

**INTEGRATED ANALYSIS: GENE EXPRESSION, SURVIVAL, BIOMARKER, & DRUG TARGET**

Submitted in partial fulfilment of the required for the award of  
Master of Science Degree in Bioinformatics and Data Science

By

NIKITHA V (43739029)



**DEPARTMENT OF BIOINFORMATICS AND DATASCIENCE**

**SCHOOL OF BIO AND CHEMICAL ENGINEERING**

**SATHYABAMA**

**INSTITUTE OF SCIENCE AND TECHNOLOGY**

**(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade “A++” by NAAC I Approved by AICTE**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,**

**CHENNAI - 600 119**

**April – 2024**



# **SATHYABAMA**

**INSTITUTE OF SCIENCE AND TECHNOLOGY**

**(DEEMED TO BE UNIVERSITY)**

**Accredited with “A” grade by NAAC I 12B Status by UGC I Approved by AICTE**

**Jeppiaar Nagar, Rajiv Gandhi Salai,**

**Chennai – 600 119**

**[www.sathyabama.ac.in](http://www.sathyabama.ac.in)**

**DEPARTMENT OF BIOINFORMATICS AND DATA SCIENCE**

## **BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the Bonafide work of NIKITHA.V (43739029) who carried out the project entitled “Integrated Analysis: Gene Expression, Survival, Biomarkers, and Drug Targets” under my / our supervision from January 2024 to April 2024.

**Internal Guide**

**External Guide**

**Head of the Department**

**Submitted for Viva voce Examination held on\_\_\_\_\_**

## **DECLARATION**

I NIKITHA V (43739029) hereby declare that the project report entitled “INTEGRATED ANALYSIS: GENE EXPRESSION, SURVIVAL, BIOMARKER, AND DRUG TARGET” done by me under the guidance of \_\_\_\_\_ (internal) and \_\_\_\_\_ (external) at sathyabama institute of science and technology is submitted in partial fulfilment of the requirements for the award of masters of Engineering / Technology degree in bioinformatics.

**DATE:**

**SIGNATURE OF THE CANDIDATE:**

**PLACE:**

## ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to board of management of sathyabama for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to Dr. \_\_\_\_\_, Dean, school of bio and chemical engineering and Dr. \_\_\_\_\_ Head of the department , Dept. of Bioinformatics 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 \_\_\_\_\_ for his/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 who were helpful in many ways for the completion of the project.

## ABSTRACT

The gene expression omnibus (GEO) at the national center for biotechnology information (NCBI) is the largest fully public repository for high-throughput molecular abundance data, primarily gene expression data. The database has a flexible and open design that allows the submission, storage and retrieval of many data types. Cutaneous melanoma is one of the most aggressive and lethal skin cancers. It is greatly important to identify prognostic biomarkers to guide the clinical management. However, it is technically challenging for untrained researchers to process high dimensional profiling data and identify potential prognostic genes in profiling datasets. Biomarker gene expression of ccRCC was retrieved from GEO database and the Cancer Genome Atlas Kidney Clear Cell Carcinoma (TCGA-KIRC) database. Gene ontology (GO) analysis and protein–protein interaction (PPI) networks of biomarker genes were constructed using the Database for Annotation, Visualization, and Integrated Discovery tool. Kaplan–Meier analysis and receiver operating characteristic curve (ROC) analysis were performed to explore the prognostic and diagnostic roles of these genes. Molecular docking interpretation is one of the crucial part before determining the result. Docking is commonly used to study the biomolecular interaction, usually for protein-ligand interaction, and to study about the molecular mechanism. Analysis of molecular interaction can help user to determine the strengthened of docking results, besides free energy of binding. In this protocol, analysis of molecular interaction as well as the surface characteristic of receptor was discussed in detail. In addition, the visualization to obtain suitable pictures for publications also included. The entire protocol will spend more of less 2 hours.

## INDEX

| S.NO      | CONTENT                           | PAGE.NO |
|-----------|-----------------------------------|---------|
| <b>1.</b> | <b>MINI PROJECT</b>               |         |
|           | INTRODUCTION                      | 8       |
|           | AIM AND OBJECTIVE                 | 12      |
|           | MATERIALS AND METHODS             | 13      |
|           | RESULT DISCUSSION                 | 18      |
|           | CONCLUSION                        | 31      |
| <b>2.</b> | <b>NEWSARTICLE<br/>REVIEW</b>     |         |
| <b>3.</b> | <b>PALGARISM CHECK<br/>REPORT</b> |         |
| <b>4.</b> | <b>MATLAB<br/>CERTIFICATE</b>     |         |

| <b>S.NO</b> | <b>FIGURES</b>                                                                                     | <b>PAGE NO</b> |
|-------------|----------------------------------------------------------------------------------------------------|----------------|
| <b>1</b>    | Volcano plot                                                                                       | 18             |
| <b>2</b>    | Network build with top ten gene using cytoscape and created column filter with taxonomy ID (9606). | 19             |
| <b>3</b>    | Hub genes are ranked using cytohubba                                                               | 20             |
| <b>4</b>    | Top 10 ranked hub gene were plot as pie chart                                                      | 20             |
| <b>5</b>    | KM PLOT                                                                                            | 22             |
| <b>6</b>    | Auto cutoff plot                                                                                   | 22             |
| <b>7</b>    | Median survival                                                                                    | 22             |
| <b>8</b>    | Disease-specific expression of the biomarker                                                       | 26             |
| <b>9</b>    | 3D Structure                                                                                       | 27             |
| <b>10</b>   | 2D Structure                                                                                       | 28             |

## **INTRODUCTION:**

### **1. DIFFERENTIAL EXPRESSION ANALYSIS**

Modern molecular biology data present major challenges for the statistical methods that are used to detect differential expression, such as the requirement of multiple testing procedures and increasingly, empirical Bayes or similar methods that share information across all observations to improve inference.

Gene expression refers to the appearance of a characteristic or effect in the phenotype that can be attributed to a particular gene. The cost of DNA sequencing continues to decrease at a staggering rate. As it does, sequencing technologies become more and more attractive as platforms for studying gene expression.

High throughput technologies, such as next generation sequencing, microarrays, mass spectrometry proteomics, and metabolomics, are capable of evaluating the expression levels of thousands of genes, proteins, or metabolites in an individual run. Differential expression analysis, the comparison of expression across conditions, has become the primary tool for finding biomarkers, drug targets, and candidates for further research. Typically, gene expression data have been analyzed on a gene-by-gene basis, without regard for complex interactions and association mechanisms.

Melanoma, a highly invasive type of skin cancer that penetrates the entire dermis layer, is associated with increased mortality rates. Excessive exposure of the skin to sunlight, specifically ultraviolet radiation, is the underlying cause of this malignant condition. The appearance of unique skin moles represents a visible clue, referred to as the “ugly duckling” sign, indicating the presence of melanoma and its association with cellular DNA damage.

Melanoma is the deadliest skin tumour and is prone to distant metastases. The incidence of melanoma has increased rapidly in the past few decades, and current trends indicate that this growth is continuing. This study was aimed to explore the molecular mechanisms of melanoma pathogenesis and discover underlying pathways and genes associated with melanoma. We used high-throughput expression data to study differential expression profiles of related genes in melanoma. The microarray dataset for melanoma skin cancer was obtained

from the GEO database, and thorough data analysis and quality control measures were performed to identify differentially expressed genes (DEGs).

## **2. SURVIVAL ANALYSIS USING KM PLOT**

Kaplan-Meier estimate is one of the best options to be used to measure the fraction of subjects living for a certain amount of time after treatment. In clinical trials or community trials, the effect of an intervention is assessed by measuring the number of subjects survived or saved after that intervention over a period of time. The time starting from a defined point to the occurrence of a given event, for example death is called as survival time and the analysis of group data as survival analysis.

Survival time are data that measure the time to a certain event such as death, failure, response, relapse, divorce or the development of a given disease. Therefore survival analysis is a statistical technique for analyzing data on the occurrence of events especially in cohort study.

Survival time can be length of remission, time to disappearance of a tumor, time to death and the time from the start of treatment to the response. Survival time has two important components that must be unambiguously defined: a starting point and an endpoint reached either when the event of interest occurs or when the follow-up time has ended. Survival data may include survival time, response to a given treatment, and patient characteristics related to survival, response and the development of disease. These data can be derived from clinical and epidemiologic studies of humans who have acute or chronic disease. Unlike other statistical methods such as logistic regression, among others, survival analysis considers censoring and time.

The reproducibility and accuracy of survival probabilities, median survival times and hazard ratios based on reconstructed KM data was assessed by comparing published statistics (survival probabilities, medians and hazard ratios) with statistics based on repeated reconstructions by multiple observers.

In the June 1958 issue of the Journal of the American statistical association (JASA) Edward L Kaplan and Paul Meier published a manuscript describing an innovative statistical method to estimate survival rates when there are incomplete survival observations included in the database.

### 3. BIOMARKER SCREENING

Biomarkers provide a powerful and dynamic approach to understanding the spectrum of malignancies with applications in observational and analytic epidemiology, randomized clinical trials, screening, diagnosis and prognosis. A biomarker used to show that a biological response, potentially beneficial or harmful, has occurred in an individual who has been exposed to a medical product or an environmental agent.

High-throughput gene expression data provides new resources and opportunities. For instance, Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases provide a huge amount of data for discovery of biomarkers.

Lung cancer is the most frequently occurring cancer in the world and continually leads in mortality among cancers. The overall 5-year survival rate for lung cancer has risen only 4% (from 12% to 16%) over the past 4 decades, and late diagnosis is a major obstacle in improving lung cancer prognosis. Survival of patients undergoing lung resection is greater than 80%, suggesting that early detection and diagnosis of cancers before they become inoperable and lethal will greatly improve mortality. Lung cancer biomarkers can be used for screening, detection, diagnosis, prognosis, prediction, stratification, therapy response monitoring, and so on.

Biomarkers in the medical field can be classified into various categories based on their purpose, as follows:

- (1) **Diagnostic purpose:** screening, early detection, and diagnostic biomarkers
- (2) **Prediction/staging purpose:** predictive, risk stratification, staging biomarkers
- (3) **Prognostic purpose:** recurrence, metastatic, prognostic biomarkers
- (4) **Treatment purpose:** companion, treatment selection biomarkers, therapy response prediction

Among these biomarkers, we will focus in this review on the diagnostic and prognostic purpose biomarkers, primarily the screening, early detection, and diagnostic biomarkers, as well as some predictive biomarkers

.

## **4. DRUG TARGET SCREENING & BINDING AFFINITY ASSESSMENT**

Protein-ligand docking is a structure-based computational method, which is used to predict the small molecule binding modes and binding affinities with protein receptors. Two docking software, Discovery Studio and AutoDock, are used for docking comparison of 195 protein-ligand complexes from PDBind dataset. Discovery Studio has a higher accuracy (successfully docked complexes, within 5 RMSD of the native complex structures) than AutoDock.

Molecular docking is a computer simulation procedure to predict the conformation of a receptor- ligand complex. Molecular docking methodology explores the behavior of small molecules in the binding site of a target protein. As more protein structures are determined experimentally using X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, molecular docking is increasingly used as a tool in drug discovery.

### **Receptor-Ligand Interactions:**

DS LigandFit - Gain direct insight into the complementary features of ligands and their potential as lead candidates. DS LigandFit lets you easily dock ligands into the binding site of receptors using shape-based searching and Monte Carlo sampling of ligands. Parameters are customizable, and your settings can be saved and shared with other users. DS LigandFit can be parallelized in multi-core machines or compute clusters for virtual high-throughput screening.

Molecular docking of small molecules to a biological target includes imaginative sampling of possible poses of the ligand in the specified groove or pocket of target candidate in an order to establish the optimal binding geometry.

Information gained from the preferred orientation of bound molecules may be employed to predict the energy profiling (such as binding free energy), strength and stability (like binding affinity and binding constant) of complexes. This can be done using scoring function of molecular docking. Now days, molecular docking is often utilized to forecast the binding orientation of small molecules (drug candidates) to their biomolecular target (such as protein, carbohydrate and nucleic acid) with the aim to determine their tentative binding parameters. This establishes raw data for the rational drug designing (structure-based-drug development) of new agents with better efficacy and more specificity.

## **AIM AND OBJECTIVE**

To comprehensively investigate the molecular landscape of lung cancer, elucidating dysregulated genes, molecular pathways, and potential therapeutic targets, while also assessing the prognostic significance of treatment modalities, including the drug doxorubicin.

### **OBJECTIVE**

1. Conduct a differential expression analysis using GEO2R to identify genes dysregulated in lung cancer compared to normal lung tissue.
2. Perform functional analysis utilizing Cytoscape and ClueGO to elucidate the biological processes and pathways associated with the dysregulated genes.
3. Utilize the KM plotter to analyze survival plots
4. Explore potential biomarkers and therapeutic targets for lung cancer through integrated analysis of gene expression data and survival outcomes
5. Assess the binding affinity of doxorubicin with potential therapeutic targets identified in lung cancer using molecular docking methods, providing insights into its mechanism of action and therapeutic potential

## **MATERIALS AND METHODS**

- 1. GEO**
- 2. CYTOSCAPE**
- 3. CYTOHUBBA**
- 4. CLUEGO**
- 5. KM PLOTTER**
- 6. THERAPEUTIC BIOMARKER (ThMAR)**
- 7. DRUG BANK**
- 8. PUBCHEM**
- 9. SWISS TARGET PREDICTION**
- 10.DISCOVERY STUDIO**

### **1. GEO**

The gene expression omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) is an international public repository for high-throughput microarray and next-generation sequence functional genomic data sets submitted by the research community. The resource supports archiving of raw data, processed data and metadata which are indexed, cross-linked and searchable. All data are freely available for download in a variety of formats GEO also provides several web-based tools and strategies to assist user to query, analyse and visualize data.

### **2. CYTOSCAPE**

Cytoscape is a free software package for visualizing, modeling and analyzing molecular and genetic interaction networks. Five major steps are described: (1) obtaining a gene or protein network, (2) displaying the network using layout algorithms, (3) integrating with gene expression and other functional attributes, (4) identifying putative complexes and functional modules and (5) identifying enriched gene ontology annotations in the network.

Cytoscape is an open source software project for integrating biomolecular interaction networks with high-throughput expression data and other molecular states into a unified conceptual framework. Although applicable to any system of molecular components and interactions, cytoscape is most powerful when used in conjunction with large databases of protein-protein, protein-DNA, and genetic interactions that are increasingly available for humans and model organisms.

Cytoscape software core provides basic functionality to layout and query the network, to visually integrate the network with expression profiles, phenotypes and other molecular states and to link the network to databases of functional annotations.

### **3. CYTOHUBBA**

Cytoscape plugin cytohubba for ranking nodes in a network by their network features. Cytohubba provides 11 topological analysis methods including degree, edge percolated components Maximum Neighborhood component density of maximum neighborhood components maximal clique centrality and six centralities (bottleneck, EcCentricity, closeness, betweenness, and stress) based on shortest paths. Among the eleven methods, the new proposed method, MCC has a better performance on the precision of predicting essential protein from the yeast PPI network.

### **4. CLUEGO**

ClueGO has two major features: it can be either used for the visualization of terms corresponding to a list of genes, or the comparison of functional annotations of two clusters. ClueGO offers the possibility to calculate enrichment/depletion tests for terms and groups as left-sided (Enrichment), right-sided (Depletion) or two-sided (Enrichment/ Depletion) tests based on the hypergeometric distribution. Furthermore it provides options to calculate mid-P-values and doubling for two-sided tests to deal with discreteness and conservatism effect as suggested. To correct the P-values for multiple testing several standard correction methods are proposed.

When comparing two clusters, another original feature of ClueGO allows to switch the visualization of the group on the network to the cluster distribution over the terms. Besides the network, ClueGO provides overview charts showing the groups and their leading term as well as detailed term histograms for both, cluster specific and common terms.

Like BinGO, ClueGO can be used in conjunction with GOrize for functional analysis of a cytoscape gene network. The created network, charts and analysis result can be saved as project in a specified folder and used for further analysis.

## 5. KM PLOTTER

The Kaplan Meier plotter is capable of assessing the correlation between the expression of all genes (mRNA, miRNA, protein, & DNA) and survival in 35k+ samples from 21 tumor types. Applied statistical tools include Cox proportional hazards regression and the computation of the False Discovery Rate. With 18,000 analyses per day, the KM-plotter is a worldwide reference for the discovery and validation of survival biomarkers.

## 6. THE THERAPEUTIC BIOMARKERS (ThMARs)

Therapeutic biomarker (ThMAR), distinct from traditional diagnostic and prognostic biomarker, has emerged to be critical in drug development and clinical practice of therapies. In this study, a database titled ‘TheMarker’ was constructed to provide the comprehensive data on ThMARs. It is unique in systematically providing five types of ThMAR to realize the collective consideration among various stages of drug discovery, comprehensively describing ThMAR data for the largest number of drugs among existing databases, and extensively covering the widest range of diseases. The data provided in TheMarker are highly expected to have great implications and significant impacts on both drug development and clinical practice.

### DATA AVAILABILITY:

TheMarker is freely accessible without any login requirement at: <https://idrblab.org/themarker>.

## 7. DRUG BANK

DrugBank ([www.drugbank.ca](http://www.drugbank.ca)) is a web-enabled database containing comprehensive molecular information about drugs, their mechanisms, their interactions and their targets. The rich, high quality, primary-sourced content found in DrugBank has allowed it become one of the world's most widely used reference drug resources.

## 8. PUBCHEM

PubChem (<https://pubchem.ncbi.nlm.nih.gov>) (1–6) is a public repository for information on chemical substances and their biological activities. Since launched in 2004 as a component of the Molecular Libraries Roadmap Initiatives of the US National Institutes of Health (NIH), PubChem has rapidly grown to a key chemical information resource that serves scientific

communities in many areas such as cheminformatics, chemical biology, medicinal chemistry and drug discovery.

## **9. SWISS TARGET PREDICTION**

SwissTargetPrediction is a web-based tool, on-line since 2014, to perform ligand-based target prediction for any bioactive small molecule.

## **10.DISCOVERY STUDIO**

Discovery Studio is a suite of software for simulating small molecules and macromolecules systems.

## **METHODS**

Expression profiling (GSE83583) data were obtained from GEO DataSets. Differentially expressed genes were identified using GEO2R. Gene ontology and pathway enrichment analyses were performed for the MDEGs. Protein–protein interaction (PPI) networks were established by Cytoscape. Further, the hub genes derived from the PPI networks were verified by The Cancer Genome Atlas (TCGA) database.

As a statistical analysis method that combines the outcome of an event with the time that it takes to occur, survival analysis can be used to explore whether a gene can significantly affect a patients survival time. The Skin cancer survival analysis of the hub genes is mainly carried out with the Kaplan Meier plotter (<http://kmplot.com/>) online tool, which can be used to explore whether the high expression of a given gene can significantly affect the survival time of patients. In addition, Kaplan Meier plotter also serve as an excellent tool for visualization of survival analysis results.

Distinct from the traditional diagnostic/prognostic biomarker (adopter as the indicator of disease state/process), the therapeutic biomarker (ThMAR) has emerged to be very crucial in the clinical development and clinical practice of all therapies. There are five types of ThMAR that have been found to play indispensable roles in various stages of drug discovery, such as: pharmacodynamics biomarker essential for guaranteeing the pharmacological effects of a therapy, safety biomarker critical for assessing the extent or likelihood of therapy-induced toxicity, monitoring biomarker indispensable for guiding clinical management by serially measuring patients status, predictive biomarker crucial for maximizing the clinical outcome

of a therapy for specific individuals, and surrogate endpoint fundamental for accelerating the approval of a therapy.

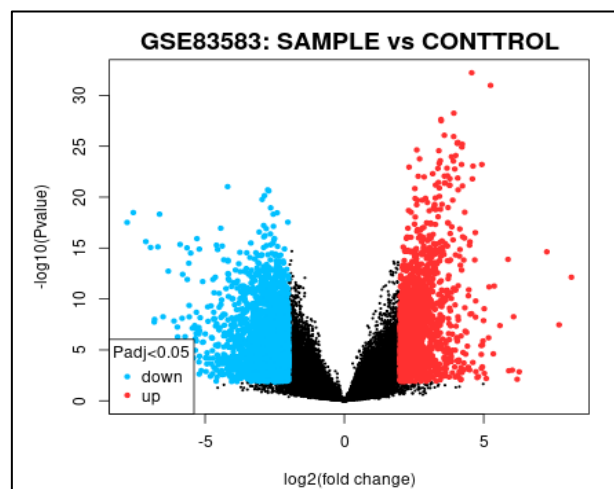
Molecular docking analysis is a structural bioinformatics approach commonly used for drug discovery, identifying potential inhibitors for a given biological target. Molecular docking also unravels interaction modes between macromolecules and drug candidates. Molecular docking analysis includes three steps: macromolecule structure, preparation of small molecule structure, and evaluation of binding affinity between ligand and receptor. Postdocking studies such as molecular dynamics (MD) analysis are frequently used to understand better ligands' dynamic manner and stability within the receptors' binding site (e.g., proteins' active site). Moreover, MD enables assessing the flexibility of residues inside the catalytic site of proteins

## RESULT AND DISCUSSION:

### EXPERIMENTAL 1: DIFFERENTIAL GENE EXPRESSION

- **GEO DATASET:** GSE83583
- **GENE LIST:** PRAME, MGP, PRAME, SPP1, MGP, CXCL8, SPP1, SERPINA3, TRIM9, PTPRZ1.
- **FILTER:** Taxonomy ID (9606)
- **TOTAL GENES:** 47306
- **UPREGULATED GENE:** 1866
- **DOWNREGULATED GENE:** 45,440
- **SAMPLE:** 84

a)



**FIG 1: volcano plot**

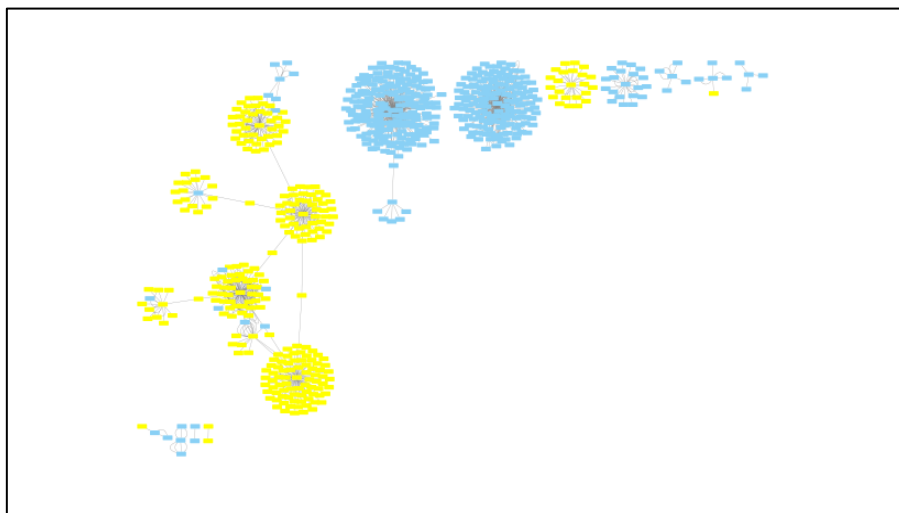
The **volcano plot**, which is commonly used in genomic studies to visualize differential expression between two different conditions. Here's a brief interpretation of the plot:

- **Dataset:** The plot is based on the dataset **GSE83583** and compares gene expression between a **sample** and a **control** group.

- **X-Axis:** Represents the  **$\log_2(\text{fold change})$**  in gene expression, with negative values indicating downregulation and positive values indicating upregulation in the sample compared to the control.
- **Y-Axis:** Shows the  **$-\log_{10}(\text{P-value})$** , which indicates the statistical significance of the gene expression changes. Higher values mean more significant changes.
- **Color Coding:**
  - **Blue dots** signify genes that are significantly **downregulated** in the sample.
  - **Red dots** signify genes that are significantly **upregulated** in the sample.
  - **Black dots** represent genes with no significant change in expression.
- **Significance Threshold:** The annotation “ $\text{Padj} < 0.05$ ” means that the results have been filtered to only include genes with an adjusted p-value less than 0.05, which is a common threshold for statistical significance in such analyses.

This type of plot helps researchers quickly identify genes that are significantly differentially expressed and may warrant further investigation.

**b)**



**Fig 2: Network build with top ten gene using cytoscape and created column filter with taxonomy ID (9606).**

c)

|    | A           | B          | C           | D                         | E            | F             | G           | H             | I             | J            | K |
|----|-------------|------------|-------------|---------------------------|--------------|---------------|-------------|---------------|---------------|--------------|---|
| 1  | DMNC        | MCC        | MNC         | DEGREE                    | EPC          | BOTTLENECK    | CLOSENESS   | BETWENESS     | STRESS        | EccENTRALITY |   |
| 2  | CXCL8       | SAF41      | SPCC594.05c | SPCC594.05c               | SPCC594.05c  | OIP-4         | SPCC594.05c | OIP-4         | OIP-4         | SAF41        |   |
| 3  | OIP-4       | BSPI       | OIP-4       | SAF41                     | SPAC23H3.05c | BSPI          | SAF41       | SPCC594.05c   | RP11-297A16.3 | SPCC895.07   |   |
| 4  | TRIM9       | hCGBP      | TRIM9       | hCGBP                     | ash2         | S13           | BSPI        | BSPI          | TRIM9         | YDL160C      |   |
| 5  | SPCC594.05c | SERPINA3   | CXCL8       | BSPI                      | SPBC337.03   | SPCC594.05c   | OIP-4       | hCGBP         | hCGBP         | YLR418C      |   |
| 6  | SPAC1610.01 | Trim9      | SPAC1610.01 | OIP-4                     | SPBC1734.15  | ZNF408        | hCGBP       | TRIM9         | ZNF408        | LEO1         |   |
| 7  | YDL160C     | Mgp        | YDL160C     | TRIM9                     | arp42        | hCGBP         | SPCC895.07  | SAF41         | BSPI          | CSL3         |   |
| 8  | YLR418C     | phosphacan | YLR418C     | SERPINA3                  | SPBC1604.20c | SAF41         | SPAC1610.01 | RP11-297A16.3 | CXCL8         | KAT5         |   |
| 9  | MAT1        | spp1       | MAT1        | Trim9                     | ask1         | MOX2          | SPBC25H2.15 | S13           | S13           | YJL183W      |   |
| 10 | BBC1        | CXCL8      | BBC1        | histone methyltransferase | dad2         | RP11-297A16.3 | SPAC227.01c | ZNF408        | p300          | L000001601   |   |
| 11 | SPBC25H2.15 | Trim9      | SPBC25H2.15 | CXCL8                     | SPAC3H1.12c  | TRIM9         | ask1        | MOX2          | Trim9         | L000003094   |   |

FIG 3: Hub genes are ranked using cytohubba

d)

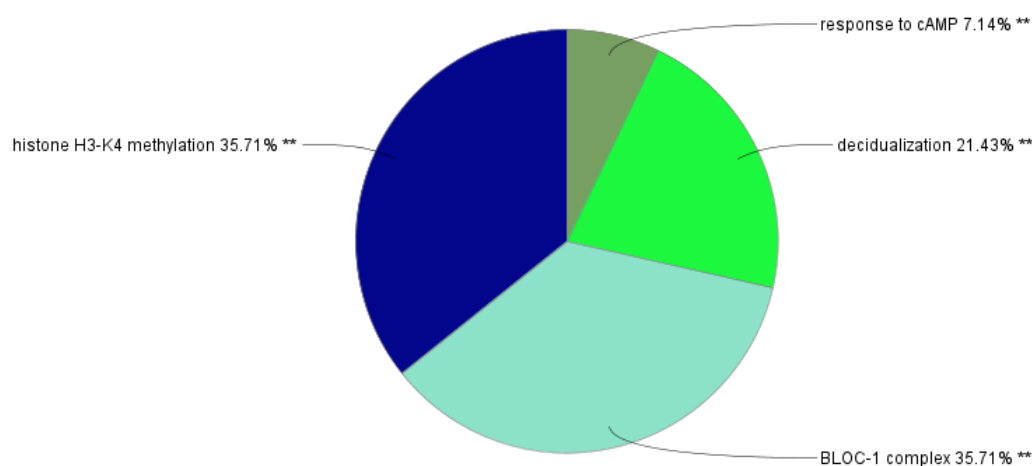


FIG 4: Top 10 ranked hub gene were plot as pie chart

The pie chart from a bioinformatics analysis tool, displaying the distribution of certain biological processes or components. Here's a brief analysis:

- **Histone H3-K4 Methylation:** Represents 35.71% of the chart, suggesting a significant presence of this epigenetic modification in the sample.
- **Response to cAMP:** Occupies 7.14%, indicating a smaller proportion of processes related to cellular response to cyclic AMP.
- **Decidualization:** Accounts for 21.43%, which could be relevant in studies of endometrial transformation or pregnancy.
- **BLOC-1 Complex:** Also makes up 35.71%, pointing to its potential importance in the biological sample being analysed

### e) ENRICHMENT ANALYSIS

| S.NO | ID         | TERM NAME                 | ONTOLOGY                                                        | FUCTIONS                                                                                                                                                                                                                                              |
|------|------------|---------------------------|-----------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1.   | GO:0051591 | Response to cAMP          | GO_BiologicalProcess-EBI-UniProt-GOA-ACAP-ARAP_25.05.2022_00h00 | cAMP, or cyclic adenosine monophosphate, plays several crucial roles within cells: cell signalling, metabolism regulation, ion channel regulation, gene expression.                                                                                   |
| 2.   | GO:0046697 | Decidualization           | GO_BiologicalProcess-EBI-UniProt-GOA-ACAP-ARAP_25.05.2022_00h00 | Decidualization is a complex process that occurs in the endometrium (the lining of the uterus) during the menstrual cycle and early pregnancy. Its primary function is to prepare the endometrium for embryo implantation and support early pregnancy |
| 3.   | GO:0031083 | BLOC-1 Complex            | GO_CellularComponent-EBI-UniProt-GOA-ACAP-ARAP_25.05.2022_00h00 | The Biogenesis of Lysosome-related Organelles Complex 1 (BLOC-1) is a protein complex with diverse functions, primarily associated with intracellular trafficking and organelle biogenesis                                                            |
| 4.   | GO:0051568 | Histone H3-K4 Methylation | GO_BiologicalProcess-EBI-UniProt-GOA-ACAP-ARAP_25.05.2022_00h00 | Histone H3-K4 methylation is a post-translational modification of histone proteins that plays a crucial role in the regulation of chromatin structure and gene expression.                                                                            |

## EXPERIMENT 2:

### SURVIVAL ANALYSIS

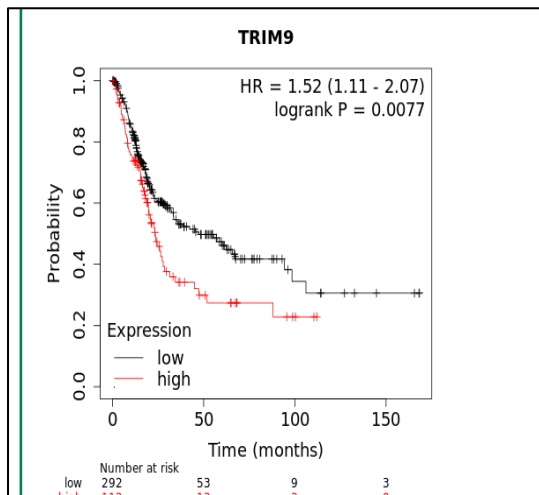


FIG 5: KM plot

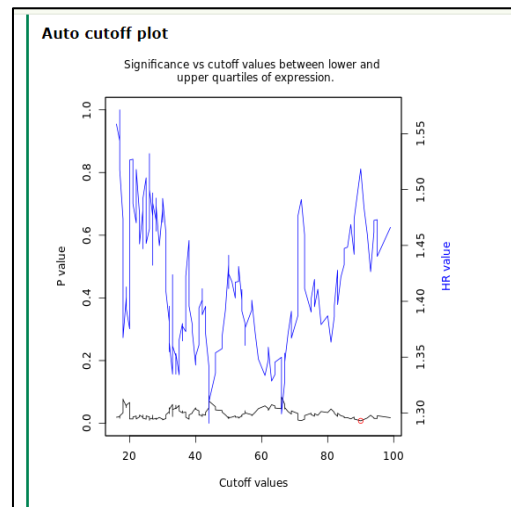


FIG 6: Auto cutoff plot

| Median survival                |                                 |
|--------------------------------|---------------------------------|
| Low expression cohort (months) | High expression cohort (months) |
| 47.33                          | 23.53                           |

FIG 7: Median survival

The Kaplan-Meier survival curve graph, which is used in medical research to estimate the survival probability over time. Here's a brief analysis of the image:

- **TRIM9 Expression:** The graph compares the survival probability between two groups based on the expression levels of TRIM9, a gene that may be involved in cancer progression.
- **Survival Curves:** The black curve represents low expression of TRIM9, while the red curve represents high expression. The separation between the curves suggests a difference in survival outcomes.
- **Statistical Significance:** The hazard ratio (HR) of 1.52 indicates that high expression of TRIM9 is associated with a 52% higher hazard of the event occurring compared to low expression. The logrank P-value of 0.0077 shows that this difference is statistically significant.

- **Time and Risk:** The x-axis shows time in months, and the table below indicates the number of individuals at risk over time for the low expression group.

The median survival times for two cohorts based on their expression levels. Here's a brief interpretation:

- **Low Expression Cohort:** The median survival time is **47.33 months**, which suggests that patients with low expression levels of a certain biomarker or gene have a longer median survival time.
- **High Expression Cohort:** The median survival time is significantly shorter, at **23.53 months**, indicating that high expression levels are associated with a worse survival outcome.

### EXPERIMENT 3: BIOMARKER SCREENING

#### a) GENERAL INFORMATION OF THE BIOMARKER

| BIOMARKER ID | BIOMARKER NAME         | CLASS   | GENE NAME  | BIOMARKER TYPE | REPRESENTATIVE DRUG   |
|--------------|------------------------|---------|------------|----------------|-----------------------|
| B3HTD6       | C-Reactive protein     | protein | CRP, PTX1  | PSMR           | Radiation therapy     |
| BH7DL8       | Interleukin 6          | Protein | IL6, IFNB2 | PSMR           | Anti-PD1/PDL1 Therapy |
| BITVG9       | C-C motif chemokines 2 | Protein | CCL2, MCP1 | PSM            | Cisplatin             |
| B6CUE3       | Interleukin 10         | protein | IL10       | PSM            | Anti-PD1/PDL1 Therapy |

#### b) SAFETY BIOMARKER (SAF)

##### 1. BENZALKONIUM

**DRUG TYPE:** Small molecule

**INDICATION:** Bacterial infection [ ICD – 11 : IA00 – IC4Z ]

**MOLECULAR FORMULA:** C<sub>22</sub>H<sub>40</sub>N<sup>+</sup>

## **2. IPILIMUMAB**

**DRUG TYPE:** Antibody

**INDICATION:** Colorectal Cancer [ICD-11: 2B91]

Hepatocellular Carcinoma [ICD-11: 2C12]

Non-small cell lung cancer [ICD-11: 2C25]

## **3. NIVOLUMAB**

**DRUG TYPE:** Antibody

**INDICATION:** Melanoma [ICD-11: 2C30]

Hodgkin lymphoma [ICD-11: 2B30]

Urothelial carcinoma [ICD-11: 2C92]

Colorectal cancer [ICD-11: 2B91]

Small cell lung cancer [ICD-11: 2C25]

Non-small cell lung cancer [ICD-11: 2C25]

Hepatocellular carcinoma [ICD-11: 2C12]

## **4. PERBROLIZUMAB**

**DRUG TYPE:** Antibody

**INDICATION:** Melanoma [ICD-11: 2C30]

Hodgkin lymphoma [ICD-11: 2B30]

Urothelial carcinoma [ICD-11: 2C92]

Triple-negative breast cancer [ICD-11: 2C60]

Cervical cancer [ICD-11: 2C77]

Gastric or gastroesophageal junction adenocarcinoma [ICD-11: 2B71]

Merkel cell carcinoma [ICD-11: 2C34]

Microsatellite instability-high or mismatch repair deficient solid tumor [ICD-11: 2A00-2F9Z]

Tumor mutational burden-high solid tumor [ICD-11: 2A00-2F9Z]

Non-small cell lung cancer [ICD-11: 2C25]

Small cell lung cancer [ICD-11: 2C25]

Hepatocellular carcinoma [ICD-11: 2C12]

Primary mediastinal large B-cell lymphoma [ICD-11: 2A81]

Endometrial carcinoma [ICD-11: 2C76]

Head and neck squamous cell carcinoma [ICD-11: 2D60]

## 5. TRASTUZUMAB

**DRUG TYPE:** Antibody

**INDICATION:** Breast cancer [ICD-11: 2C60]

Gastric and gastroesophageal cancer [ICD-11: 2B71]

### c) APPLICATION OF BIOMARKER IN DIAGNOSIS AND PROGNOSIS:

| DIAGNOSIS                                                                                                                                                                                                                                                                                                                                                                                                | PROGNOSIS                                                                                                                                                                                                                                                                                                                                                                     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"><li>➤ Acute respiratory illness</li><li>➤ Alzheimer's disease</li><li>➤ Breast cancer</li><li>➤ Chronic heart failure</li><li>➤ Colorectal cancer</li><li>➤ Congestive heart failure</li><li>➤ Cutaneous melanoma</li><li>➤ Heart failure</li><li>➤ Heart failure with preserved ejection fraction</li><li>➤ Huntington's disease</li><li>➤ Maternal obesity</li></ul> | <ul style="list-style-type: none"><li>➤ Acute decompensated heart failure</li><li>➤ Acute heart failure</li><li>➤ Angioimmune blastin T- CELL lymphoma</li><li>➤ Chronic heart failure</li><li>➤ Colon cancer</li><li>➤ Colorectal cancer</li><li>➤ Congestive heart failure</li><li>➤ Decompensated heart failure</li><li>➤ Heart failure</li><li>➤ Ovarian cancer</li></ul> |

#### d) TISSUE-SPECIFIC EXPRESSION OF THE BIOMARKER

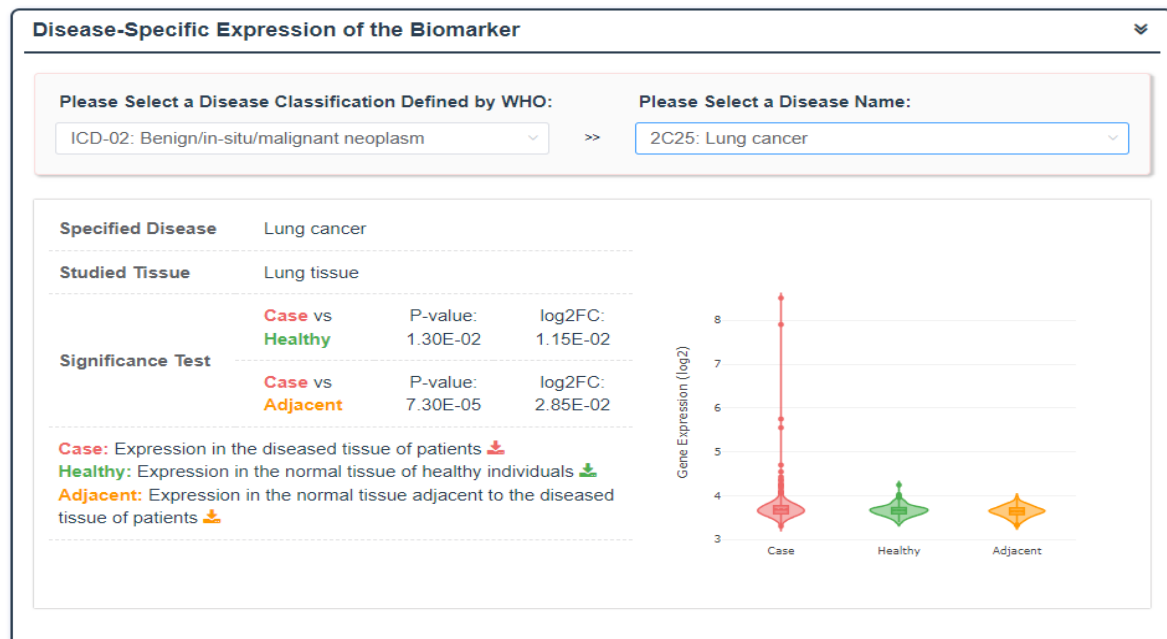


FIG 8: Disease-specific expression of the biomarker

#### EXPERIMENT 4:

**DISEASE:** Lung cancer

**DRUG NAME:** Sorafenib

**PROTEIN:** 2Y9Q

**TOP 10 GENES:** NTRK1, P33, PIG61, RP5-823N20.1, Dr8, DER/ Faint little ball, PPPIR47, PRKM1, LYN, D –CbIL.

a)

| NAME   | -PLP1 | -PLP2 | JAIN | -PMF   | DOCK SCORE |
|--------|-------|-------|------|--------|------------|
| 216239 | 98.39 | 94.24 | 0.22 | 123.15 | 71.659     |

b)

| TARGET                                  | COMMON NAME | UNIPROT ID | TARGET CLASS                        |
|-----------------------------------------|-------------|------------|-------------------------------------|
| Serine threonine-protein kinase RAF     | RAF1        | 904049     | Kinase                              |
| Homodomain-interacting protein kinase 4 | HIPK4       | Q8NE63     | Kinase                              |
| Serine/threonine-protein kinase TAO2    | TAOK2       | Q9UL54     | Kinase                              |
| Cyclin- dependent kinase like 3         | CDKL3       | Q81VW4     | Kinase                              |
| Serotonin 2b (5- HT2b) receptor         | HTR2B       | P41595     | Family AG Protein- coupled receptor |
| Tyrosine- protein kinase PYN            | FYN         | PO6241     | Kinase                              |

c)

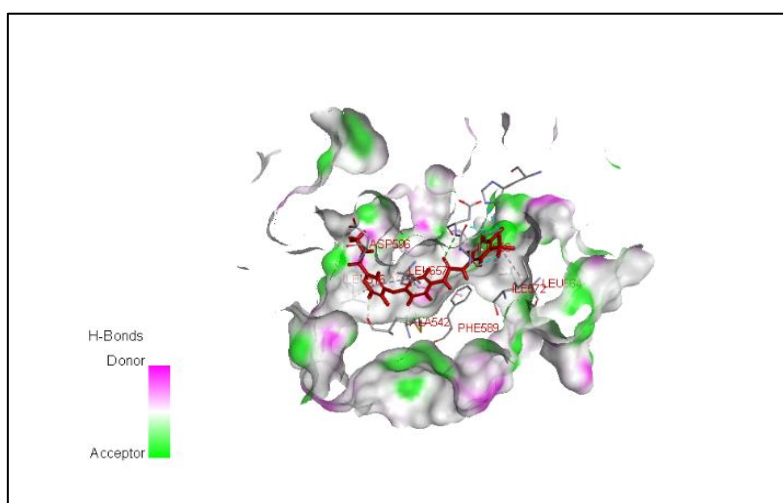
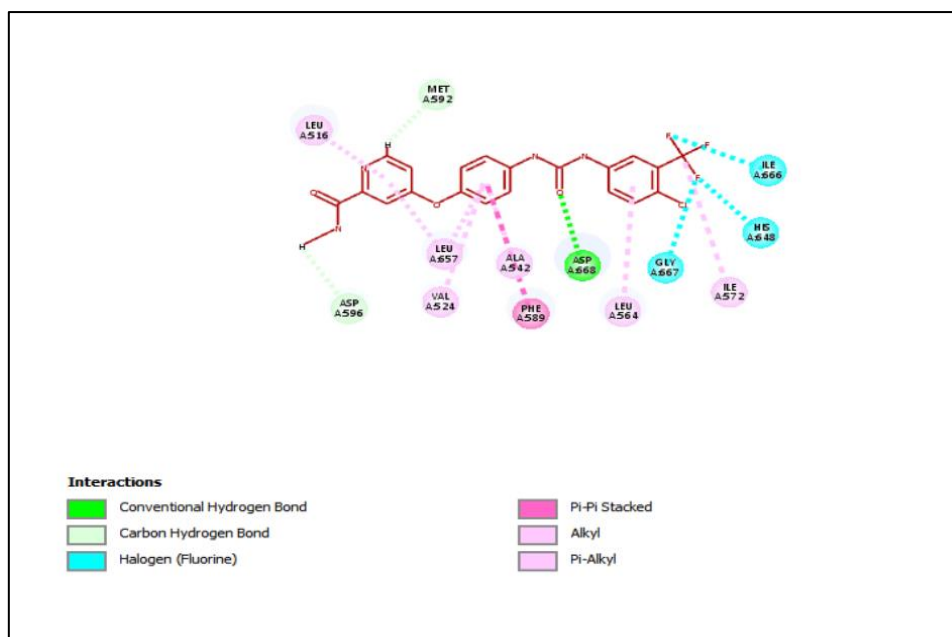


FIG 9: 3D Structure



**FIG 10: 2D Structure**

The molecular interaction diagram, which is commonly used in biochemistry and molecular biology to illustrate the interactions between amino acids and other molecules like ligands or proteins. Here's an interpretation based on the image details:

- **Amino Acid Interactions:** The diagram shows various amino acids such as LEU A:516, MET A:592, ILE A:566, etc., and how they interact with a central molecule.
- **Bond Types:** Different types of bonds are color-coded to indicate their nature:
  - **Green lines** represent conventional hydrogen bonds, which are critical for the stability of molecular structures.
  - **Pink lines** indicate Pi-Pi Stacked interactions, which occur between aromatic rings and contribute to the three-dimensional structure of proteins.
  - **Red lines** show carbon hydrogen bonds, which are weaker than conventional hydrogen bonds but still play a role in molecular shape.
  - **Blue dotted lines** represent halogen (fluorine) interactions, which can influence the binding affinity of molecules.
- **Legend:** The legend at the bottom left of the image provides a key to understanding the color coding of the interactions.

## **d)INTERACTION BETWEEN RESIDUE AND DRUG**

### **1. GLY13**

The interaction between Glycine 13 (Gly13) and PRKM1 refers to a specific molecular interaction involving the protein kinase M1 (PRKM1), also known as PKC $\mu$ . Glycine 13 is a residue within the protein sequence of PRKM1.

PRKM1 is a member of the protein kinase C (PKC) family, which plays crucial roles in various cellular processes, including cell proliferation, differentiation, apoptosis, and cytoskeletal organization. PKC $\mu$  is a unique member of the PKC family, characterized by its unique domain structure and substrate specificity.

The interaction between Gly13 and PRKM1 likely refers to a structural or functional aspect of PRKM1 that involves this particular glycine residue. Glycine residues are often found at key positions within protein structures due to their small size and flexible nature, allowing them to contribute to protein conformation and function.

### **2. LYS129**

The interaction between Lysine 129 (Lys129) and PRKM1, also known as PKC $\mu$ , would involve a specific molecular interaction between the amino acid residue Lys129 and the protein kinase M1.

PRKM1, a member of the protein kinase C (PKC) family, is involved in various cellular processes such as cell proliferation, differentiation, apoptosis, and cytoskeletal organization. Lys129, as a specific amino acid residue within the PRKM1 protein sequence, could play a crucial role in its structure, function, or regulation.

The interaction between Lys129 and PRKM1 might serve several purposes:

- Structural stability
- Substrate binding
- Catalytic activity
- Regulatory mechanisms

### 3. H47

The interaction between Histidine 47 (H47) and PRKM1, also known as PKC $\mu$ , refers to a specific molecular interaction involving the histidine residue at position 47 within the protein sequence of PKC $\mu$ .

PRKM1 is a member of the protein kinase C (PKC) family, involved in various cellular processes such as cell proliferation, differentiation, apoptosis, and cytoskeletal organization. Histidine 47, being a specific amino acid residue within the PRKM1 protein sequence, could play a significant role in its structure, function, or regulation.

The interaction between H47 and PRKM1 could serve several purposes:

- Metal Ion co-ordination
- Substrate recognition
- Ph sensing
- Conformational stability

## **CONCLUSION:**

### **Identification of Hub Genes Based on the lung Cancer PPI Network**

As a Java plugin for Cytoscape, cytoHubba facilitate the analysis and visualization of molecular interaction networks based on 11 topological algorithms and thus enable us to find essential nodes from a network. Since MCC algorithm has a good performance in predicting PPI network hub genes, we therefore apply MCC algorithm to identify the melanoma hub genes in current study. A total of 10 hub genes were finally identified based on the MCC algorithm and ranked according to the degree value. The top10 hub genes are TRIM9, OIP4, CXCL8, SPCC594.05c, YDL160C, SAF41, BSPI, S13, SPBC25H2.15, HCGBP.

## **SURVIVAL ANALYSIS:**

The Kaplan Meier plotter can be used to assess the effect of a lot of genes on patient survival among 21 cancer types, such as ovarian (n=2,190), gastric (n=1,440), and lung (n=3,452) cancer. High expression hub genes of TRIM9 have been demonstrated to be correlated with significant increase of survival melanoma patients.

## **BIOMARKER SCREENING:**

Biomarker screening is a fundamental aspect of modern healthcare and biomedical research, allowing for the detection, monitoring, and treatment of diseases. Biomarkers, which can be various molecules or molecular signatures found in biological samples, serve as indicators of physiological or pathological processes. Key aspects of biomarker screening include the identification of suitable biomarker candidates, the development of sensitive detection methods, and the validation of biomarker performance in diverse patient populations.

Advancements in high-throughput technologies, such as genomics, proteomics, metabolomics, and imaging, have significantly accelerated biomarker discovery and screening processes. These technologies enable the simultaneous analysis of numerous biomarkers, providing valuable insights into disease mechanisms and treatment responses.

Moreover, machine learning algorithms and bioinformatics tools play crucial roles in analyzing complex biomarker data and identifying patterns associated with disease states or treatment outcomes. This interdisciplinary approach enhances our ability to utilize biomarkers effectively in clinical decision-making and personalized medicine.

Looking ahead, continued innovation in biomarker screening methodologies, coupled with collaborative efforts among researchers, clinicians, regulatory agencies, and industry partners, will drive progress in disease diagnosis, patient care, and therapeutic development. Ultimately, biomarker screening holds the promise of revolutionizing healthcare delivery, improving patient outcomes, and advancing precision medicine initiatives.

## **MOLECULAR DOCKING AND DRUG DISCOVERY**

Molecular docking stands as a promising tool in the realm of lung cancer drug discovery. It enables researchers to computationally predict and analyze the interactions between potential drug candidates and target proteins relevant to lung cancer. Through this process, molecular docking facilitates the identification of lead compounds, optimization of their binding affinity, and design of novel therapeutics tailored to combat this deadly disease.

Key advantages of molecular docking in lung cancer drug discovery include its efficiency in screening large chemical libraries, prioritizing lead compounds, and providing valuable insights into the molecular mechanisms underlying lung cancer progression and therapy resistance. Advancements in computational methodologies further enhance the accuracy and reliability of molecular docking predictions, while integration with other techniques enables a comprehensive understanding of lung cancer biology.

Looking ahead, molecular docking holds the potential to accelerate the discovery and development of personalized treatments for lung cancer, addressing unmet medical needs and improving patient outcomes. By leveraging computational approaches, molecular docking contributes to the advancement of precision medicine, striving towards a future where lung cancer patients receive tailored therapies that offer improved survival rates and quality of life.

## REFERENCES:

- Lung Cancer Biomarkers . Volume 72, 2015, Pages 107-170. Je-Yoel Cho Hoseok I
- Protein-ligand Docking Application and Comparison using Discovery Studio and AutoDock .Wang, Qi.
- Molecular docking algorithms. Volume 9, issue 12, 2008
- An overview of molecular docking. 23 May 2016. Shweta Agarwal and Ranjana Mehrotra
- Biomarker Screening and Prognostic Significance Analysis for Renal Cell Carcinoma. 07 Sep 2021. Xiangui Meng, Hongwei Yuan, Weiquan Li, Wen Xiao & Xiaoping Zhang
- Online Survival Analysis Software to Assess the Prognostic Value of Biomarkers Using Transcriptomic Data in Non-Small-Cell Lung Cancer. December 18, 2013 . Bala'zs Gyo'rfi, Pawel Surowiak<sup>2</sup> , Jan Budczies , Andra's La'nczky.
- TheMarker: a comprehensive database of therapeutic biomarkers. 2024 Jan 5;52(D1):D1450-D1464. Zhang Y, Zhou Y, Zhou Y, Yu X, Shen X, Hong Y, Zhang Y, Wang S, Mou M, Zhang J, Tao L, Gao J, Qiu Y, Chen Y, Zhu F.
- PubChem Substance and Compound databases . Volume 44, Issue D1, 4 January 2016, Pages D1202–D1213 Sunghwan Kim, Paul A. Thiessen, Evan E. Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A. Shoemaker, Jiyao Wang, Bo Yu, Jian Zhang, Stephen H. Bryant.
- SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules, *Nucleic Acids Research*, Volume 47, Issue W1, 02 July 2019, Pages W357–W364, Antoine Daina, Olivier Michielin, Vincent Zoete,.
- Identification of Potentially Therapeutic Target Genes in Ovarian Cancer via Bioinformatic Approach," 2021 *IEEE 9th International Conference on Bioinformatics and Computational Biology (ICBCB)*, Taiyuan, China, 2021, pp. 96-101, L. Chengzhang and X. Jiucheng.
- Identification of Potentially Therapeutic Target Genes in Ovarian Cancer via Bioinformatic Approach. May 2021 Li Chengzhang's Xu Jiucheng
- edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, Volume 26, Issue 1, January 2010, Pages 139–140, Mark D. Robinson, Davis J. McCarthy, Gordon K. Smyth.
- MultiRankSeq: Multiperspective Approach for RNAseq Differential Expression Analysis and Quality Control . 27 May 2014. Yan Guo, Shilin Zhao, Fei Ye, Quanhu Sheng, and Yu Shyr.
- The 2013 *Nucleic Acids Research* Database Issue and the online Molecular Biology Database Collection, *Nucleic Acids Research*, Volume 41, Issue D1, 1 January 2013, Pages D1–D7, Xosé M. Fernández-Suárez, Michael Y. Galperin.

Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation, *Nucleic Acids Research*, Volume 40, Issue 10, 1 May 2012, Pages 4288–4297, Davis J. McCarthy, Yunshun Chen, Gordon K. Smyth.

Differential Expression Analysis for Pathways Published: March 14, 2013 Winston A. Haynes , Roger Higdon, Larissa Stanberry, Dwayne Collins, Eugene Kolker

Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2, 2366–2382 (2007). Cline, M., Smoot, M., Cerami, E. *et al.*

Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks Paul Shannon<sup>1</sup>, Andrew Markiel<sup>1</sup>, Owen Ozier<sup>2</sup>, Nitin S. Baliga<sup>1</sup>, Jonathan T. Wang<sup>2</sup>, Daniel Ramage<sup>2</sup>, Nada Amin<sup>2</sup>, Benno Schwikowski Trey Ideker.

NCBI GEO: archive for functional genomics data sets—update, *Nucleic Acids Research*, Volume 41, Issue D1, 1 January 2013, Pages D991–D995, Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L. Robertson, Nadezhda Serova, Sean Davis, Alexandra Soboleva.

*cytoHubba*: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 8 (Suppl 4), S11 (2014). Chin, CH., Chen, SH., Wu, HH. *et al.*

# PHARMACOGENOMICS

## “BALANCING GENETIC PRIVACY AND PUBLIC HEALTH”

APRIL 12 2024



### INTRODUCTION TO PHARMACOGENOMICS:

The fields of pharmacogenomics (PGx), transcriptomics, proteomics, epigenomics, and metabolomics are swiftly advancing branches of -omics, offering promising avenues for discovering biomarkers to inform treatments and uncovering new disease mechanisms. Pharmacogenomics focuses on DNA variations, while epigenomics investigates heritable changes unrelated to DNA but impacting gene expression. Pharmacogenomic research traces its roots to 1931 with Snyder's seminal discovery of the heritability of phenylthiocarbamate taste sensitivity. Key advancements occurred in the 1960s through twin studies, revealing the genetic inheritance of interindividual metabolism variations by CYP2D6 and NAT2 enzymes. Subsequent decades, particularly the 1980s and 1990s, witnessed significant findings in common genetic variations within genes responsible for drug metabolism and transport. The decoding of the human genome in the early 2000s propelled rapid growth in pharmacogenomics.

The human body relies on oxygen for metabolic functions, with the lungs serving to continuously exchange gases with the external environment. However, the lungs also inadvertently inhale various pollutants such as pollen, germs, dust, smoke, and toxins, which can lead to the development of asthma, chronic obstructive pulmonary disease (COPD), lung cancer, and other respiratory disorders. Symptoms like breathlessness, chronic coughing, and hypoxia significantly impact patients' quality of life and productivity, with lung diseases being a leading cause of infant mortality. These conditions can affect different parts of the respiratory system, including the trachea, bronchi, bronchioles, alveoli, pleurae, and pleural cavity, disrupting the muscles and nerves involved in breathing. Lung disorders encompass a range of bacterial, viral, and respiratory syncytial virus infections like influenza, pneumococcal disease, COPD, idiopathic pulmonary fibrosis (IPF), and lung cancer. Failure to manage these disorders promptly can lead to severe health complications and even pose life-threatening risks. Patients with conditions like asthma or COPD experience various daily challenges, including strained relationships, anxiety, reduced physical activity, low self-esteem, fatigue, and depression. Respiratory disorders profoundly diminish patients' productivity and quality of life due to symptoms like dyspnea and persistent coughing. Studies indicate that approximately 34% and 40% of COPD patients suffer from anxiety and depression, respectively. This chapter serves as an overview of respiratory disorders and their pharmacological treatments.

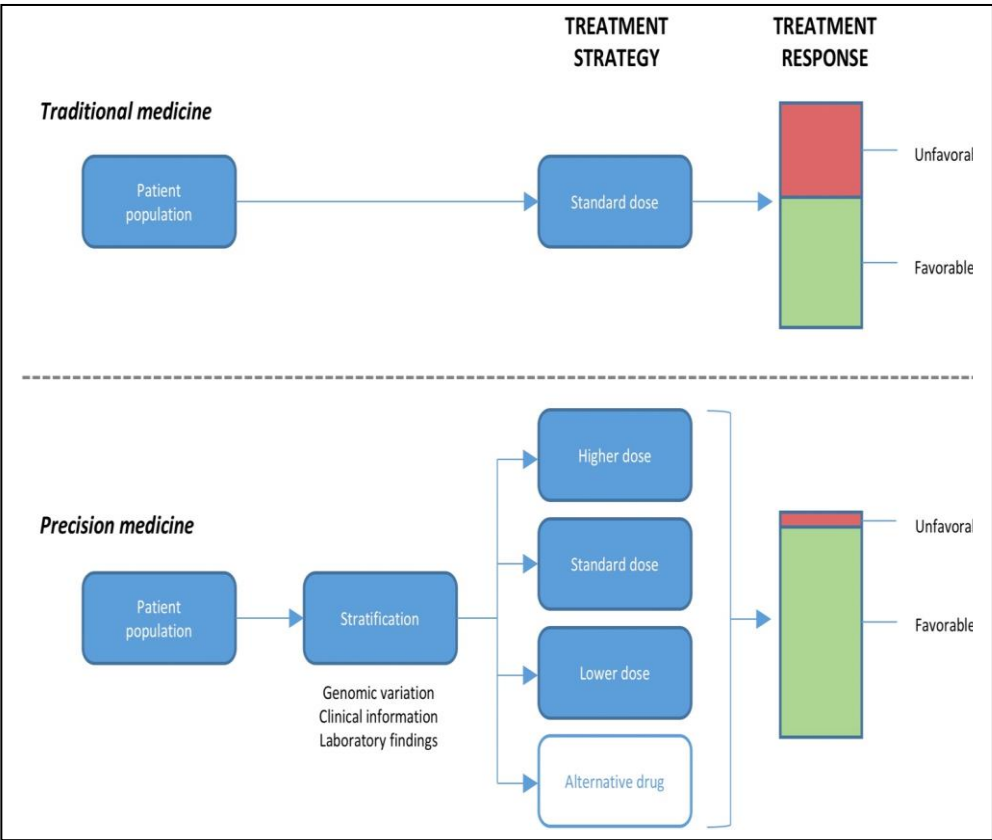
### CASE STUDY: The significance of Pharmacogenomics in Rare Diseases

Rare diseases have emerged as an increasingly significant focus in public health due to their combined occurrence and often grave nature. Initiatives like the Orphan Drug Act aim to boost the development of therapeutics for these conditions. However, despite the rigorous pre-market studies required for drug approval, these efforts often fail to fully explain the diversity in treatment responses, resulting in medications sometimes being more harmful or less effective than anticipated. Introducing a "Goal Line" to delineate the spectrum of phenotypic variations in response to medication, with "Goal Posts" marking the extreme points of this spectrum as (1) Super-Response, indicating an exceptional therapeutic effect, and (2) Serious Harm, can facilitate the investigation of pharmacogenomics behind these extremes. This exploration holds promise for new therapeutic developments, informing drug policy, and enhancing understanding of disease pathophysiology. In rare disease research, where cohorts are smaller than ideal, a combination of "small data" and "big data" approaches to data collection and analysis is essential for robust results. This paper underscores the importance of studying drug response alongside other rare disease research efforts and emphasizes the necessity for global collaboration in rare disease pharmacogenomics.

The variability in drug response from person to person presents a significant challenge both in clinical practice and drug development, often resulting in therapeutic failures or adverse drug reactions (ADRs) among individuals or specific patient subgroups. Aside from the high incidence of ADRs and their associated health risks, considerable financial burdens are also incurred. Various factors contribute to the risk of drug inefficacy or toxicity, including drug interactions, patient age, renal and liver function, other health conditions, and lifestyle factors like smoking and alcohol use. Additionally, recent years have highlighted the substantial impact of genetic factors on modifying drug responses or predisposing individuals to ADRs. Genetic polymorphisms in drug-metabolizing enzymes, drug receptors, and drug transporters have been linked to variations in drug efficacy and toxicity. Notably, migraine is recognized as a polygenic and multifactorial disorder, with genetic investigations implicating genes such as CACNA1A and DRD2 in its susceptibility.

PHARMACOGENOMICS COMMONLY USED  
ICU MEDICATIONS:

(1) Propofol, a widely utilized anesthetic and sedative in the ICU, undergoes glucuronidation and hydroxylation to form its inactive metabolite. The hydroxylation rate, which contributes significantly to individual variability, is primarily mediated by the CYP2B6 enzyme. (2) Midazolam, a frequently employed benzodiazepine for sedation in emergency departments and ICUs, undergoes metabolism primarily via the CYP3A4/3A5 enzyme to produce its active metabolite, alpha 1-hydroxymidazolam. (3) *Lorazepam*. Lorazepam is a benzodiazepine derivative with antianxiety and sedative-hypnotic properties. Lorazepam is mainly cleared by direct glucuronidation. (4) Diazepam, a benzodiazepine commonly employed for sedation or anticonvulsant therapy in the ICU, is metabolized by CYP 3A4 and 2C19 enzymes to produce its active metabolites, notably desmethyldiazepam.



PERSONALIZED MEDICINE: PHARMACOGENOMICS  
AND DRUG DEVELOPMENT

Personalized medicine aims to deliver the appropriate drug, at the correct dose, to the right patient. Pharmacogenomics (PGx) seeks to identify genetic variations that can impact drug efficacy and safety. Together, these fields encompass a broad spectrum, spanning from fundamental drug discovery research to understanding the genetic basis of drug interactions and responses, advancing novel drug development, conducting genetic assessments of patients, and managing clinical care accordingly. Ultimately, the goal of Pharmacogenomics is to anticipate how a patient's genetic makeup will influence their response to a specific drug, thereby optimizing medical treatment. By predicting individual drug responses, it becomes feasible to enhance therapeutic outcomes and reduce the occurrence of adverse effects.



APPLICATION OF PHARMACOGENOMICS

Numerous factors can influence a patient's response to a medication, encompassing intrinsic elements like age, gender, race/ethnicity, genetic makeup, existing health conditions, organ function impairments, as well as physiological changes such as pregnancy and lactation. Extrinsic factors like smoking, dietary habits (including food, beverages, and supplements), and concurrent medication usage further contribute to this complexity. The interplay of enzyme, transporter, and receptor genotypes, alongside other variables like concurrent medications and health conditions, can significantly impact the risk-to-benefit balance for individual patients. This commentary examines the optimal timing for acquiring genomic information during drug development and its integration into labeling and treatment standards to tailor drug therapy, thereby advancing the principles of personalized medicine.

REFERENCE:

1. The Role of Pharmacogenomics in Rare Diseases - Published: 14 March 2024
2. Pharmacogenomics Beyond Single Common Genetic Variants: The Way Forward - July 28, 2023

INTEGRATION OF MACHINE LARNING AND  
PHARMACOGENOMIC BIOMARKERS FOR  
PREDICTING RESPONSE TO ANTIDEPRESSANT  
TREATMENT

In conclusion, the strategic integration of pharmacogenomics and machine learning shows great potential in achieving accurate prediction of therapeutic outcomes for antidepressant treatment in patients with Major Depressive Disorder (MDD), facilitating their application in clinical settings. Additionally, it offers the opportunity to categorize depressed patients into subgroups, identifying those more likely to respond favorably to specific treatments. Advancements toward these objectives will be realized through the integration of new pharmacogenomic findings into predictive algorithms, either independently or in conjunction with other multi-omic and relevant clinical parameters. However, caution is warranted regarding potential methodological challenges associated with machine learning and artificial intelligence techniques for outcome prediction, particularly the risk of overfitting predictive models. Addressing these challenges will necessitate the collaboration of multidisciplinary teams comprising clinical, pharmacogenomics, informatics, and computer science experts, as well as access to large volumes of high-dimensional data. Furthermore, cooperative efforts among research teams and peer reviewers with expertise in statistical and machine learning methodologies will be crucial.

3. Pharmacogenomics: state of the research and perspectives in clinical application - May 2003
  4. Chapter 9 - Pharmacogenomics of pulmonary and respiratory diseases - 2023, Pages 223-246
  5. Pharmacogenomics of Medications Commonly Used in the Intensive Care Unit - 04 December 2018
  6. Pharmacogenomics and Global Precision Medicine in the Context of Adverse Drug Reactions: Top
  7. 10 Opportunities and Challenges for the Next Decade -1 October 2016
  8. Personalized Medicine: Pharmacogenomics and Drug Development
  9. Integration of machine learning and pharmacogenomic biomarkers for predicting response to antidepressant treatment: can computational intelligence be used to augment clinical assessments? - 27 Sep 2019
-

# Nikitha vinoth

*by* OfficialTurnitin

---

**Submission date:** 28-Apr-2024 11:37PM (UTC+0900)

**Submission ID:** 2362315444

**File name:** PHARMACOGENOMICS-NEWS-ARTICLE.docx (325.09K)

**Word count:** 707

**Character count:** 4734

# PHARMACOGENOMICS

## "BALANCING GENETIC PRIVACY AND PUBLIC HEALTH"

APRIL 12 2024



### INTRODUCTION TO PHARMACOGENOMICS:

The fields of pharmacogenomics (PGx), transcriptomics, proteinomics, epigenomics, and metabolomics are swiftly advancing branches of -omics, offering promising avenues for discovering biomarkers to inform treatments and uncovering new disease mechanisms. Pharmacogenomics focuses on DNA variations, while epigenomics investigates heritable changes unrelated to DNA but impacting gene expression. Pharmacogenomic research traces its roots to 1931 with Snyder's seminal discovery of the heritability of phenylthiocarbamate taste sensitivity. Key advancements occurred in the 1960s through twin studies, revealing the genetic inheritance of interindividual metabolism variations by CYP2D6 and NAT2 enzymes. Subsequent decades, particularly the 1980s and 1990s, witnessed significant findings in common genetic variations within genes responsible for drug metabolism and transport. The decoding of the human genome in the early 2000s propelled rapid growth in pharmacogenomics.

The human body relies on oxygen for metabolic functions, with the lungs serving to continuously exchange gases with the external environment. However, the lungs also inadvertently inhale various pollutants such as pollen, germs, dust, smoke, and toxins, which can lead to the development of asthma, chronic obstructive pulmonary disease (COPD), lung cancer, and other respiratory disorders. Symptoms like breathlessness, chronic coughing, and hypoxia significantly impact patients' quality of life and productivity, with lung diseases being a leading cause of infant mortality. These conditions can affect different parts of the respiratory system, including the trachea, bronchi, bronchioles, alveoli, pleurae, and pleural cavity, disrupting the muscles and nerves involved in breathing. Lung disorders encompass a range of bacterial, viral, and respiratory syncytial virus infections like influenza, pneumococcal disease, COPD, idiopathic pulmonary fibrosis (IPF), and lung cancer. Failure to manage these disorders promptly can lead to severe health complications and even pose life-threatening risks. Patients with conditions like asthma or COPD experience various daily challenges, including strained relationships, anxiety, reduced physical activity, low self-esteem, fatigue, and depression. Respiratory disorders profoundly diminish patients' productivity and quality of life due to symptoms like dyspnea and persistent coughing. Studies indicate that approximately 34% and 40% of COPD patients suffer from anxiety and depression, respectively. This chapter serves as an overview of respiratory disorders and their pharmacological treatments.

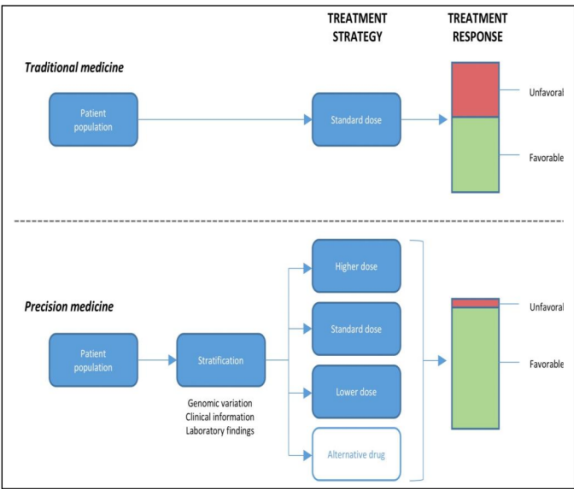
### CASE STUDY: The significance of Pharmacogenomics in Rare Diseases

Rare diseases have emerged as an increasingly significant focus in public health due to their combined occurrence and often grave nature. Initiatives like the Orphan Drug Act aim to boost the development of therapeutics for these conditions. However, despite the rigorous pre-market studies required for drug approval, these efforts often fail to fully explain the diversity in treatment responses, resulting in medications sometimes being more harmful or less effective than anticipated. Introducing a "Goal Line" to delineate the spectrum of phenotypic variations in response to medication, with "Goal Posts" marking the extreme points of this spectrum as (1) Super-Response, indicating an exceptional therapeutic effect, and (2) Serious Harm, can facilitate the investigation of pharmacogenomics behind these extremes. This exploration holds promise for new therapeutic developments, informing drug policy, and enhancing understanding of disease pathophysiology. In rare disease research, where cohorts are smaller than ideal, a combination of "small data" and "big data" approaches to data collection and analysis is essential for robust results. This paper underscores the importance of studying drug response alongside other rare disease research efforts and emphasizes the necessity for global collaboration in rare disease pharmacogenomics.

The variability in drug response from person to person presents a significant challenge both in clinical practice and drug development, often resulting in therapeutic failures or adverse drug reactions (ADRs) among individuals or specific patient subgroups. Aside from the high incidence of ADRs and their associated health risks, considerable financial burdens are also incurred. Various factors contribute to the risk of drug inefficacy or toxicity, including drug interactions, patient age, renal and liver function, other health conditions, and lifestyle factors like smoking and alcohol use. Additionally, recent years have highlighted the substantial impact of genetic factors on modifying drug responses or predisposing individuals to ADRs. Genetic polymorphisms in drug-metabolizing enzymes, drug receptors, and drug transporters have been linked to variations in drug efficacy and toxicity. Notably, migraine is recognized as a polygenic and multifactorial disorder, with genetic investigations implicating genes such as CACNA1A and DRD2 in its susceptibility.

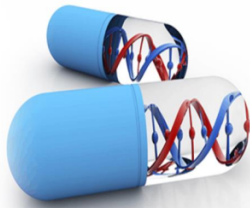
PHARMACOGENOMICS COMMONLY USED ICU MEDICATIONS:

(1) Propofol, a widely utilized anesthetic and sedative in the ICU, undergoes glucuronidation and hydroxylation to form its inactive metabolite. The hydroxylation rate, which contributes significantly to individual variability, is primarily mediated by the CYP2B6 enzyme. (2) Midazolam, a frequently employed benzodiazepine for sedation in emergency departments and ICUs, undergoes metabolism primarily via the CYP3A4/3A5 enzyme to produce its active metabolite, alpha 1-hydroxymidazolam. (3) Lorazepam. Lorazepam is a benzodiazepine derivative with antianxiety and sedative-hypnotic properties. Lorazepam is mainly cleared by direct glucuronidation. (4) Diazepam, a benzodiazepine commonly employed for sedation or anticonvulsant therapy in the ICU, is metabolized by CYP 3A4 and 2C19 enzymes to produce its active metabolites, notably desmethyldiazepam.



PERSONALIZED MEDICINE: PHARMACOGENOMICS AND DRUG DEVELOPMENT

Personalized medicine aims to deliver the appropriate drug, at the correct dose, to the right patient. Pharmacogenomics (PGx) seeks to identify genetic variations that can impact drug efficacy and safety. Together, these fields encompass a broad spectrum, spanning from fundamental drug discovery research to understanding the genetic basis of drug interactions and responses, advancing novel drug development, conducting genetic assessments of patients, and managing clinical care accordingly. Ultimately, the goal of Pharmacogenomics is to anticipate how a patient's genetic makeup will influence their response to a specific drug, thereby optimizing medical treatment. By predicting individual drug responses, it becomes feasible to enhance therapeutic outcomes and reduce the occurrence of adverse effects.



INTEGRATION OF MACHINE LEARNING AND PHARMACOGENOMIC BIOMARKERS FOR PREDICTING RESPONSE TO ANTIDEPRESSANT TREATMENT

In conclusion, the strategic integration of pharmacogenomics and machine learning shows great potential in achieving accurate prediction of therapeutic outcomes for antidepressant treatment in patients with Major Depressive Disorder (MDD), facilitating their application in clinical settings. Additionally, it offers the opportunity to categorize depressed patients into subgroups, identifying those more likely to respond favorably to specific treatments. Advancements toward these objectives will be realized through the integration of new pharmacogenomic findings into predictive algorithms, either independently or in conjunction with other multi-omic and relevant clinical parameters. However, caution is warranted regarding potential methodological challenges associated with machine learning and artificial intelligence techniques for outcome prediction, particularly the risk of overfitting predictive models. Addressing these challenges will necessitate the collaboration of multidisciplinary teams comprising clinical, pharmacogenomics, informatics, and computer science experts, as well as access to large volumes of high-dimensional data. Furthermore, cooperative efforts among research teams and peer reviewers with expertise in statistical and machine learning methodologies will be crucial.

APPLICATION OF PHARMACOGENOMICS

Numerous factors can influence a patient's response to a medication, encompassing intrinsic elements like age, gender, race/ethnicity, genetic makeup, existing health conditions, organ function impairments, as well as physiological changes such as pregnancy and lactation. Extrinsic factors like smoking, dietary habits (including food, beverages, and supplements), and concurrent medication usage further contribute to this complexity. The interplay of enzyme, transporter, and receptor genotypes, alongside other variables like concurrent medications and health conditions, can significantly impact the risk-to-benefit balance for individual patients. This commentary examines the optimal timing for acquiring genomic information during drug development and its integration into labeling and treatment standards to tailor drug therapy, thereby advancing the principles of personalized medicine.

## ORIGINALITY REPORT

10%

SIMILARITY INDEX

3%

INTERNET SOURCES

10%

PUBLICATIONS

0%

STUDENT PAPERS

## PRIMARY SOURCES

1

Arjun P Athreya, Ravishankar Iyer, Liewei Wang, Richard M Weinshilboum, William V Bobo. "Integration of machine learning and pharmacogenomic biomarkers for predicting response to antidepressant treatment: can computational intelligence be used to augment clinical assessments?", Pharmacogenomics, 2019

Publication

7%

2

[www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)

Internet Source

3%

Exclude quotes Off

Exclude bibliography On

Exclude matches Off



## Course Completion Certificate

Nikitha V

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

**MATLAB Onramp**



DIRECTOR, TRAINING SERVICES

12 November 2023



## Course Completion Certificate

Nikitha V

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

Object-Oriented Programming Onramp



DIRECTOR, TRAINING SERVICES

13 November 2023



## Course Completion Certificate

Nikitha V

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

**MATLAB Fundamentals**



DIRECTOR, TRAINING SERVICES

17 November 2023



## Course Completion Certificate

**Nikitha V**

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

**Machine Learning Onramp**



DIRECTOR, TRAINING SERVICES

15 December 2023



## Course Completion Certificate

Nikitha V

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

**Machine Learning with MATLAB**

A handwritten signature in black ink, which appears to read 'Craig Santos'. The signature is written in a cursive style and is positioned above a thin horizontal line.

DIRECTOR, TRAINING SERVICES

19 December 2023



## Course Completion Certificate

Nikitha V

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

Deep Learning Onramp



DIRECTOR, TRAINING SERVICES

20 December 2023



## Course Completion Certificate

Nikitha V

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

**Deep Learning with MATLAB**

A handwritten signature in black ink, which appears to read 'Craig Santos'. The signature is written in a cursive style and is positioned above a thin horizontal line.

DIRECTOR, TRAINING SERVICES

20 December 2023



## Course Completion Certificate

Nikitha V

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

Computer Vision Onramp



DIRECTOR, TRAINING SERVICES

20 December 2023



## Course Completion Certificate

Nikitha V

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

Reinforcement Learning Onramp



DIRECTOR, TRAINING SERVICES

23 December 2023



## Course Completion Certificate

Nikitha V

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

Statistics Onramp

A handwritten signature in black ink, appearing to read 'Craig Santos', written over a thin horizontal line.

DIRECTOR, TRAINING SERVICES

4 March 2024



## Course Completion Certificate

**Nikitha V**

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

**MATLAB for Data Processing and Visualization**



DIRECTOR, TRAINING SERVICES

5 March 2024



## Course Completion Certificate

Nikitha V

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

Image Processing Onramp



DIRECTOR, TRAINING SERVICES

5 March 2024



## Course Completion Certificate

**Nikitha V**

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

**Image Processing with MATLAB**



DIRECTOR, TRAINING SERVICES

10 March 2024



## Course Completion Certificate

Nikitha V

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

Signal Processing with MATLAB

A handwritten signature in black ink, which appears to read 'Craig Santos'. The signature is written in a cursive style with a horizontal line underneath it.

DIRECTOR, TRAINING SERVICES

12 March 2024



## Course Completion Certificate

**Nikitha V**

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

**Wireless Communications Onramp**



DIRECTOR, TRAINING SERVICES

13 March 2024



## Course Completion Certificate

Nikitha V

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

**Optimization Onramp**



DIRECTOR, TRAINING SERVICES

13 March 2024



## Course Completion Certificate

Nikitha V

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

**MATLAB Programming Techniques**



DIRECTOR, TRAINING SERVICES

13 March 2024



## Course Completion Certificate

**Nikitha V**

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

**Solving Nonlinear Equations with MATLAB**



DIRECTOR, TRAINING SERVICES

20 March 2024