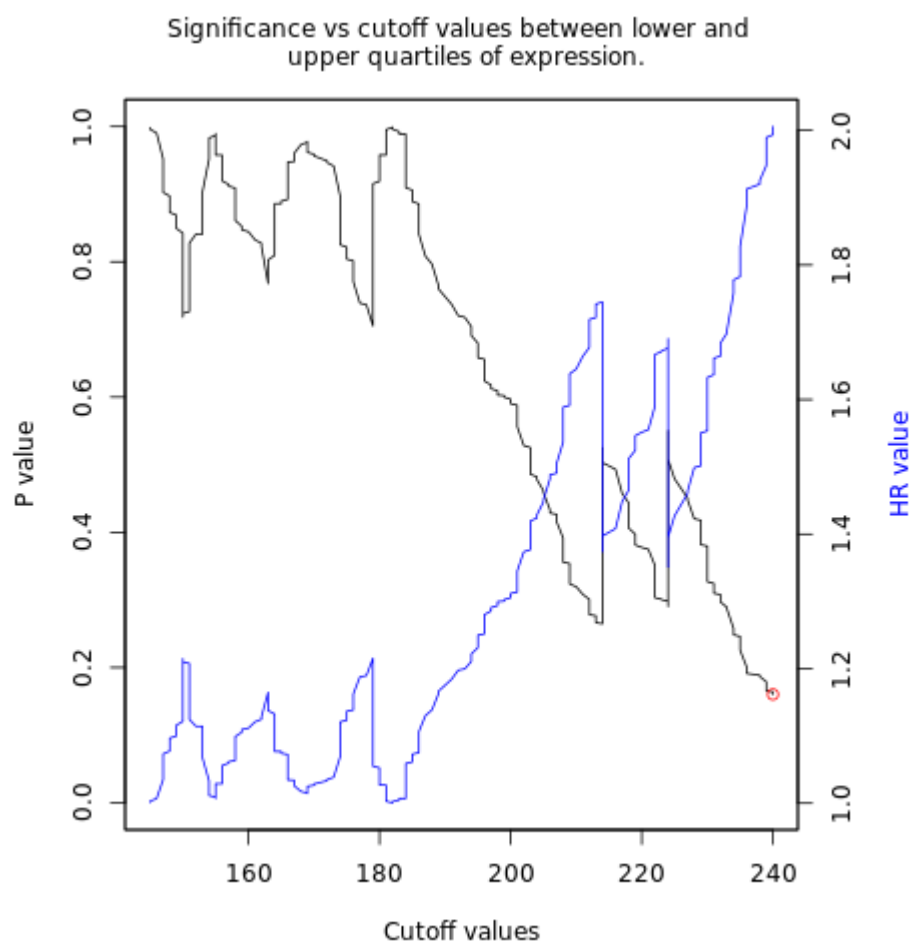


Fig:2.2

Low expression cohort (months)	High expression cohort (months)
NA	NA

3.Biomarkers: identifying biological markers for disease Diagnosis and therapeutic monitoring

3.1 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.

3.2 Aim :

Identify the disease markers related to tuberculosis and categorize them into pharmacodynamic Biomarker, Monitoring biomarker, safety biomarkers

3.3 Procedure:

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

3.4 RESULTS:

The biological marker are identified for disease diagnosis.

3.5 INTERPRETATION:

General Information of the Biomarker

BIOMARKER ID	BIOMARKER NAME	CLASS	GENE NAME	BIOMARKER TYPE	REPRESENTATIVE DRUG	REPRESENTATIVE DISEASE
B3HTD6	C-Reactive protein	protein	CRP;PTX1	PRD SAF PDY MOI	Anti-Tuberculosis	Tuberculosis

Drug Name: Isoniazid
Benzalkonium
Nivolumab
Pembrolizumab
Trastuzumab

Application of the Biomarker in Diagnosis and Prognosis:

Diagnosis: Acute respiratory illness, Alzheimer's disease.

Prognosis: Acute decompensated heart failure, Acute heart failure.

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

General Information of the Biomarker (Biomarker ID: B3HTD6)

Biomarker Name

C-reactive protein

Biomarker Class

Protein

Gene Name

CRP; PTX1

Synonyms

C reactive protein; PTX1; C-reactive protein

view as list

Biomarker Type

PDY

SAF

MOI

PRD

SUR

Sequence

> B3HTD6 C-reactive protein

MEKLLCFLVL TSLSHAFGQT DMSRKAFVFP KESDTSYVSL KAPLTKPLKA FTVCLHFYTE

LSSTRGYSIF SYATKRQDNE ILIFWSKDIG YSFTVGGSEI LFEVPEVTVA PVHICTSWES

Show More

Function

Displays several functions associated with host defense: it promotes agglutination, bacterial capsular swelling, phagocytosis and complement fixation through its calcium-dependent binding to phosphorylcholine. Can interact with DNA and histones and may scavenge nuclear material released from damaged circulating cells.

External Links

Gene ID

1401

UniProt ID

CRP_HUMAN

HGNC ID

HGNC:2367

Show More

Related Biomarker

mRNA biomarker

CRP mRNA (B7X3OM)

Fig:3.1

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

4.1 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 quantify 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.

4.2 Aim:

Drug Isoniazid targets screening and binding assessment.

4.3 Materials and Methods:

Swiss Target Prediction:

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 organisations 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 chemi-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.

4.4 PROCEDURE

Disease/Disorder ----- associated FDA approved drugs was chosen and smiled notation was retrieved from Pubchem using following link

<https://pubchem.ncbi.nlm.nih.gov/>

Swisstarget home page was accessed using following link

<http://www.swisstargetprediction.ch/index.php>

Predicted the targets (Top five) and drug interaction was analyzed using discovery studio Receptor ligand interaction module

4.5 Results:

KasA gene is predicting target for Isoniazid drug

4.6 Conclusion:

ENTRY	ENTRYNAME	PROTEIN NAME	GENE NAME	ORGANISM
P68363	TBA1B_HUMAN	Tubulin alpha-1B	TUBA1B	Homosapiens(human)

		chain		
--	--	-------	--	--

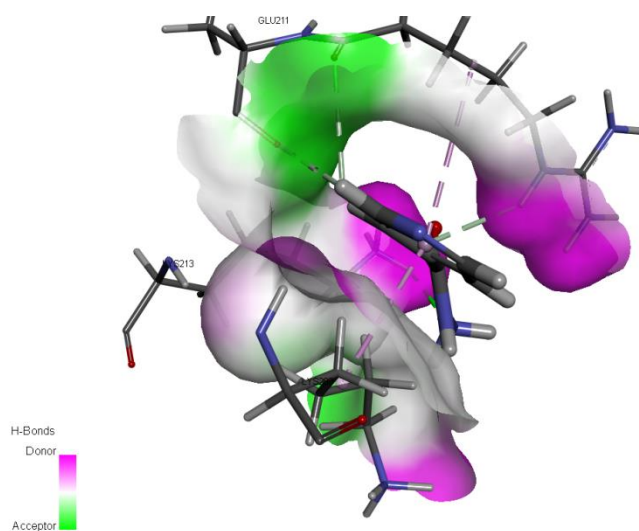


Fig:4.1

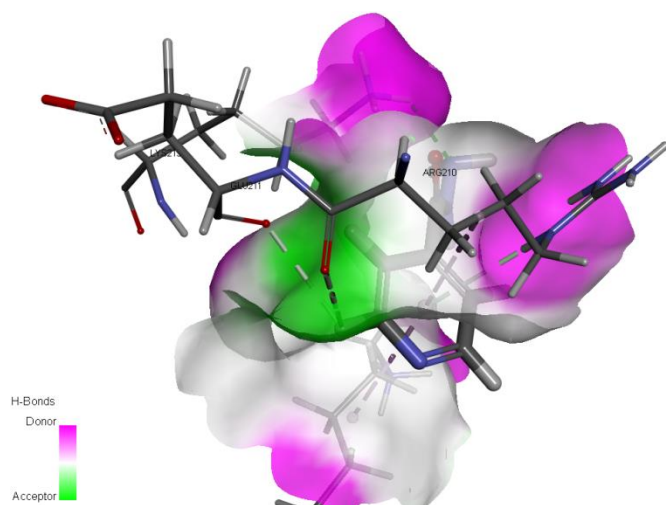


Fig:4.2

name	-plp1	-plp2	jain	-pmf	dockscore
2E4H					
3767	28.58	23.16	-0.12	37.4	23.061
3767	29.59	23.54	-0.26	36.41	22.745
3767	27.58	22.78	-0.18	33.72	22.661
3767	28.81	23.2	-0.21	34.54	22.661
3767	30.44	23.49	-0.4	35.98	22.636
3767	33.28	25.89	-0.24	33.55	22.615
3767	33.28	25.89	-0.24	33.55	22.615
3767	33.21	25.81	-0.23	33.28	22.614
3767	33.21	25.81	-0.23	33.28	22.614
3767	29.72	23.92	-0.18	34.24	22.593

Molecule

name	distance	types	category	from	to
A:LYS213:HZ1 - 3767:O1	1.88063	Hydrogen Bond	Conventional Hydrogen Bond	A:LYS213:HZ1	3767:O1
A:LYS213:HZ3 - 3767:N4	2.15835	Hydrogen Bond	Conventional Hydrogen Bond	A:LYS213:HZ3	3767:N4
3767:H14 - A:ARG210:O	3.01144	Hydrogen Bond	Carbon Hydrogen Bond	3767:H14	A:ARG210:O
3767:H14 - A:GLU211:O	2.90623	Hydrogen Bond	Carbon Hydrogen Bond	3767:H14	A:GLU211:O
A:ARG210:HE - 3767	2.71791	Hydrogen Bond	Pi-Donor Hydrogen Bond	A:ARG210:HE	3767
3767 - A:ARG210	4.4292	Hydrophobic	Pi-Alkyl	3767	A:ARG210
3767 - A:LYS280	4.03534	Hydrophobic	Pi-Alkyl	3767	A:LYS280

Non bond

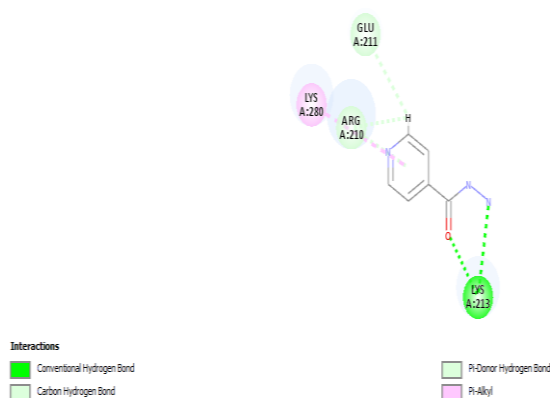


Fig:4.3

Isoniazid is a medication primarily used to treat tuberculosis (TB). It functions by inhibiting the synthesis of mycolic acids, essential components of the cell walls of the bacteria that cause TB.

Isoniazid could indirectly influence the amino acid GluA:211 through downstream effects on cellular processes. For example, if the protein in question is involved in the metabolism or transport of isoniazid, alterations in its function due to drug interaction could indirectly impact the GluA:211 residue.

The interaction between isoniazid and LYSa:280 likely refers to the potential effect of isoniazid on lysine residue 280 (LYSa:280) within a protein or enzyme. Isoniazid is a medication primarily used to treat tuberculosis by inhibiting the synthesis of mycolic acids, essential components of the bacterial cell wall. However, isoniazid is also known to have interactions with various cellular proteins and enzymes.

MATLAB CERTIFICATES:

Signal processing onramp

Non linear equations

Image processing onramp

Matlab fundamentals

Matlab programming techniques

Machine learning

Data processing and visualization

Deep learning onramp

Deep learning with matlab

Reinforcement learning onramp

Computer vision onramp

Machine learning onramp

Object oriented programming onramp

Matlab onramp

Image processing matlab

Statistics onramp

Simulink onramp

1. Matlab Fundamentals:



Course Completion Certificate

Gowsalya S

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

MATLAB Fundamentals

DIRECTOR, TRAINING SERVICES

31 March 2024

Matlab, short for Matrix Laboratory, is a high-level programming language and interactive environment primarily used for numerical computing and data analysis. It excels in matrix and vector operations, making it ideal for engineering and scientific applications. With its vast library of built-in functions and toolboxes, Matlab offers extensive capabilities for mathematical modeling, simulation, and algorithm development. Its intuitive syntax allows for rapid prototyping and easy implementation of complex algorithm.

2. Deep Learning With Matlab:



Course Completion Certificate

Gowsalya S

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

Deep Learning with MATLAB

DIRECTOR, TRAINING SERVICES

29 March 2024

Learn the basics of deep learning for image classification problems in MATLAB. Through MATLAB, you can grasp the foundational concepts of deep learning, including feedforward and convolutional neural networks, recurrent neural networks, and deep reinforcement learning. MATLAB offers a user-friendly environment with built-in functions and tools for developing, training, and evaluating deep learning models.

3.Solving Nonlinear Equations With Matlab



Course Completion Certificate

Gowsalya S

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

Solving Nonlinear Equations with MATLAB

DIRECTOR, TRAINING SERVICES

24 March 2024

Learning to solve nonlinear equations with MATLAB introduces fundamental numerical methods for finding roots or solutions to equations where the relationship between variables is nonlinear. MATLAB provides various techniques for solving such equations efficiently and accurately. One basic method is the use of built-in functions like `fsolve`, which employs iterative algorithms such as Newton's method or the Trust Region method to find solutions.

4. Matlab for Data Processing&Visualization



Course Completion Certificate

Gowsalya S

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

MATLAB for Data Processing and Visualization

RAY L. SANTOS
DIRECTOR, TRAINING SERVICES

14 March 2024

MATLAB serves as a powerful platform for data processing and visualization, offering a comprehensive toolkit for analyzing and presenting data in various fields. With its extensive library of functions and toolboxes, MATLAB enables users to preprocess raw data efficiently, including cleaning, filtering, and transforming datasets to extract meaningful insights. Its intuitive syntax and interactive environment facilitate exploratory data analysis, allowing users to interactively manipulate and visualize data to uncover patterns and trends. MATLAB's rich plotting capabilities enable the creation of a wide range of visualizations, from basic plots and histograms to advanced 3D visualizations and interactive dashboards.

5. Machine Learning With MATLAB



Course Completion Certificate

Gowsalya S

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

Machine Learning with MATLAB

Craig L. Hunter

DIRECTOR, TRAINING SERVICES

14 March 2024

Machine learning teaches computers to do what comes naturally to humans: learn from experience. Machine learning algorithms use computational methods to “learn” information directly from data without relying on a predetermined equation as a model.

6. Signal Processing With MATLAB



Course Completion Certificate

Gowsalya S

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

Signal Processing with MATLAB

Craig L. Hunter

DIRECTOR, TRAINING SERVICES

14 March 2024

In signal processing, Generate sampled and synthesized signals from the command line and visualize them. Create noise signals for a given specification. Perform signal processing operations like resampling, modulation, and correlation,

7. Matlab Programming Techniques



Course Completion Certificate

Gowsalya S

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

MATLAB Programming Techniques

Craig L. Hunter

DIRECTOR, TRAINING SERVICES

12 March 2024

Matlab programming techniques encompass a set of strategies and best practices aimed at optimizing code efficiency, readability, and maintainability within the Matlab environment. These techniques include modularization, where code is organized into functions and scripts to promote reusability and reduce redundancy. Vectorization is another crucial technique, emphasizing the use of array operations instead of iterative loops to enhance computational efficiency.

8.Statistics Onramp



Course Completion Certificate

Gowsalya S

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

Statistics Onramp

DIRECTOR, TRAINING SERVICES

2 March 2024

The Statistics Onramp in MATLAB is an interactive tutorial designed to introduce users to fundamental concepts and techniques in statistics using MATLAB. It covers a wide range of topics, including descriptive statistics, probability distributions, hypothesis testing, and regression analysis. The Onramp provides a hands-on learning experience with interactive exercises and quizzes to reinforce understanding.

9. Deep Learning Onramp



Course Completion Certificate

Gowsalya S

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

Deep Learning Onramp

Craig L. Santos
DIRECTOR, TRAINING SERVICES

9 January 2024

The Deep Learning Onramp in MATLAB offers a structured and interactive learning experience focused on introducing users to the principles and applications of deep learning. It provides a series of tutorials, exercises, and quizzes designed to cover fundamental concepts such as neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs). Through the Onramp, users gain hands-on experience in building and training deep learning models using MATLAB's extensive set of tools and functions.

10. Image Processing Onramp



Course Completion Certificate

Gowsalya S

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

Image Processing Onramp

DIRECTOR, TRAINING SERVICES

20 December 2023

The Image Processing Onramp in MATLAB offers an interactive tutorial designed to introduce users to the fundamental concepts and techniques of image processing. It provides a structured learning path covering topics such as image enhancement, segmentation, feature extraction, and morphological operations. Through hands-on exercises and examples, users gain practical experience in applying various image processing algorithms to solve real-world problems.

11.Reinforcement Learning Onramp



Course Completion Certificate

Gowsalya S

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

Reinforcement Learning Onramp


DIRECTOR, TRAINING SERVICES

20 December 2023

The Reinforcement Learning Onramp in MATLAB is an interactive tutorial designed to introduce users to the fundamentals of reinforcement learning (RL). Reinforcement learning is a machine learning paradigm where an agent learns to make decisions by interacting with an environment to maximize cumulative rewards. The Onramp provides a structured learning path that covers essential concepts such as Markov decision processes, policy iteration, and Q-learning algorithms.

12. Computer Vision Onramp



Course Completion Certificate

Gowsalya S

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

Computer Vision Onramp


DIRECTOR, TRAINING SERVICES

19 December 2023

The Computer Vision Onramp in MATLAB is an interactive learning platform designed to introduce users to the fundamentals of computer vision. It offers a structured curriculum covering essential concepts such as image processing, feature detection, object recognition, and deep learning techniques applied to visual data.

13. Machine Learning Onramp



Course Completion Certificate

Gowsalya S

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

Machine Learning Onramp

Craig L. Hunter
DIRECTOR, TRAINING SERVICES

19 December 2023

The Machine Learning Onramp provided by MATLAB is an interactive tutorial designed to introduce users to the fundamentals of machine learning. It offers a structured learning path covering essential concepts such as supervised and unsupervised learning, classification, regression, and clustering. Through hands-on exercises and practical examples, users gain proficiency in using MATLAB's built-in functions and toolboxes for data preprocessing, feature extraction, model training, and evaluation.

14. Object oriented Programmin Onramp



Course Completion Certificate

Gowsalya S

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

Object-Oriented Programming Onramp


DIRECTOR, TRAINING SERVICES

19 December 2023

The Object-Oriented Programming (OOP) Onramp in MATLAB provides an introduction to the principles and practices of object-oriented programming within the MATLAB environment. Object-oriented programming is a paradigm that revolves around the concept of objects, which encapsulate data and behavior. In MATLAB, OOP enables users to create classes and objects, allowing for the organization of code into modular, reusable components. The Onramp covers fundamental OOP concepts such as classes, objects, inheritance, encapsulation, and polymorphism, providing a structured approach to learning these concepts through interactive exercises and examples

15. Matlab Onramp



Course Completion Certificate

Gowsalya S

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

MATLAB Onramp

Craig L. Hunter
DIRECTOR, TRAINING SERVICES

10 November 2023

MATLAB Onramp is an interactive introductory tutorial designed to familiarize users with the basics of MATLAB, a high-level programming language and environment widely used in engineering, science, and beyond. It provides a structured learning path, guiding users through essential concepts such as variables, arrays, and basic operations.

16. Signal Processing Onramp



Course Completion Certificate

Gowsalya S

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

Signal Processing Onramp



DIRECTOR, TRAINING SERVICES

1 April 2024

The Signal Processing Onramp is an interactive tutorial provided by MATLAB that serves as an entry point for individuals looking to learn about signal processing. This onramp offers a structured learning experience, guiding users through fundamental concepts and techniques in signal processing.

17. Simulink Onramp



Course Completion Certificate

Gowsalya S

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

Simulink Onramp


DIRECTOR, TRAINING SERVICES

1 April 2024

The Simulink Onramp provides an introductory learning experience for individuals looking to delve into the world of dynamic system modeling and simulation using Simulink, a graphical programming environment in MATLAB. Through interactive tutorials and exercises, users gain a foundational understanding of Simulink's interface, tools, and capabilities. They learn to build and simulate dynamic systems using block diagrams, representing components and their interconnections graphically.