Annals of Bangladesh Agriculture
Journal homepage: gau.edu.bd/aba

ORIGINAL ARTICLE

A multi-stage computational and bioinformatics framework for the identification and validation of hub toxicogenomic biomarkers

Mohammad Nazmol Hasan[1*], Mohammad Shah Alam[2] and Md Mamunur Rahman[3]

[1] Department of Agricultural and Applied Statistics, Gazipur Agricultural University, Gazipur 1706, Bangladesh
[2] Department of Anatomy and Histology, Gazipur Agricultural University, Gazipur, 1706, Bangladesh
[3] Department of Entomology, Gazipur Agricultural University, Gazipur, 1706, Bangladesh.

ARTICLE INFO.

Keywords:
chemical toxicity, toxicogenomic biomarker, statistical methods, machine learning approaches, bioinformatics approaches, protein-protein interaction network.

Received : 28 August 2025
Revised : 24 October 2025
Accepted : 24 December 2025
Published : 05 January 2026

Citation:

Hasan, M. N., M. S. Alam and M. M. Rahman. 2026. A multi-stage computational and bioinformatics framework for the identification and validation of hub toxicogenomic biomarkers. *Ann. Bangladesh Agric.* 29(2): 93-109

ABSTRACT

Chemical toxicity is challenging to mitigate, necessitating a revisit to seed compound screening. Safety is crucial in approving drugs, pesticides, and cosmetics, necessitating the identification of safety biomarkers, such as toxicogenomic biomarkers (ToxBG), to predict potential toxicity. In this regard, we proposed a sequence of computational and bioinformatics approaches to identify key/hub ToxBG (HToxBG) for predicting chemical toxicity. In this sequence, we initially identified ToxBGs using statistical approaches, such as t-test, Wilcoxon signed-rank test (WSR-test), and linear model for microarray data analysis (LIMMA), based on the chemically treated and control samples of gene expression data collected from the online database "Toxygates." In the treatment group, rat samples were treated with chemicals (acetaminophen, bromobenzene, coumarin, methapyrilene, and nitrofurazone) with three dose levels, and gene expression data were collected at multiple time points. These statistical approaches, including the t-test, WSR-test, and LIMMA, identified 3,856, 3,232, and 3,377 ToxBGs, respectively. Of these, 2,877 were common and considered second-stage ToxBGs. This study validated the second-stage ToxBGs using four machine learning (ML) approaches. Among these ML approaches, the support vector machine (SVM) achieved higher accuracy in classifying treated and control samples, yielding sensitivity of 0.98, specificity of 0.97, accuracy of 0.98, and AUC (0.99) compared to other methods. The second-stage ToxBGs were also co-clustered with their associated chemicals. The protein-protein interaction (PPI) network analysis predicted that the second-stage ToxBGs were enriched in the biological pathways that perform important functions. Additionally, these ToxBGs were also enriched in different diseases like liver cirrhosis, HIV coinfection, gastric cancer, generalized hypotonia, neoplasm of the liver, etc. Out of 2877 common ToxBGs, 160 key/hub ToxBGs (HToxBGs) have been identified, 70 genes associated with disease states, and 90 involved in critical biological pathways, enabling the study of chemical toxicity. Therefore, the proposed sequence of computational and bioinformatics approaches can be used to identify HToxBGs and predict chemical toxicity.

*Corresponding Author: Department of Agricultural and Applied Statistics, Gazipur Agricultural University, Gazipur 1706, Bangladesh. Email: nazmol.stat.bioin@gau.edu.bd

https://doi.org/10.3329/aba.v29i2.84008

Introduction

Toxicogenomics examines the role of genes and chemicals, medications, or environmental stressors in the development of disease in people, animals, and plants by combining transcript, protein, and metabolite profiling with traditional toxicology. Several toxicants' actions and illness-causing effects have been made clear by the patterns of changed molecular expression brought on by particular exposures or disease consequences (Waters and Fostel, 2004; Afshari *et al*., 2011; Hasan *et al*., 2018). This health hazards are due to the toxicity of the toxins (small molecules, peptides, or proteins), environmental stressors (drought, heat, salinity, heavy metal, biotic stress) and chemical agents (drugs, gasoline, alcohol, pesticides, fuel oil, and cosmetics) in organism (NRC, 2007; Afshari *et al*., 2011; Hasan *et al*., 2019). Last century's industrial activities significantly increased the quantity of heavy metals to which people are exposed. Arsenic, cadmium, chromium, lead, and mercury have been the most frequently found heavy metals to cause poisoning in humans. Chromium, cadmium, and arsenic are among the harmful metals that can lead to genomic instability. They have been thought to be carcinogenic due to defects in DNA repair after the three metals produce oxidative stress and DNA damage (Azeh Engwa *et al*., 2019; Balali-Mood *et al*., 2021). Pesticides used in Chile's agriculture sector have neurotoxic effects that raise the risk of Parkinson's and Alzheimer's disease in workers who are exposed to high levels of these chemicals. (Lucero *et al*., 2019). Similarly, noise, emotional stress, and physicochemical agents are examples of environmental risk factors that significantly affect human health. This exposure to the environment may cause 14 alterations in non-coding RNAs (ncRNAs) as well as epigenetic reprogramming (Miguel *et al*., 2020). The unintended consequence of a medicine that causes death or morbidity with symptoms severe enough to make a patient seek medical attention and/or need hospitalization is known as a drug-induced disease (Krueger, 2006). The most frequent cause of acute liver failure in the western world is still drug-induced liver damage (DILI). The medicine in question must be stopped immediately upon the onset of DILI, particularly if there is jaundice and/or increased transaminases (Laster and Satoskar, 2015). The availability of sensitive, specific, and broadly applicable biomarkers of toxic effects, and the term ToxBG refers to genes induced by toxicants. Toxicogenomics has been applied at every level of chemical risk assessment, and it is believed that changes in gene expression may be employed as biomarkers of harmful effects (Hasan *et al*., 2018). Successively, functional analysis of these ToxBGs can efficiently predict the extent of toxicity, probable health hazards, and disease-causing ability that will appear over time. Therefore, the identification of ToxBGs and their functional analysis will guide the prediction of chemical/drug/environmental stress toxicity before phenotypic changes appear, allowing for preventive measures to be taken. To address these issues in this study, an attempt is made to identify ToxBGs using statistical and ML approaches, and their functional analysis using integrated bioinformatics analysis.

Materials and Methods

Gene expression data collection from chemically treated and control rat samples

To examine the toxicity of chemicals, we use genome-wide gene expression data from the Japanese Toxicogenomics Project (TGP) (Uehara *et al.*, 2010). Here, we consider genome-wide *in vivo* gene expression data from treatment and control samples of the rat's (*Rattus norvegicus*) liver. In the treatment samples, a homogeneous group of rats was treated with three dose levels of chemicals: acetaminophen, bromobenzene, coumarin, methapyrilene, and nitrofurazone. Thereafter, gene expression data from treated rat liver (*in vivo* conditions) were collected at four time intervals (3 hours, 6 hours, 9 hours, and 24 hours), which constitute 60 treated samples. Additionally, in the experiment, there was a control sample against each of the treated samples, which constituted 60 control samples. We have downloaded gene expression data for these samples from the online toxicogenomic database "Toxygates" (Nyström-Persson *et al.*, 2013, 2017) (https://toxygates.nibiohn.go.jp/toxygates/#columns).

Identification and validation of the ToxBG using statistical and ML approaches

In this section, firstly, we have identified the ToxBGs using statistical approaches. The two-sample t-test under the assumption of equal variances, WSR-test, and LIMMA (Ritchie *et al.*, 2015) were used for the identification of ToxBGs. The downloaded gene expression data were analyzed based on the mentioned statistical approaches using the base package and "limma" of the R programming language software. ToxBGs identified by the two-sample t-test, WSR-test, and LIMMA were considered as the first-stage ToxBG. However, the common genes identified by these methods were declared as the second-stage ToxBG. On the other hand, for the validation of the identified second-stage ToxBGs, we have used different ML approaches like Linear Discriminant Analysis (LDA) (Ye and Wang, 2006), Logistic Regression (LR) (Boateng and Abaye, 2019), Support Vector Machine (SVM) (Schölkopf, 2003; Fernandes de Mello and Antonelli Ponti, 2018; Mohsin Abdulazeez *et al.*, 2020), and Random Forest (RF) (Dong *et al.*, 2020). We have evaluated the performance of these ML approaches using the evaluation metric accuracy, area under the curve (AUC), sensitivity, and specificity. All the ML and evaluation approaches were analyzed using R packages "caret" and "pROC", respectively. Since the chemicals with similar characteristics are associated with a set of ToxBGs (Hamadeh *et al.***,** 2002; Hasan *et al.*, 2019, 2025). Therefore, finding the chemicals and their associated ToxBGs is another way to interpret the toxicity of chemicals. In this study, we have used robust hierarchical co-clustering (Hasan *et al.*, 2025) to identify the gene-chemical association using rhcolcust (Badsha *et al.*, 2020) package in R.

Bioinformatics approaches

To understand the characteristics and functions of the second-stage ToxBGs, an integrated bioinformatics analysis was done. This bioinformatics analysis was also used to narrow down the number of or to find the most important or third/final stage of ToxBGs. The bioinformatics analysis includes functional and pathway enrichment analysis, PPI network analysis, and disease enrichment analysis.

Functional and pathway enrichment analysis of ToxBG

The Kyoto Encyclopedia of Genes and Genomes (KEGG) and Functional Gene Ontology (GO) pathway enrichment analysis is a popular method for identifying the pathways, molecular functions (MF), biological processes (BP), and cellular components (CC) (Kanehisa *et al*., 2016). BP is a change or series of changes that take place throughout the cell's granularity period and are mediated by one or more gene products for various biological purposes (Carbon *et al.*, 2021). Gene products' biochemical actions are known as MFs. A gene product's active location within a cell is known as the CC (Carbon *et al.*, 2021). A set of experimentally verified pathway maps known as the KEGG pathway illustrates our understanding of the networks of molecular interactions, reactions, and relationships involved in metabolism, cellular functions, genetic information processing, organismal systems, environmental information processing, human diseases, and drug development (Kanehisa *et al.*, 2023). We performed ToxBG functional and pathway enrichment analysis using the NetworkAnalyst tool with GO and KEGG databases (Xia *et al*., 2014). To assess the statistical significance of the functional enrichment analysis, Fisher's exact test was used, with a cut-off adjusted p-value<0.05. Once more, we used three well-known tools, DAVID (Huang *et al*., 2007), EnrichR (Chen *et al*., 2013), and Metascape (Zhou *et al*., 2019) to perform functional and pathway enrichment analysis utilizing GO and KEGG databases. And finally, we suggested a common, significantly enriched term (i.e., a term that is statistically significant and enriched in every tool) to ensure the reliability of the results.

PPI network analysis of ToxBG

PPIs are the physical attraction of two or more protein molecules brought on by biochemical events that are guided by the hydrophobic effect, electrostatic forces, and hydrogen bonds. In most cases, a protein cannot function without interacting with one or more other proteins (Seychell and Beck, 2021). According to Braun and Gingras (2012) (Braun and Gingras, 2012), the PPIs aid in the creation of bigger protein complexes that carry out particular tasks. Numerous molecular and biological activities are carried out by it, including protein function, cell-to-cell contacts, metabolic and developmental control, the occurrence of illness, and the invention of therapies. An undirected graph is used to depict a PPI network, with nodes denoting proteins and their interactions denoted by edges. The top-ranked hub protein is a node that has the most important interactions, linkages, or edges with other nodes. Consequently, HToxBGs/proteins can be investigated using the PPI network analysis of ToxBG. To identify HToxBG that enriched to functional pathways and diseases through PPI network. The PPI network of ToxBG was built in this study using the STRING database (Szklarczyk *et al*., 2019). The PPI network was visualized, and topological studies were conducted using Cytoscape 3.8.0 and NetworkAnalyst (Xia *et al*., 2014). A medium confidence score of 900 was utilized as the PPI cutoff value. Using a topological degree of measurement (> 25), the HToxBGs in the PPI network are located.

Results

First stage ToxBG identification by statistical approaches

The t-test identified 3856 biomarker genes, the WSR-test identified 3232 biomarker genes, and LIMMA identified 3377 biomarker genes, and they were considered as the first stage ToxBG for measuring the toxicity of the mentioned chemicals. The common 2877 biomarker gene for all the mentioned test-statistic was considered as the second-stage ToxBG (Figure 1). In Figure 1, we showed the t-test, WSR-test, and LIMMA identified first-stage ToxBGs and common second-stage ToxBGs. The co-cluster (Figure 2) showed the association between the second-stage ToxBGs and chemicals. From the bottom-left corner of Figure 2, the ToxBGs and chemicals were clustered chronologically according to ascending order of numeric.

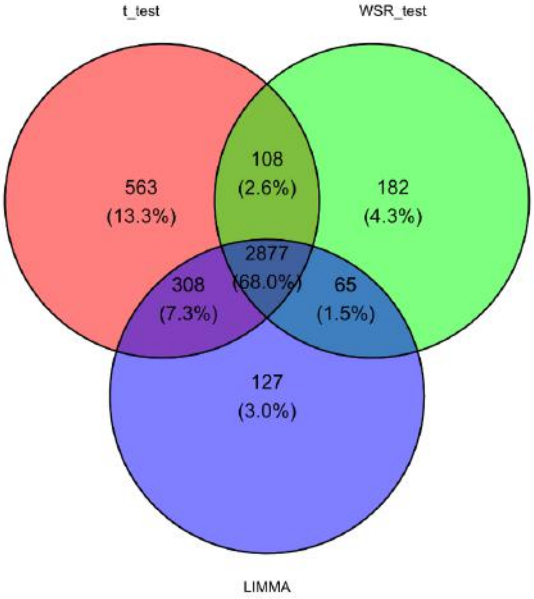


Figure 1. Venn diagram of the ToxBGs identified by the t-test, WSR-test, and LIMMA, based on p-value < 0.01, and common biomarker genes identified by all the tests.

Validation of the identified ToxBG by the ML approaches

We validated the second-stage common ToxBGs by the ML approaches like LDA, LR, SVM, and RF. In this regard, accuracy, AUC, sensitivity, and specificity were used as performance evaluation metrics. In Table 1, we presented different performance evaluation scores against different ML approaches. The performance evaluation scores (accuracy = 0.9806, AUC = 0.9995, sensitivity = 0.9833, specificity = 0.9722) were highest for the

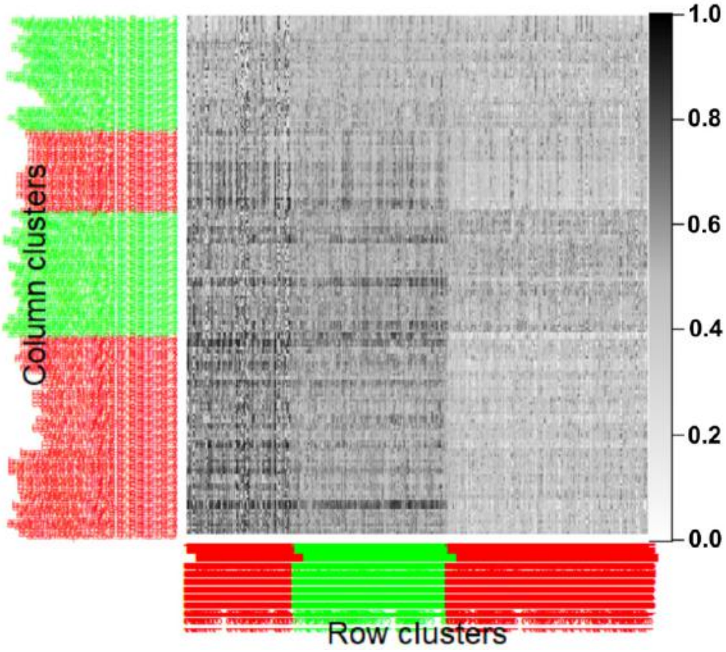


Figure 2. Co-cluster or association between second-stage ToxBGs and chemicals with different doses and time points. In the figure, the horizontal axis aligns gene clusters, and the vertical axis aligns chemical clusters with different doses and time points.

SVM (Table 1). Therefore, we can conclude that the SVM is the better ML approach for classifying the case (chemically treated sample) and control sample based on the second-stage common ToxBGs. Additionally, based on the results of accuracy, AUC, sensitivity, and specificity (Table 1), it could be concluded that the identified ToxBGs can efficiently classify the treated and control group of samples, and the identified second-stage ToxBGs are the candidate HToxBGs.

Identification of the third stage ToxBG using bioinformatics approaches

The statistical approaches identified 2877 common second-stage ToxBGs, which is a very large number. Measuring the toxicity of chemicals based on this large number of ToxBGs is very challenging. Therefore, we narrowed down this large number of ToxBGs based on the hub gene identification technique using the KEGG pathway enrichment and disease enrichment analysis. These bioinformatics approaches are described in the subsequent sections.

Identification of the third stage ToxBG using KEGG pathway enrichment analysis

The pathway enrichment analysis was done using the DAVID online platform for functional enrichment and pathway analysis. The second stage ToxBGs were significantly

Table 1. Second stage common ToxBGs validation by the performance evaluation metrics of the ML approaches

ML Approaches	Performance evaluation metrics			
	Accuracy	AUC	Sensitivity	Specificity
RF	0.8694	0.9654	0.9759	0.5500
LR	0.5333	0.5544	0.5537	0.4722
SVM	0.9806	0.9995	0.9833	0.9722
LDA	0.9125	0.9609	0.9574	0.7778

enriched in the rno03008: Ribosome biogenesis in eukaryotes, rno01100: Metabolic pathways, rno03015: mRNA surveillance pathway, rno03020: RNA polymerase, rno00100: Steroid biosynthesis, rno04216: Ferroptosis, rno03013: Nucleocytoplasmic transport, rno00900: Terpenoid backbone biosynthesis, and rno00480: Glutathione metabolism pathways. Thereafter, protein-protein interaction network analysis was done for ToxBGs enriched in the mentioned pathway using the string database (Figure 3). The results of the string database were used to find HToxBGs for declaring the third stage ToxBGs. The 90 third-stage HToxBGs were presented in Table 2, and the network of HToxBGs for each of the significant pathways was given in Figure 4.

Identification of the third-stage ToxBG using disease enrichment analysis

The common ToxBGs identified by the statistical approaches were then analyzed using the string database and Cytoscape to get

Table 2. HToxBGs for significantly enriched pathways that were extracted from the protein-protein interaction network

rno03008: Ribosome biogenesis in eukaryotes	rno01100: Metabolic pathways	rno03015: mRNA surveillance pathway	rno03020: RNA polymerase	rno00100: Steroid biosynthesis	rno04216: Ferroptosis	rno03013: Nucleocytoplasmic transport	rno00900: Terpenoid backbone biosynthesis	rno00480: Glutathione metabolism
Nat10	Sqle	Ncbp1	Polr2c	Msmo1	Hmox1	Nup153	Fdps	Gsr
Nmd3	Cyp51	Eif4a3	Polr2f	Lss	Ftl1	Ranbp2	Hmgcr	Gss
Nop58	Hmgcr	Rnps1	Polr2e	Cyp51	Tfrc	Eif4a3	Mvd	Gstp1
Utp18	Hmgcs1	Upf3b	Polr1a	Fdft1	Steap3	Nup93	Idi1	Gstm1
Utp14a	Fdps	Casc3	Polr3e	Dhcr24	Slc40a1	Nup54	Mvk	Hpgds
Wdr3	Fdft1	Magoh	Polr1c	Tm7sf2	Slc7a11	Nup160	Acat2	Gstm3
Pwp2	Acat2	Alyref	Polr3c	Nsdhl	Pcbp2	Nup58	Pmvk	Mgst2
Wdr43	Mvd	Cpsf6	Crcp	Sqle	Gclc	Alyref	Hmgcs1	Gstm4
Gnl3	Lss	Clp1	Polr3d	Sc5d	Gclm	Ncbp1	Dhdds	Gsta3
Mphosph10	Mvk	Ddx39b	Polr1b	Ebp	Tp53	Pom121	Nus1	Prdx6

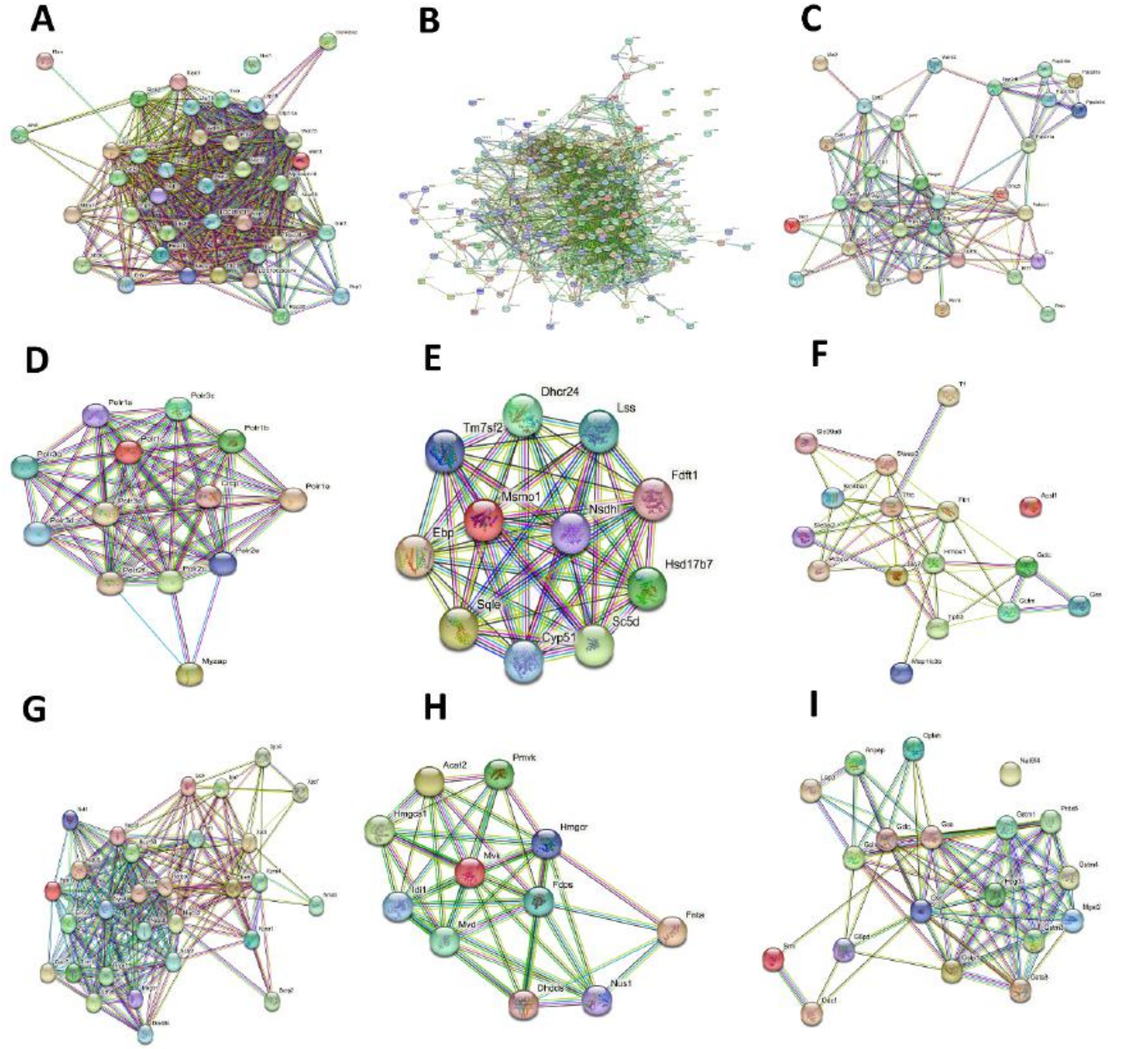


Figure 3. PPI network analysis using the string database of the different significant KEGG pathway-enriched ToxBGs. The enrichment analysis was done using the DAVID online bioinformatics database tool. In figure A) rno03008: Ribosome biogenesis in eukaryotes, B) rno01100: Metabolic pathways, C) rno03015: mRNA surveillance pathway, D) rno03020: RNA polymerase, E) rno00100: Steroid biosynthesis, F) rno04216: Ferroptosis, G) rno03013: Nucleocytoplasmic transport, H) rno00900: Terpenoid backbone biosynthesis, and I) rno00480: Glutathione metabolism.

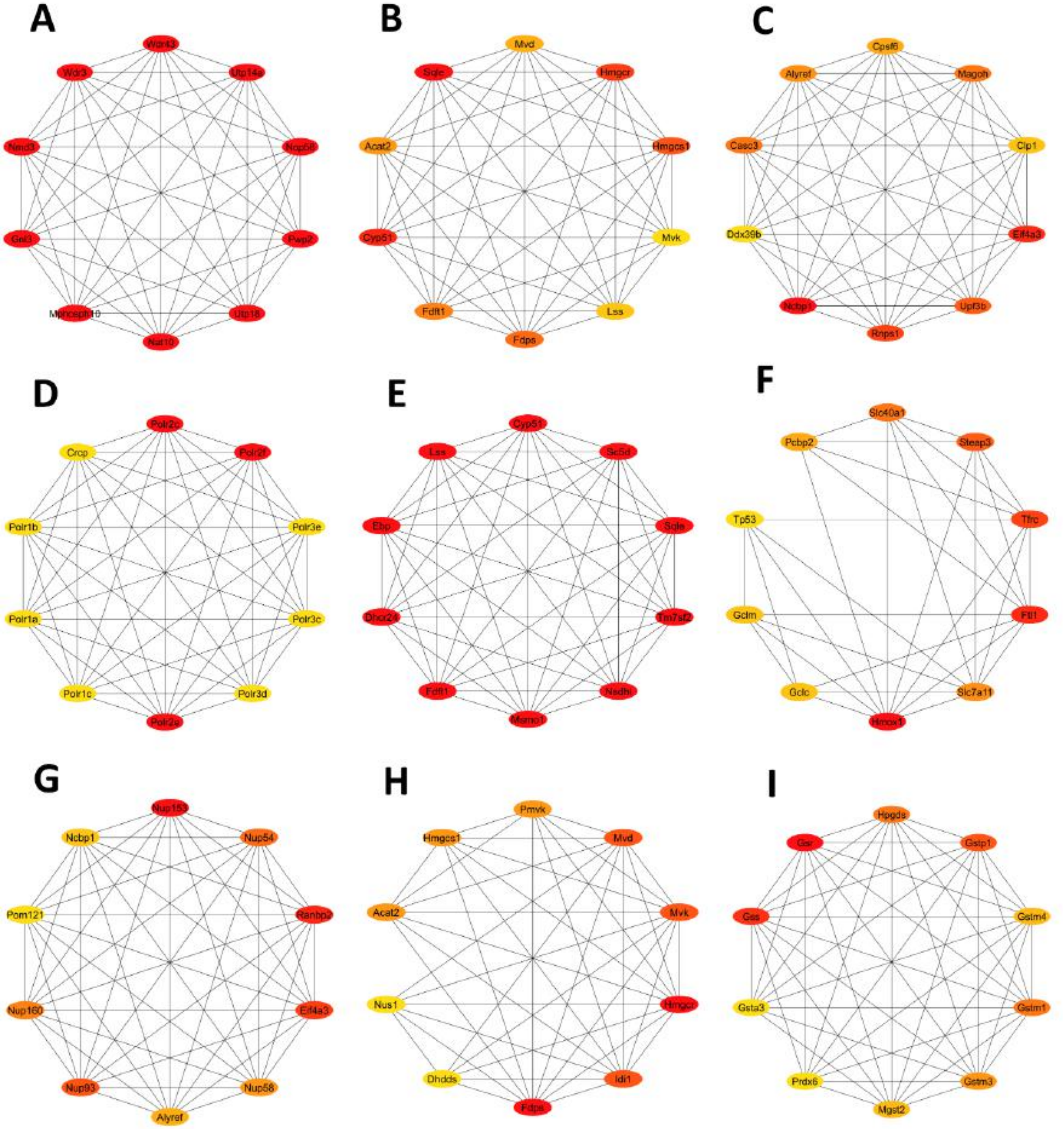
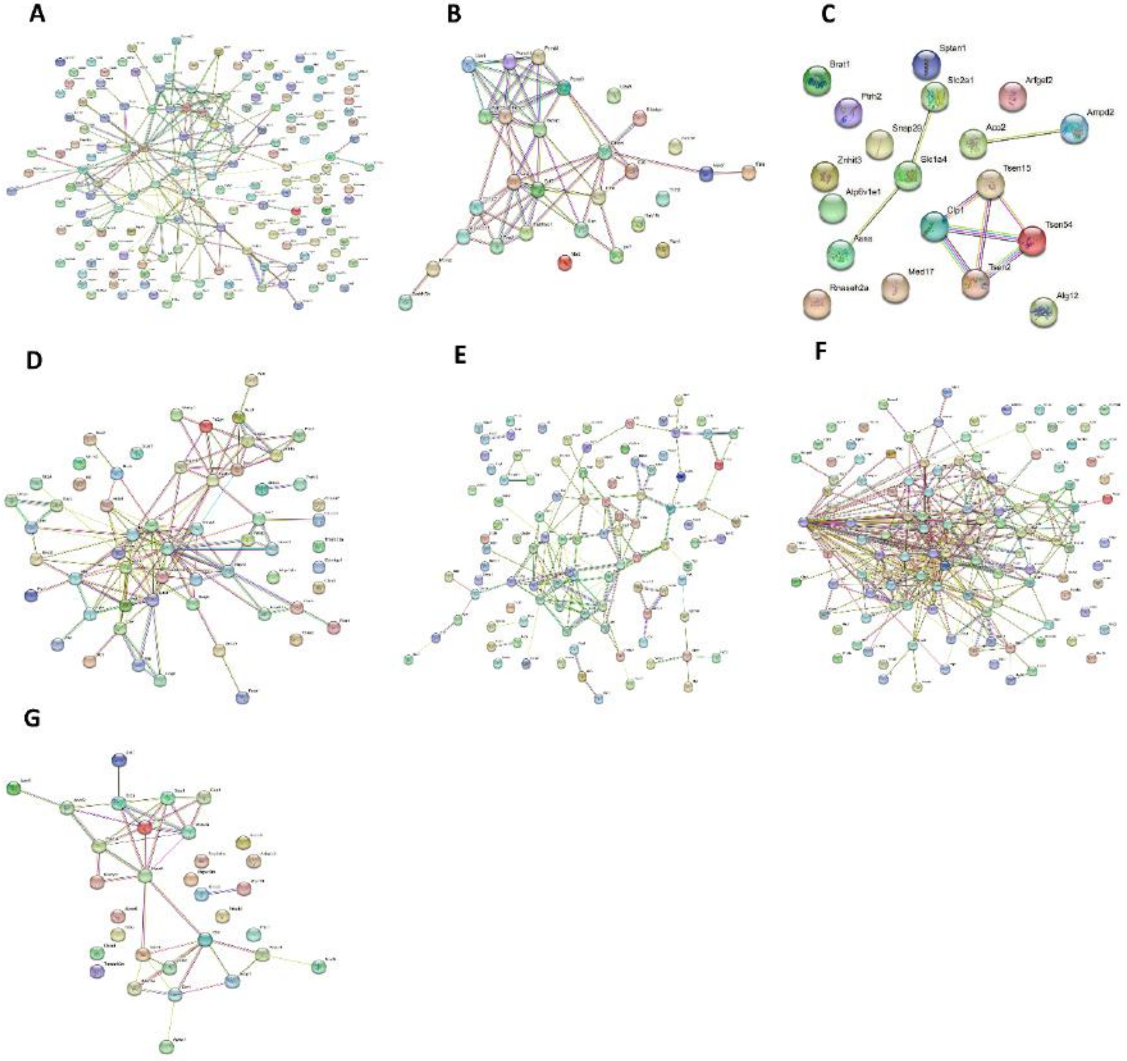


Figure 4. The PPI network of the HToxBGs for different pathways. In the figure A) rno03008: Ribosome biogenesis in eukaryotes, B) rno01100: Metabolic pathways, C) rno03015: mRNA surveillance pathway, D) rno03020: RNA polymerase, E) rno00100: Steroid biosynthesis, F) rno04216: Ferroptosis, G) rno03013: Nucleocytoplasmic transport, H) rno00900: Terpenoid backbone biosynthesis, and I) rno00480: Glutathione metabolism.

the third stage ToxBGs that create diseases. In Figure 5, we presented the PPI network of the second-stage ToxBGs that significantly enriched different diseases. Similarly, in Table 3 and Figure 6, we showed the third stage HToxBGs. We identified 70 HToxBGs in the third stage for disease causation.



Figurere 5. PPI network analysis using the string database of the different significant diseases enriched by ToxBGs. The enrichment analysis was done using the Enrichr online bioinformatics database tool. In the figure, A) Liver Cirrhosis, B) HIV Coinfection, C) Progressive microcephaly, D) Hereditary Diffuse Gastric Cancer, E) Generalized hypotonia, F) Malignant neoplasm of liver, and G) Disease Exacerbation.

Table 3. HToxBGs for significantly enriched diseases that were extracted from the PPI network

Liver Cirrhosis	HIV Coinfection	Progressive microcephaly	Hereditary Diffuse Gastric Cancer	Generalized hypotonia	Malignant neoplasm of the liver	Disease Exacerbation
Aldh1a1	*Cct5*	*Tsen15*	*Tp53*	*Pc*	*Tp53*	*Rps6*
Hmox1	*Cct3*	*Tsen2*	*Myc*	*Suox*	*Ccnd1*	*Nop56*
Igf1	*Psma5*	*Clp1*	*Ccnd1*	*Glul*	*Myc*	*Myc*
Pxdn	*Psmc5*	*Tsen54*	*Cdh1*	*Acaca*	*Igf1*	*Bop1*
Hmgcr	*Eif4a1*	*Slc1a4*	*Rps6*	*Acat2*	*Esr1*	*Wdr46*
Myc	*Psme3*	*Aco2*	*Jun*	*Aco2*	*Jun*	*Pa2g4*
Hao1	*Psmd6*	*Ampd2*	*Hmox1*	*Hlcs*	*Hsp90aa1*	*Rrp9*
Cxcl12	*Odc1*	*Asns*	*Casp8*	*Atic*	*Cdh1*	*Esr1*
Srebf1	*Psmd13*	*Slc2a1*	*Mapk1*	*Aldh5a1*	*Hmox1*	*Cdh1*
Abat	*Psmd3*	*Tsen15*	*Fgfr2*	*Tsfm*	*Kdr*	*Kdm1a*

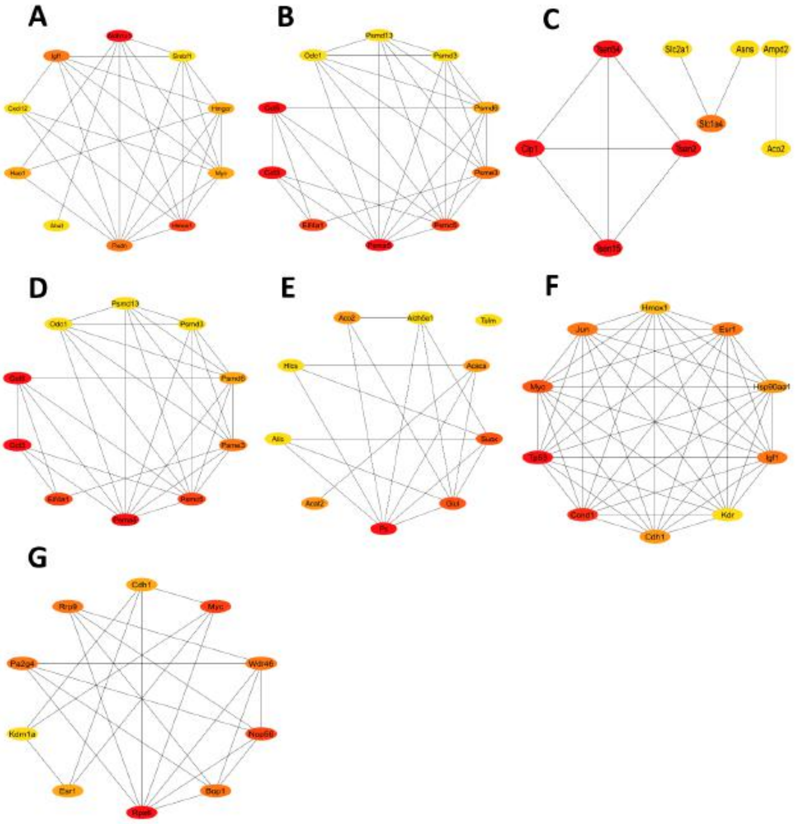


Figure 6. PPI network of 10 HToxBGs using Cytoscape of different significant diseases enriched by the ToxBGs. The enrichment analysis was done using the Enrichr online bioinformatics database tool. In the figure, A) Liver Cirrhosis, B) HIV Coinfection, C) Progressive microcephaly, D) Hereditary Diffuse Gastric Cancer, E) Generalized hypotonia, F) Malignant neoplasm of liver, and G) Disease Exacerbation.

Final stage ToxBG declaration

In the second-stage, we identified 2877 common ToxBGs using statistical approaches, namely the two-sample t-test, WSR-test, and LIMMA. The identified ToxBGs IDs were then converted to the official gene name. We functionally annotated these ToxBGs using the DAVID online database platform and Enrichr online database to identify which ToxBGs were significantly enriched in the KEGG pathways and diseases. PPI network analysis of the significantly enriched ToxBGs to the pathways and diseases was done using the string database, and 10 HToxBGs for each of the significant pathways and diseases were discovered using Cytoscape. A total of 90 HToxBGs for the nine significantly enriched pathways and 70 of HToxBGs for the seven significantly enriched diseases were discovered at the final stage of ToxBGs identification. Finally, we declared 90 HToxBGs for pathway enrichment analysis and 70 of HToxBGs for disease enrichment analysis, totaling 160 ToxBGs as the final stage ToxBGs for predicting chemicals/drugs toxicity.

Discussion

Pharmaceutical, pesticide, and environmental chemical researchers are very interested in the early prediction of chemical/drug adverse effects because toxicity is one of the main causes of drug attrition. The study of chemical toxicity requires an understanding of the regulatory pathways and cell signaling that a drug candidate affects. (Barel and Herwig, 2018; Füzi *et al*., 2021). The identified ToxBGs or HToxBGs enriched in the chemical-treated perturbed pathway rno03008: Ribosome biogenesis in eukaryotes is essential to the molecular life of all cells. The synthesis of ribosomes is also one of the most energy-intensive cellular functions. The strictly controlled process of ribosome biogenesis is closely related to other essential biological functions, such as cell division and growth (Thomson *et al*., 2013). A metabolic pathway in biochemistry is a connected set of chemical events that take place inside a cell. Metabolites are the reactants, products, and intermediates of an enzymatic reaction that are altered by a series of chemical reactions that are catalyzed by enzymes (Boyle, 2005). In addition to being necessary for energy consumption, these metabolic pathways are also necessary for specific effector functions such as phagocytosis, degranulation, chemotaxis, reactive oxygen species (ROS) generation, and neutrophil extracellular traps (Stojkov *et al*., 2022). Our identified ToxBGs were enriched in the ToxBGs metabolic pathway. The identified ToxBGs were also enriched in another important pathway, rno00480: Glutathione metabolism. The most prevalent low molecular weight thiol is glutathione (also known as gamma-glutamyl-cysteinyl-glycine, or GSH), and the main redox pair in mammalian cells is GSH/glutathione disulfide. Protein glutathionylation, signal transduction, cytokine production and immunological response, DNA and protein synthesis, gene expression, cell proliferation and apoptosis, and antioxidant defense are all regulated by glutathione. Glutathione deficiency contributes to oxidative stress, which is a major factor in aging and the etiology of numerous illnesses, such as kwashiorkor, seizures, Alzheimer's disease, Parkinson's disease, liver disease, cystic fibrosis, sickle cell anemia, HIV, AIDS, cancer, heart attacks, strokes, and diabetes (Wu *et al*., 2004). The rest of the pathways significantly enriched by the ToxBGs are also important for the rat and human stabilizing biological conditions, and up- or downregulation of the ToxBGs in the respective pathways creates diseases and other health hazards. On the other hand, the liver, a part of the gastrointestinal tract, is one of the most important organs in the human body that performs over 500 functions to promote physiological homeostasis (Faccioli *et al*., 2022). Conversely, the identified ToxBGs or HToxBGs were enriched in chemical-induced diseases. Among these, severe liver scarring (fibrosis), loss of organ function, and dire consequences associated with portal hypertension (high blood pressure in the hepatic portal vein and its branches) are characteristics of cirrhosis (Fallowfield *et al.*, 2021). Thus, chronic hepatitis, liver cirrhosis, and hepatocellular cancer are frequently caused by hepatitis B virus (HBV) infection. Ten percent of individuals with HIV also have chronic co-infection with HBV due to common mechanisms of transmission. Comparing HIV/HBV coinfection to chronic HBV mono-infection, the former hastens the development of cirrhosis, end-stage liver disease, or hepatocellular carcinoma (Cheng *et al*., 2021). An inactivating mutation in the E-cadherin gene (CDH1) on chromosome 16 is the most common cause of hereditary diffuse gastric cancer (HDGC), an inherited genetic disease (Stewart and Wild 2014). A person's risk of stomach cancer is greatly increased if they inherit an inactive copy of the CDH1 gene. To prevent this cancer, people with these mutations frequently choose to have a preventive gastrectomy, which involves removing the stomach entirely (Stewart and Wild 2014). Mutations in CDH1 are also associated with a high risk of lobular breast cancers, and may be associated with a mildly elevated risk of colon cancer (Van der Post *et al*., 2015). The identified ToxBGs or HToxBGs were enriched in other chemically induced diseases also (Table 3). On the other hand, hub genes are those that interact with numerous other genes in the gene network and are frequently essential for biological processes and gene regulation. In addition, hub genes were described as the most closely associated with disease. Therefore, the proposed sequence of computational and bioinformatics approaches can be applied to identify and evaluate the HToxBGs, or the final stage of ToxBGs, for predicting the potential toxicity of chemicals or drugs.

Conclusion

Finally, we can conclude that the differential expression of ToxBGs may perturb the respective pathway, which causes diseases. ToxBGs that are directly enriched in diseases—their differential expression is responsible for those diseases. On the other hand, HToxBGs also play the key role in regulating their neighboring genes that regulate the disease state. Thus, the suggested sequence of computational and bioinformatics techniques can be used to detect and assess HtoxBGs and to forecast the possible toxicity of chemicals or medications.

Acknowledgement

This research was financially supported by the Ministry of Science & Technology, Government of the People's Republic of Bangladesh, grant number: SRG-224553.

Conflict of Interest

The authors affirm that they have no conflict of interest to disclose.

Author Contribution

Conceptualization, Mohammad Nazmol Hasan; methodology, Mohammad Nazmol Hasan; software, Mohammad Nazmol Hasan; validation, Mohammad Nazmol Hasan, Mohammad Shah Alam, and Md Mamunur Rahman; data curation, Mohammad Nazmol Hasan, Mohammad Shah Alam, and Md Mamunur Rahman; writing—preparation of the initial draft, Mohammad Nazmol Hasan; writing, review and editing, Mohammad Nazmol Hasan, Mohammad Shah Alam, and Md Mamunur Rahman; visualization, Mohammad Nazmol Hasan, Mohammad Shah Alam, and Md Mamunur Rahman; supervision, Mohammad Nazmol Hasan; project administration, Mohammad Nazmol Hasan; revenue acquisition, Mohammad Nazmol Hasan. All authors have reviewed the manuscript in its current form and given their approval.

References

Afshari, C. A., H. K. Hamadeh and P. R. Bushel. 2011. The evolution of bioinformatics in toxicology: Advancing toxicogenomics. *Toxicol. Sci*. 120.

Engwa G. A., P. U. F. Okeke, N. F. Nwalo and M. Unachukwu. 2019. *Mechanism and Health Effects of Heavy Metal Toxicity in Humans*. In: Poisoning in the Modern World - New Tricks for an Old Dog?

Badsha, M. B., M. N. Hasan and M. N. H. Mollah. 2020. rhcoclust: Robust Hierarchical Co-Clustering to Identify Significant Co-Cluster. *CRAN Contrib*. Packag.

Balali-Mood, M., K. Naseri, Z. Tahergorabi, Z., M. R. Khazdair and M. Sadeghi. 2021. Toxic Mechanisms of Five Heavy Metals: Mercury, Lead, Chromium, Cadmium, and Arsenic. *Front. Pharmacol*. 12.

Barel, G and R. Herwig. 2018. Network and pathway analysis of toxicogenomics data. *Front Genet* 9:. https://doi.org/10.3389/fgene.2018.00484.

Boateng, E. Y. and D. A. Abaye. 2019. A Review of the Logistic Regression Model with Emphasis on Medical Research. *J Data Anal Inf Process* 07:https://doi.org/10.4236/jdaip.2019.74012.

Boyle. J. 2005. *Lehninger principles of biochemistry* (4th ed.): Nelson, D., and Cox, M. Biochem Mol Biol Educ 33:. https://doi.org/10.1002/bmb.2005.494033010419.

Braun, P. and A. C. Gingras. 2012. History of protein-protein interactions: From egg-white to complex networks. *Proteomics* 12.

Carbon, S., E. Douglass, B. M. Good et al. 2021. The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res* 49:. https://doi.org/10.1093/nar/gkaa1113.

Chen, E. Y., C. M. Tan, Y. Kou et al. 2013. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14:. https://doi.org/10.1186/1471-2105-14-128.

Cheng, Z., P. Lin and N. Cheng. 2021. HBV/HIV Coinfection: Impact on the Development and Clinical Treatment of Liver Diseases. *Front. Med*. 8.

Dong, X., Z. Yu, W. Cao et al. 2020. A survey on ensemble learning. *Front. Comput. Sci*. 14

Faccioli LAP, Dias ML, Paranhos BA, dos Santos Goldenberg RC (2022) Liver cirrhosis: An overview of experimental models in rodents. *Life Sci*. 301.

Fallowfield, J. A., M. Jimenez-Ramos and A. Robertson. 2021. Emerging synthetic drugs for the treatment of liver cirrhosis. *Expert Opin. Emerg. Drugs* 26.

Mello F. R. and P. M. Antonelli. 2018. *Introduction to Support Vector Machines*. In: Machine Learning.

Füzi, B., J. Gurinova, H. Hermjakob et al. 2021. Path4Drug: Data Science Workflow for Identification of Tissue-Specific Biological Pathways Modulated by Toxic Drugs. *Front Pharmacol* 12:. https://doi.org/10.3389/fphar.2021.708296.

Hamadeh, H. K., R. P. Amin, R. S. Paules and C. A. Afshari. 2002. An overview of toxicogenomics. *Curr Issues Mol Biol* 4:. https://doi.org/10.21775/cimb.004.045.

Hasan, M. N., Badsha, B. and M. N. H. Mollah. 2025. Robust hierarchical co-clustering for exploring toxicogenomic biomarkers and their chemical regulators. *Sci Rep* 1–14. https://doi.org/doi: 10.1038/s41598-025-99568-7.

Hasan, M. N., Begum, A. A., M. Rahman and M. N. H. Mollah. 2019. Robust identification of significant interactions between toxicogenomic biomarkers and their regulatory chemical compounds using logistic moving range chart. *Comput Biol Chem* 78:. https://doi.org/10.1016/j.compbiolchem.2018.12.020.

Hasan, M. N., M. M. Rana, A. A. Begum et al. 2018. Robust Co-clustering to Discover Toxicogenomic Biomarkers and Their Regulatory Doses of

Chemical Compounds Using Logistic Probabilistic Hidden Variable Model. *Front Genet* 9:. https://doi.org/10.3389/fgene.2018.00516.

Huang, D. W., B. T. Sherman, Q. Tan et al. 2007. The DAVID Gene Functional Classification Tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* 8:. https://doi.org/10.1186/gb-2007-8-9-r183.

Kanehisa, M., M. Furumichi, Y. Sato et al. 2023. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* 51:D587–D592. https://doi.org/10.1093/nar/gkac963.

Kanehisa, M., Y. Sato, M. Kawashima et al. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:. https://doi.org/10.1093/nar/gkv1070.

Krueger, C. 2006. Drug-Induced Diseases: Prevention, Detection, and Management. *J Pharm Technol* 22:. https://doi.org/10.1177/875512250602200413.

Laster, J. and R. Satoskar. 2015. Aspirin-Induced Acute Liver Injury. *ACG Case Reports J* 2:. https://doi.org/10.14309/crj.2014.81.

Lucero, B., P. A. Ceballos, M. T. Muñoz-Quezada et al. 2019. Validity and Reliability of an Assessment Tool for the Screening of Neurotoxic Effects in Agricultural Workers in Chile. *Biomed Res Int* 2019:. https://doi.org/10.1155/2019/7901760.

Miguel, V., S. Lamas and C. Espinosa-Diez. 2020. Role of non-coding-RNAs in response to environmental stressors and consequences on human health. *Redox Biol*. 37.

Abdulazeez, M. A., D. Q. Zeebaree and D. M. Abdulqader. 2020. Machine learning supervised algorithms of gene selection: A review. *ResearchgateNet* 62.

NRC. 2007. National Research Council of the National Academies: Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment. National Academies Press, Washington, DC.

Nyström-Persson, J., Y. Igarashi, M. Ito et al. 2013. Toxygates: Interactive toxicity analysis on a hybrid microarray and linked data platform. *Bioinformatics* 29:. https://doi.org/10.1093/bioinformatics/btt531.

Nyström-Persson, J., Y. Natsume-Kitatani, Y. Igarashi et al. 2017. Interactive Toxicogenomics: Gene set discovery, clustering and analysis in Toxygates. *Sci Rep* 7:. https://doi.org/10.1038/s41598-017-01500-1.

Ritchie, M. E., B. Phipson, D. Wu, et al. 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:. https://doi.org/10.1093/nar/gkv007.

Schölkopf, B. 2003. An Introduction to

Support Vector Machines. In: *Recent Advances and Trends in Nonparametric Statistics.*

Seychell, B. C. and T. Beck. 2021. Molecular basis for protein-protein interactions. *Beilstein J. Org. Chem*. 17.

Stewart, B. and C. Wild. 2014. World Cancer Report 2014. Int Agency Res Cancer 22: https://publications.iarc.who.int/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014.

Stojkov, D., L. Gigon, S. Peng et al. 2022. Physiological and Pathophysiological Roles of Metabolic Pathways for NET Formation and Other Neutrophil Functions. *Front. Immunol*. 13.

Szklarczyk, D., A. L. Gable, D. Lyon et al. 2019. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47:. https://doi.org/10.1093/nar/gky1131.

Thomson, E., S. Ferreira-Cerca and E. Hurt. 2013. Eukaryotic ribosome biogenesis at a glance. *J Cell Sci* 126:. https://doi.org/10.1242/jcs.111948.

Uehara, T., A. Ono, T. Maruyama et al. 2010. The Japanese toxicogenomics project: Application of toxicogenomics. *Mol. Nutr. Food Res*. 54.

Van der Post, R. S., I. P. Vogelaar, F. Carneiro et al. 2015. Hereditary diffuse gastric cancer: Updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. *J. Med. Genet*. 52.

Waters, M. D. and J. M. Fostel. 2004. Toxicogenomics and systems toxicology: Aims and prospects. *Nat. Rev. Genet*. 5.

Wu, G., Y. Z. Fang, S. Yang et al. 2004. Glutathione Metabolism and Its Implications for Health. *J. Nutr*. 134.

Xia, J., M. J. Benner and R. E. W. Hancock. 2014. NetworkAnalyst - Integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res* 42:. https://doi.org/10.1093/nar/gku443.

Ye, J. and T. Wang. 2006. Regularized discriminant analysis for high dimensional, low sample size data. In: Proceedings of the ACM SIGKDD I*nternational Conference on Knowledge Discovery and Data Mining*.

Zhou, Y., B. Zhou, L. Pache et al. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10:. https://doi.org/10.1038/s41467-019-09234-6.