

Published: June 30, 2023

Citation: Sheetal A, Reddy AS, Arora P, et al., 2023. Predictive Modeling of Metabolomics data for the Identification of Biomarkers in Chronic Kidney Disease. Medical Research Archives, [online] 11(6). <https://doi.org/10.18103/mra.v11i6.3894>

Copyright: © 2023 European Society of Medicine. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI:

<https://doi.org/10.18103/mra.v11i6.3894>

ISSN: 2375-1924

RESEARCH ARTICLE

Predictive Modeling of Metabolomics data for the Identification of Biomarkers in Chronic Kidney Disease

Ambulge Sheetal¹, Ambati Sekhar Reddy¹, Pooja Arora^{1*}, Veena Puri², Prasad V. Bharatam³

¹Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), Panjab University, Sector 14, Chandigarh - 160014, India.

²Center for Systems Biology and Bioinformatics, National Institute of Pharmaceutical Education and Research (NIPER), Panjab University, Sector 14, Chandigarh - 160014, India.

³Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Panjab University, Sector 14, Chandigarh - 160014, India.

poojaarora@niper.ac.in

Abstract:

Chronic kidney disease is a specific type of Kidney Disease in which, a gradual loss of Kidney Function over a period of 3-30 months is noticeable. Early detection is imperative to prevent this catastrophic event and initiate treatment that may mitigate renal injury. Metabolomics data in Chronic Kidney Disease carries a lot of information about biomarkers. However, it is not clear which of these biomarkers are significant, biostatistical analysis of metabolomics data might provide the clues. In this work, an attempt has been made to find novel biomarkers that may be responsible for causing Chronic Kidney Disease by employing bioinformatics and advanced computational tools. The Chronic Kidney Disease data of the patients (in stages 3 and 4) was selected and data was segregated based on renal and cardiovascular parameters. The study consisted 441 patients and 293 metabolites. Subsequently the identification of top metabolites (as biomarkers) was carried out using statistical methods like t-test, Principal component analysis and partial least square analysis. Nine biomarkers were identified from these statistical analyses i.e Galacturonic acid, p-cresol, L-serine, L-glutamine, Lactose, 2-O-Glycerol- α -D-galactopyranoside, hexa-TMS, Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester, Pseudo uridine penta-tms and Myo-inositol. The Reactivity of identified metabolites was confirmed by using quantum chemistry calculations in Gaussian software. Heat Map was constructed to find out the variations in concentrations of biomarkers in healthy and CKD patients and the showed the higher concentrations of L-serine, Galacturonic acid, L-glutamine and lower concentrations of Pseudo uridine penta-tms, Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester, 2-O-Glycerol- α -D-galactopyranoside, hexa-TMS, Myo-inositol, p-cresol, Lactose in Death Patients. The biological significance of identified top metabolites has been evaluated by identifying the metabolic pathways in which the metabolites are involved. The metabolites which were found to be toxic are pseudouridine, L-glutamine, and galactouronic acid as per the previous reported literature. The Variations in concentration of these metabolites are responsible for the Death of patient with Chronic Kidney Disease.

Introduction:

Chronic kidney disease is a progressive disease in which there is a gradual loss of [kidney function](#) over a period of many months. The leading cause of kidney failure is diabetes. High blood pressure is the second leading cause of kidney failure. The geriatric people exhibit risk factor for kidney diseases, specifically for CKD. The main causes of chronic kidney disease in children are anatomical/structural abnormalities or inherited conditions, such as polycystic kidney disease. Conditions that cause damage to the kidney filters, the glomeruli, can lead to chronic kidney disease. To help improve the quality of care for people with kidney disease, the National Kidney Foundation (NKF) divided kidney disease into five stages based on [glomerular filtration rate \(GFR\)](#). Among these, stages 3 and 4 carry GFR values in the range of 15-44ml/min.^{1,2}

The metabolome is the total complement of metabolites present in a biological sample under given genetic, nutritional or environmental conditions. Metabolomics is the systematic study of the small molecular metabolites in a cell, tissue, bio fluid, or cell culture media that are the tangible result of cellular processes or responses to an environmental stress. This data driven technology yields many insights into metabolic modelling and help pharmaceutical research, nutrition and toxicology.³

Biomarker discovery and drug safety screens are two important fields which garnered the advantages in metabolomics. In clinical practice, biomarkers are widely used as tools for diagnosis, prognosis, patient stratification,

monitoring treatment efficacy, and post treatment surveillance. The use of biomarkers improves reliability of diagnosis and ensures that patients receive effective and safe treatments. In metabolomics, biomarkers are small molecules (metabolites) that can be used to distinguish two groups of samples, typically a disease and control group using targeted or untargeted approach.⁴

A few studies were reported on the identification of biomarkers from metabolomics data using statistical methods. Silvia *et al.* (2019), reported the metabolomics biomarkers and the risk of overall mortality and end stage renal disease (ESRD) in CKD using Cox regression method and performed multivariable model for each metabolite.⁵ They suggested 16 metabolites as biomarkers, further they suggested segregated results into renal replacement (RRT) and death by using cox proportionality model. Karnovsky *et al.* (2019), carried out a study on differential network enrichment analysis which revealed novel lipid pathways in chronic kidney disease.⁶ Afshinnia *et al.* (2016 and 2018), studied the impaired β -oxidation and altered complex lipid fatty acid partitioning with advancing CKD using lipidomic data using statistical methods. It was estimated that increased abundance of saturated C16–C20 FFAs coupled with impaired β -oxidation of FFAs may be the mechanisms underpinning lipid metabolism changes that typify advancing CKD.^{7,8} Kimura *et al.* (2015), carried out a study on the Identification of biomarkers for development of end-stage kidney disease in CKD by metabolomic profiling using capillary

electrophoresis, liquid chromatography mass spectrometry (LC-MS), Cox proportional hazard models. Higher levels of 16 plasma metabolites were identified as biomarkers.⁹ Xi *et al.* (2014), published a chapter on statistical analysis and modeling of mass spectrometry based metabolomics data, in which the multivariate statistical techniques were used in metabolomics studies, ranging from biomarker selection to model building and validation.¹⁰ Ska *et al.* (2012), carried out a classification study on double-check validation of diagnostic statistics.¹¹

Yang *et al.* (2019), proposed a machine learning model to characterize CKD with metabolomics data, using an elastic-net model (ENM) and identified fifteen metabolomic biomarker and serum biomarkers.¹²

CKD has severe complications in targeted therapy thus there is a need for potential target identification for early detection to ease the process. The data available in metabolomics is complex, so there is a need for exploration of metabolomics using statistical modeling to predict the outcomes.

In this work the identification of potential biomarkers using statistical methods and metabolomics data was done. The details are provided in following sections.

Methodology:

Data collection

This was a case-control study to identify biomarker that can predict the chronic kidney disease at a very early stage. The Data on 454 participants of the ProgreDir Cohort Study,

Sao Paulo, Brazil has been used in study(<https://www.metabolomicsworkbench.org/>). Metabolites reported in this data bank were identified by GC-MS (Agilent MassHunter) and NIST libraries. The inclusion and exclusion criteria adopted for the creation of this dataset were – (i) Dataset was taken after excluding metabolites present in <50% of participants, 293 metabolites were analyzed. (ii) Stage 3 and 4 kidney disease patients were only selected for study. (iii) Data collection was focused on renal and cardiovascular parameters. (iv) All patients aged ≥ 30 years and for whom at least two measurements of creatinine ≥ 1.6 mg/dl for men and ≥ 1.4 mg/dl for women were considered as potential candidates. (v) The exclusion criteria checked were: hospitalization within the last six months, pregnancy, acute myocardial infarction within the last six months, psychiatric diseases, autoimmune diseases, ongoing immunosuppressive therapy or chemotherapy, ongoing RRT, HIV/AIDS infection, glomerulonephritis, hepatitis B or C and organ transplantations. These were the criteria used by the authors during the study i.e. "[Metabolomics biomarkers and the risk of overall mortality and ESRD in CKD: Results from the ProgreDir Cohort](#)" which was extracted from the metabolomic society database.⁵

Data analysis

The data analysis was done using diverse statistical methods. First the data was preprocessed using Z score as the dataset used was having wide range of values. The Z score is standardized score that helps to bring the values on same scale, which makes the

calculation easy. It is mostly used because it allows calculating the probability of a score occurring within a standard normal distribution and enables us to compare two scores that are from different samples (which may have different means and standard deviations).¹³ Once the standardization was done the dataset was divided into two groups (i) the first group includes the patients who have undergone renal replacement as well as who died due to kidney failure (n=124) during the study period and (ii) the second group includes the rest of patients (n=312). Then an unpaired t test was performed on both subsets to identify the significant metabolites. The t-test is used for hypothesis testing to determine whether a process has an effect on both samples or if the groups are different from each other. An unpaired t-test is a statistical procedure that compares the averages/means of two independent or unrelated groups to determine if there is a significant difference between the two. The metabolites having the significance value of less than 0.05 were considered as top metabolites and used for further analysis.¹⁴ After t test the top significant metabolites were taken further and PCA method was applied. PCA is a dimensionality reduction technique, which is a method used for obtaining important variables (in form of components) from a large set of variables available in a data set. The principal components were identified and the metabolites having values greater than 0.01 were considered top metabolites which were further validated and considered as biomarkers of CKD.¹⁵ After PCA the validation

method i.e. PLS was employed to the top metabolites obtained in t test. Partial least squares (PLS) is a technique that reduces the data to a smaller set of uncorrelated components and performs least squares regression on these components, instead of the original data. It addresses the multicollinearity problem by computing latent vectors which explains both the independent variables and the dependent variables. It is used when the goal is to predict more than one dependent variables.¹⁶

Quantum chemical study of the top metabolites from both PCA and PLS method was performed. All the geometry optimizations were carried out using hybrid DFT (Density Functional Theory)¹⁷ employing 6-31+ G(d) basis set and the B3LYP functional. Then the electrophilicity index (ω) of top 9 metabolites was calculated. The electrophilicity index (ω) measures the energy of stabilization when an optimal electronic charge transfer from the environment to the system occurs. In order to find the biological significance of the metabolites and prove the toxicity toxtree software was used.¹⁸

Results

1. Identifying biomarkers using predictive statistical modelling

1.1 Data preprocessing by Z score

The data preprocessing was done by imputing the missing values and Z standardization. The missing values were identified and they were replaced with mean values. The Z standardization was carried on for each metabolite. The values were found to be between -6.4 and 16.7.¹³

1.2. T- test results

The dataset was divided into two groups–(i) the first group consisting of 124 patients who died and/ or had undergone renal replacement, which is taken as control group (ii) the second group consisting of the remaining patients which was taken as test group. In t test the control group and test groups should have equal in number to perform analysis so the test group which were having 312 patients were further divided into 3 subsets of 124 each in which 1st subset has patients 1 to 124, 2nd has patients 125 to 248 (i.e total 124 patients) and as only 64 left out

312 and we cannot perform t test of unequal groups the 3rd subset has patients from 184 to 312. The unpaired t test was carried out between the three subset and the control group. Then a column called frequency was added to table of t-test results which implies how many times a particular metabolites had the p-value less than 0.05 in subsets. Based on the p-value the screening was done in which the 47 metabolites having more than 1 value in frequency column were selected as top metabolites and those were taken for further processing (Table S1).¹⁴

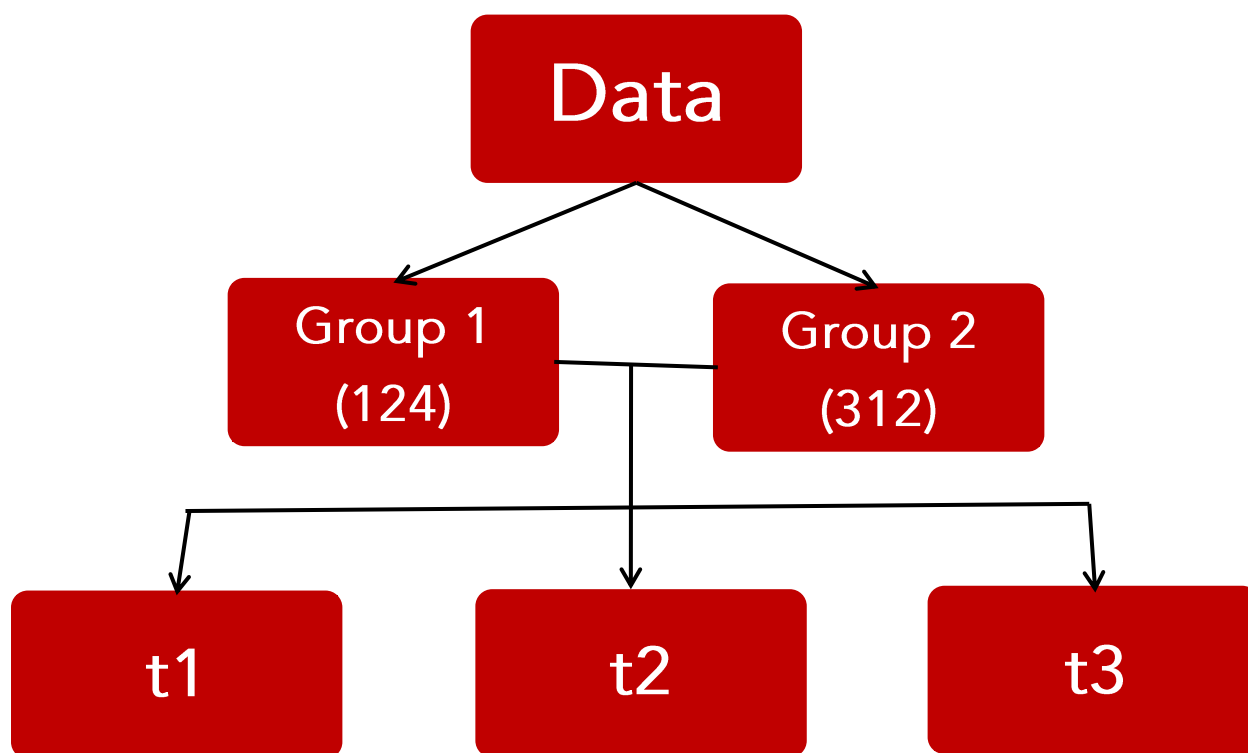


Figure 1: flow chart of groups divided during t-test

1.3. Principal Component Analysis Results

The principal component analysis was performed in statistica software using the top 47 metabolites obtained after t-test. The

results of PCA in statistica gave the significance values of metabolites in the power column which implies the probability of metabolite responsible for causing the

disease and based on the value of power the ranking is given in the variable importance column (Table S2). The top 20 ranking metabolites were considered for further analysis.¹⁵

1.4. Partial least square analysis Results

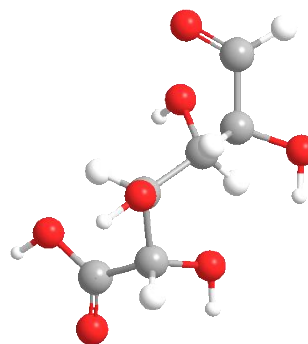
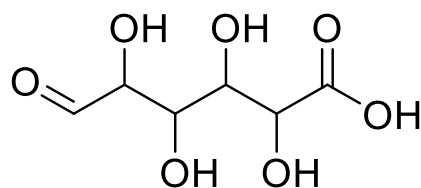
PLS analysis was performed to validate the results from PCA. The PLS was also calculated in statistica software and the variable importance of each metabolites in the dataset is given in VIP column which implies the significant values of metabolites. The ranking was provided according to the VIP column (Table S3). The metabolites with top 20 ranks were selected as significant metabolites for identification of biomarkers.¹⁶

The top metabolites obtained after both PCA and PLS were compared with the literature and the metabolites present in all three i.e. PCA, PLS and literature were considered as the potential biomarkers. They were nine metabolites which were common those are Galacturonic acid, p-cresol, L-serine, L-glutamine, Lactose, 2-O-Glycerol- α -D-galactopyranoside, hexa-TMS, Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester,

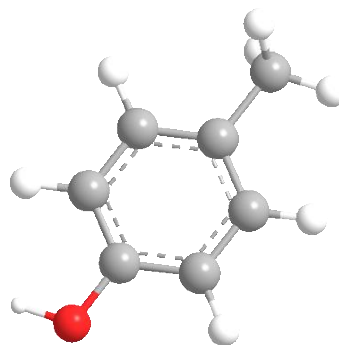
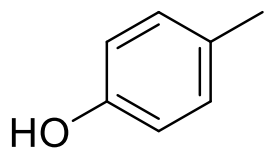
Pseudo uridine penta-TMS and Myo-inositol and these were further explored to know their toxic effects in kidney disease.

2. Electronic structure analysis of identified biomarkers

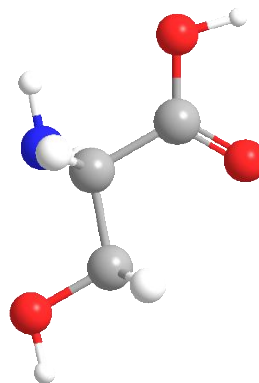
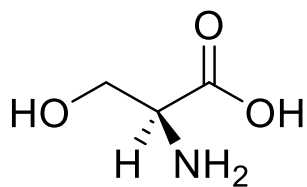
The 3D structures of the selected nine metabolites were obtained using quantum chemical geometry optimization using the DFT method B3LYP/6-31+G(d). The Gaussian software was used for quantum chemical calculations and the electrophilicity index values of the identified top metabolites were estimated (Table 1). Figure 2 shows the 2D and 3D structures. The electrophilicity values of these biomarkers are very low (< 3), hence the electrophilicity is not responsible for the observed CKD.¹⁷



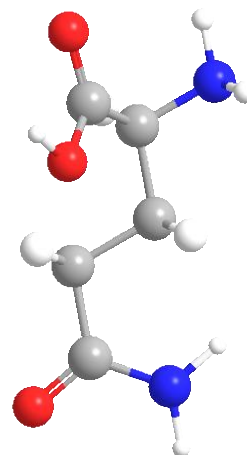
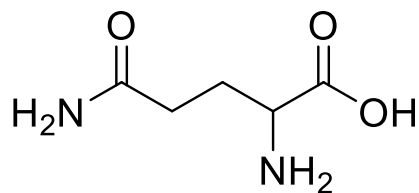
2D and 3D structure of galacturonic acid



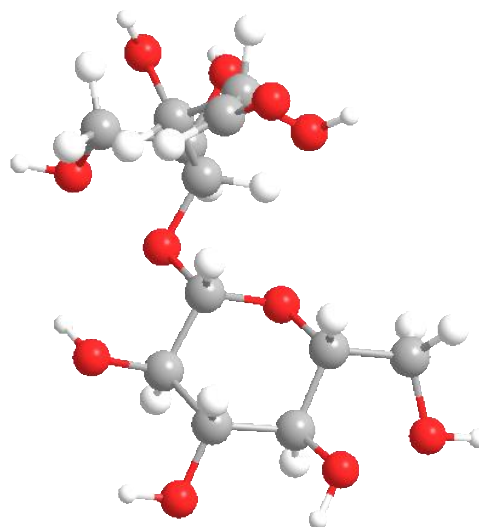
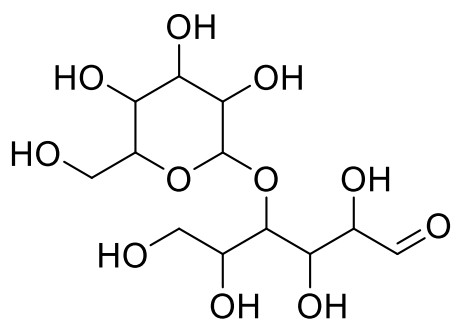
2D and 3D structure of p-cresol



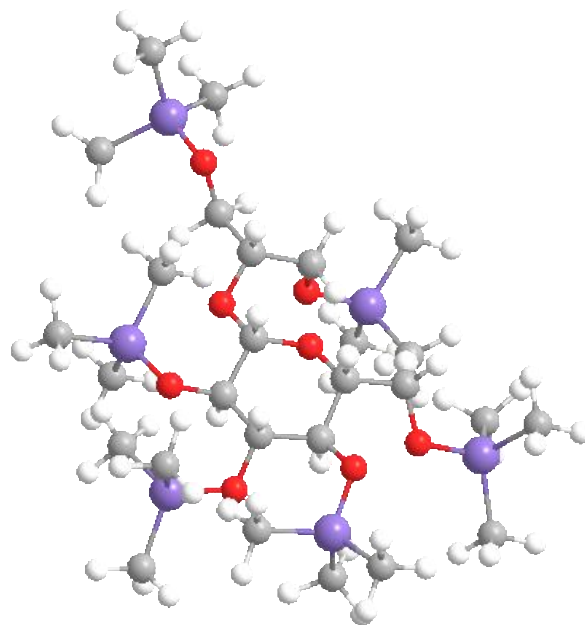
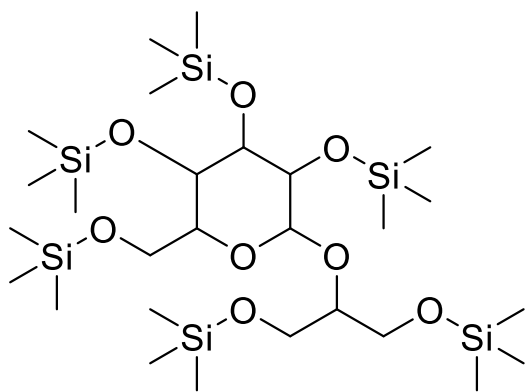
2D and 3D structure of L-serine



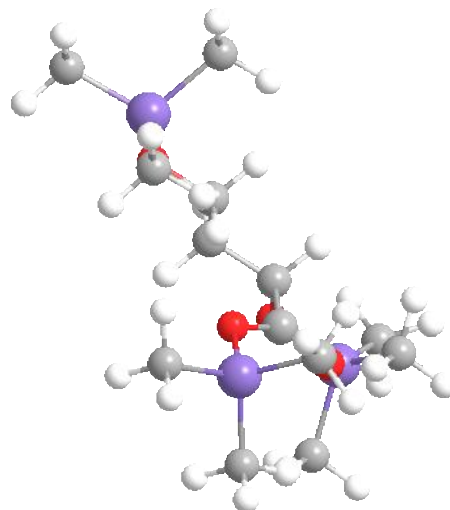
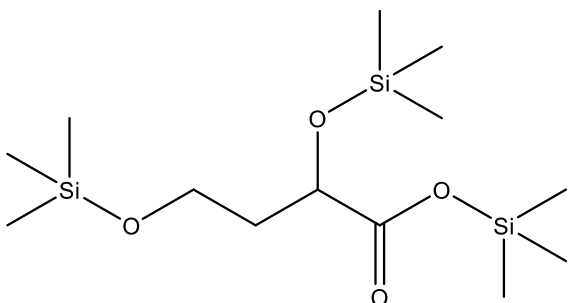
2D and 3D structure of L-glutamine



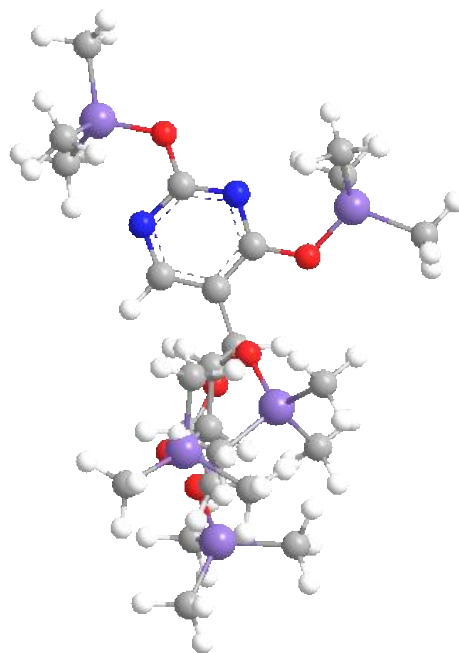
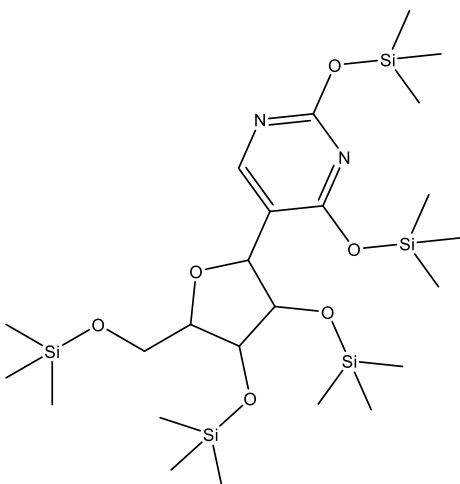
2D and 3D structure of Lactose



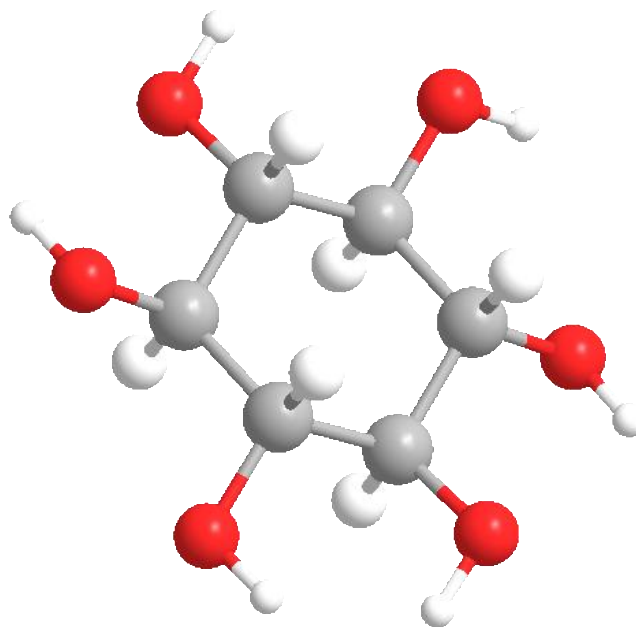
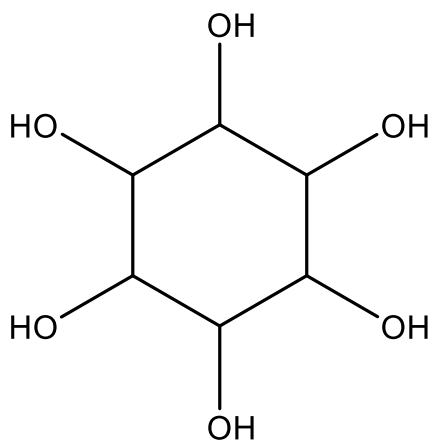
2D and 3D structure of 2-O-Glycerol-.alpha.-d-galactopyranoside, hexa-TMS



2D and 3D structure of Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester



2D and 3D structure of Pseudo uridine penta-tms



2D and 3D structure of Myo-inositol

Figure 2: 2D and 3D structures of top 9 metabolites used for quantum calculations

Metabolites name	Electrophilicity index
Galacturonic acid	1.7439
p-cresol	0.9440
L-serine	1.0742
L-glutamine	1.0704
Lactose	1.0368
2-O-Glycerol- α -D-galactopyranoside, hexa-TMS	1.0243
Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester	1.0536
Pseudo uridine penta-tms	0.9624
Myo-inositol	1.1757

Table 1: Electrophilicity index of top metabolites.

3. Toxtree

The toxtree software classifies the metabolites as toxic and non toxic based on the decision tree approach. The decision is made by the software based on the information available in

literature and classifies the metabolites into toxic or non toxic. The toxicity of top 9 metabolites according to toxtree is shown in Table 2.¹⁸

Metabolites name	Toxicity
Galacturonic acid	highly toxic
p-cresol	less toxic
L-serine	less toxic
L-glutamine	highly toxic
Lactose	less toxic
2-O-Glycerol-.alpha.-d-galactopyranoside, hexa-TMS	highly toxic
Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester	highly toxic
Pseudo uridine penta-tms	highly toxic
Myo-inositol	less toxic

Table 2: Toxicity of top metabolites according to toxtree software.

4. Heat map:

A heat map is a two-dimensional graphical representation of data where the data values are mapped to colors across a range. Each colored cell on the map corresponds to a concentration value in your data table, with samples in rows and features/compounds in columns. Heatmap is used to identify features that are unusually high/low using stronger intensities of one color to represent lower levels of the variable, and increasing intensities of a different color to represent higher levels.¹⁹

Heat map was constructed to Validate the results obtained from Statistical Methods as Molecular Docking and Molecular dynamics (MD) simulation are not possible for the metabolites.

Heat Map of CKD Biomarkers was constructed using MetaboAnalyst 5.0 Tool (Figure 3)

Up regulated and down regulated metabolites can be seen clearly in the Table 3.

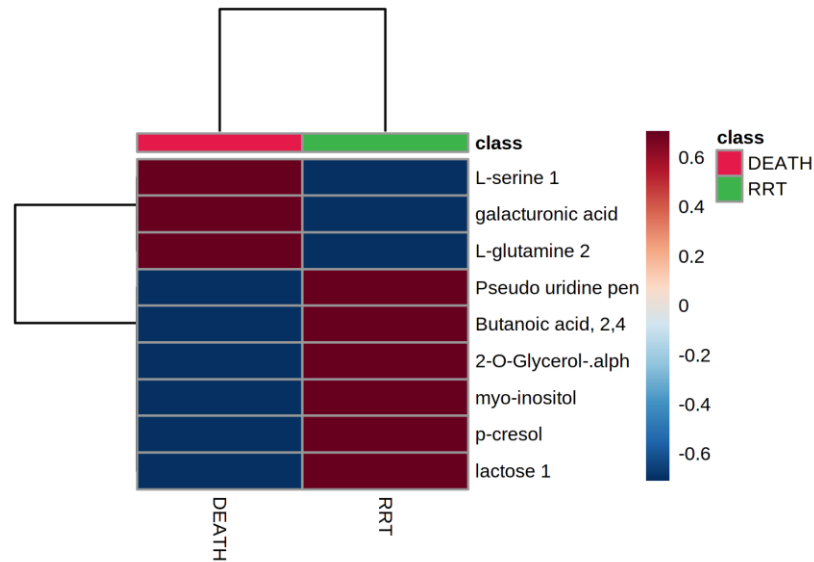


Figure 3: Heat map

Upregulated Metabolites	Downregulated Metabolites
L-serine	Pseudo uridine penta-tms
Galacturonic acid	Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester
L-glutamine	2-O-Glycerol-.alpha.-d-galactopyranoside, hexa-TMS
	Myo-inositol
	p-cresol
	Lactose

Table 3: Up regulated and Down regulated metabolites

The concentration of metabolites like L-serine, Galacturonic acid and L-glutamine are high in Death patients. Similarly for like Pseudo uridine penta-tms, Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester, 2-O-Glycerol-.alpha.-d-galactopyranoside, hexa-TMS, Myo-inositol, p-cresol and Lactose the concentration is low in Death Patients. The Variations in concentration of these metabolites are responsible for the Death of patient with CKD.

5. The metabolic pathways in which identified metabolites are involved

5.1. L-glutamine

Liua et al.(2017) suggested that purine metabolism disturbance is one of the factor responsible for chronic kidney disease. L-glutamine is one of the products in purine metabolism which may be the reason for purine metabolism disturbance and further a article published in medical hypothesis which

is entitled as “Disturbed purine nucleotide metabolism in chronic kidney disease is a risk factor for cognitive impairment” showed that

disturbances in the pathway may be the reason for many diseases.¹⁹

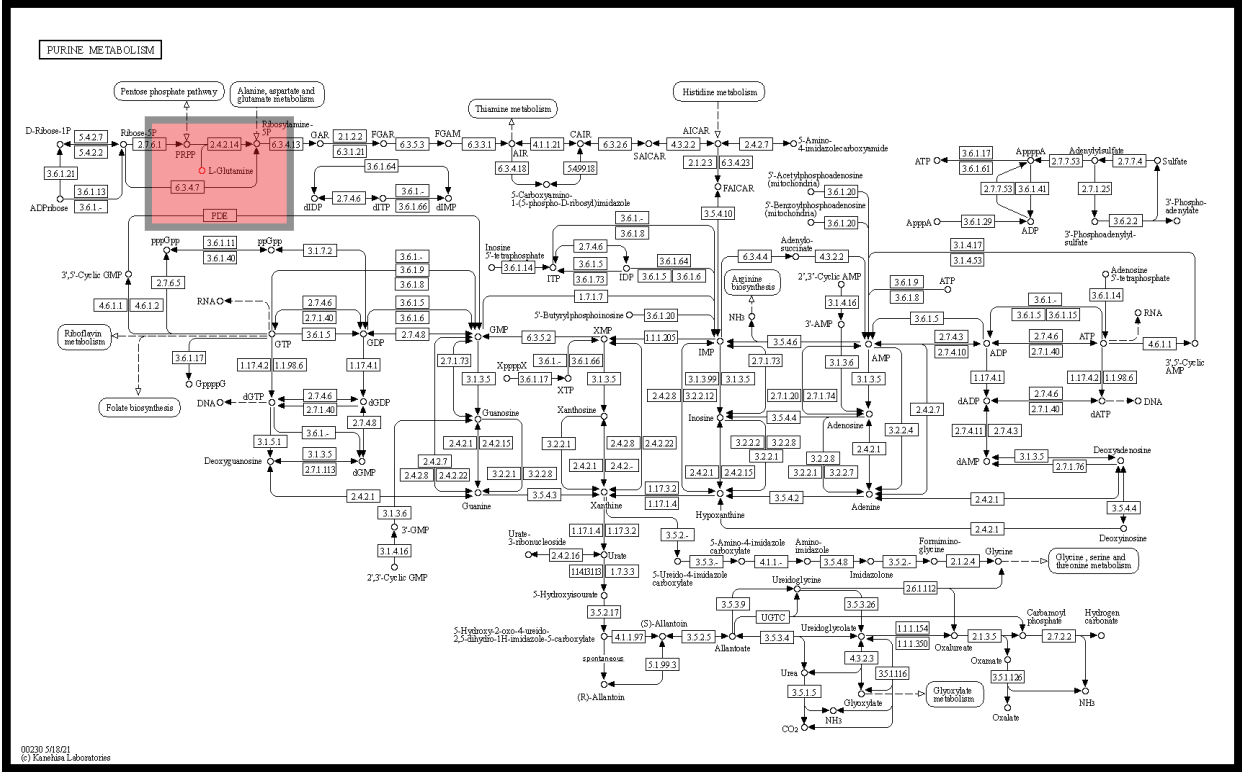


Figure 3: Purine metabolic pathway.

5.2. Galactouronic acid

Yu et al.(2020) suggested that over expression of amino acids and changes in interconversion of pentose and glucuronate pathways along

with other metabolic pathways lead to many diseases like membranous nephropathy, diabetes nephropathy and kidney disease²⁰

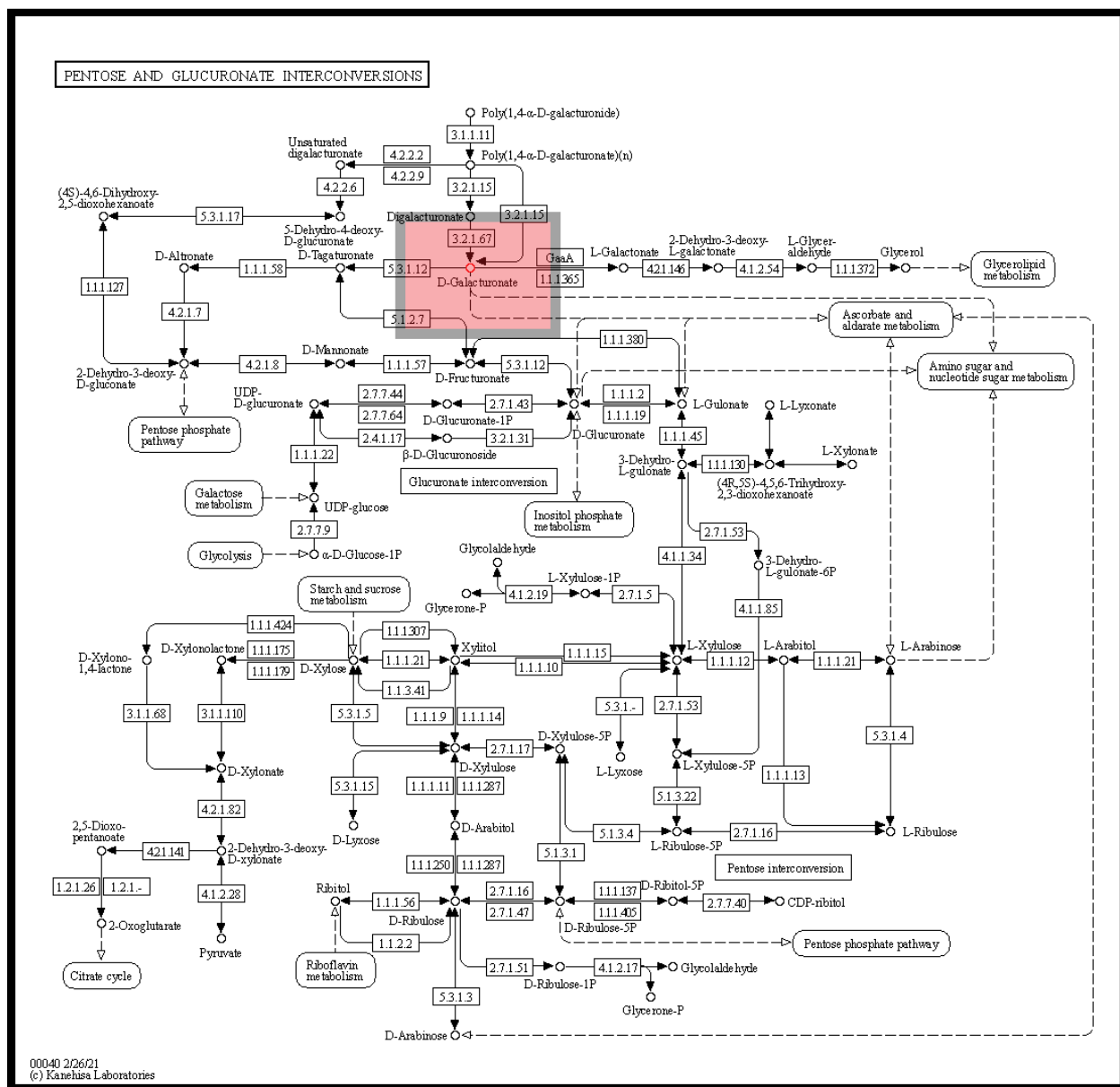


Figure 4: Pentose and glucuronate interconversion pathway.

5.3. Pseudouridine

Mazumder et al., identified the metabolites of kidney patients and suggested that the purine and pyrimidine metabolism alterations lead to

reduced kidney function causing CKD in the patients. Especially the pseudouridine present in the pyrimidine metabolism may be the cause for the failure of kidney function.²¹

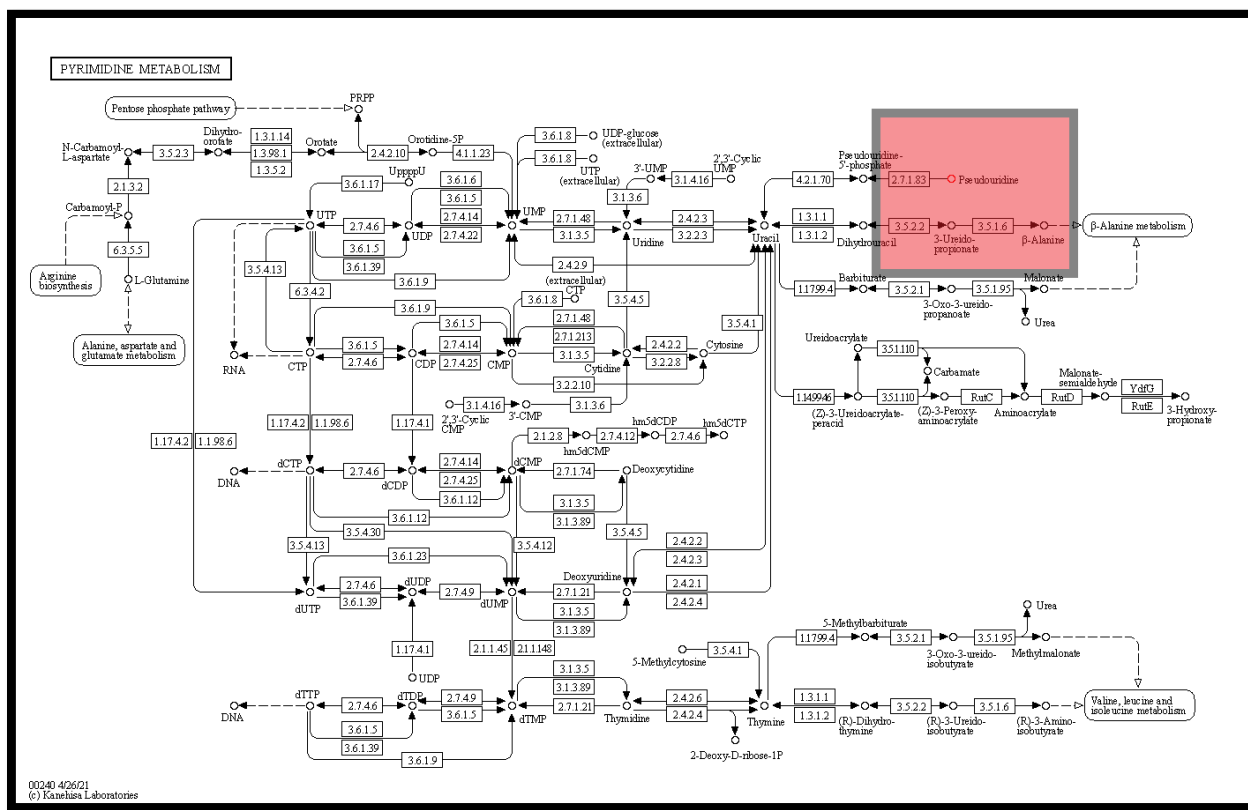


Figure 5: Pyrimidine metabolic pathway.

These above metabolic pathways were linked to kidney disease in the literature and suggested that the alterations of this caused kidney disease and the current work suggests that identified metabolites may be the cause of alterations in the pathways. There are a few

other metabolites like 2-O-Glycerol- α -D-galactopyranoside, hexa-TMS and butanoic acid, 2,4-bis(trimethylsilyloxy)-, trimethylsilyl ester which were also found to be biomarkers according to the statistical analysis but there is no information related to that in literature.



Figure 6: Flow of toxicity

DISCUSSION

There are no targets to treat kidney diseases with drugs at the global level. This makes the identification of biomarkers an important endeavor. In this study metabolomic data was statistically explored for the identification of biomarkers in CKD.

The raw dataset obtained from Metabolomics Society was preprocessed using R software, and further statistical methods such as t-tests, PCAs, and PLS were applied to identify potential biomarkers. The top ranking metabolites obtained in both PCA and PLS methods were compared with the literature and nine metabolites were found to be in common i.e Galacturonic acid, p-cresol, L-serine, L-glutamine, Lactose, 2-O-Glycerol-.alpha.-d-galactopyranoside, hexa-TMS, Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester, Pseudo uridine penta-tms and Myo-inositol. These metabolites were further explored to understand their reactivity in the body. The quantum chemical studies were performed to evaluate their structural features. Then a toxicity tool called Toxtree was used to identify toxic metabolites which showed 5 metabolites in the toxic class i.e Galacturonic acid, L-glutamine, 2-O-Glycerol-.alpha.-d-galactopyranoside, hexa-TMS, Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester and Pseudo uridine penta-tms. Heat Map is constructed to validate the results. The concentration of metabolites like L-serine, Galacturonic acid and L-glutamine are high in Death patients. Similarly for like Pseudo uridine penta-tms, Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester, 2-O-Glycerol-.alpha.-d-galactopyranoside,

hexa-TMS, Myo-inositol, p-cresol and Lactose the concentration is low in Death Patients. The Variations in concentration of these metabolites are responsible for the Death of patient with CKD. The toxic metabolites were further explored for their metabolic pathways and connected to kidney toxicity using the literature. It can be predicted that the disturbance in the identified pathways due to these nine metabolites may be the cause of observed CKD.

CONCLUSION

Our study results suggests that Galacturonic acid, p-cresol, L-serine, L-glutamine, Lactose, 2-O-Glycerol-.alpha.-d-galactopyranoside, hexa-TMS, Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester, Pseudo uridine penta-tms and Myo-inositol are the top metabolites responsible for chronic kidney disease. These findings are consistent with the literature.

Heat Map showed conclude that higher concentrations of L-serine, Galacturonic acid, L-glutamine and lower concentrations of Pseudo uridine penta-tms, Butanoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester, 2-O-Glycerol-.alpha.-d-galactopyranoside, hexa-TMS, Myo-inositol, p-cresol, Lactose in Urine of Death Patients. These Variations in concentration of these metabolites are responsible for the Death of patient with Chronic Kidney Disease.

Corresponding author:

Pooja Arora
Department of Pharmacoinformatics,
National Institute of Pharmaceutical
Education and Research (NIPER),
Sector - 67, S. A. S. Nagar (Mohali) – 160062,
Panjab University, Sector 14,
Chandigarh - 160014, India.
Email: poojaarora@niper.ac.in

Acknowledgement / Funding:

We are grateful for DBT
(BT/PR40164/BTIS/137/17/2021) for funding
this project.

Conflicts of Interest:

None

References:

1. Levey AS, Eckardt KU, Tsukamoto Y, et al. Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int.* 2005;67(6):2089-2100.
2. Rettig RA, Norris K, Nissenson AR. Chronic kidney disease in the United States: a public policy imperative. *Clin J Am Soc Nephrol.* 2008;3(6):1902-1910.
3. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol.* 2016;17(7):451-459.
4. Mayeux R. Biomarkers: potential uses and limitations. *NeuroRx.* 2004;1(2):182-188.
5. Titan SM, Venturini G, Padilha K, et al. Metabolites related to eGFR: Evaluation of candidate molecules for GFR estimation using untargeted metabolomics. *Clin Chim Acta.* 2019;489:242-248.
6. Ma J, Karnovsky A, Afshinnia F, et al. Differential network enrichment analysis reveals novel lipid pathways in chronic kidney disease. *Bioinformatics.* 2019;35(18):3441-3452.
7. Afshinnia F, Rajendiran TM, Soni T, et al. Impaired β -Oxidation and Altered Complex Lipid Fatty Acid Partitioning with Advancing CKD. *J Am Soc Nephrol.* 2018;29(1):295-306.
8. Afshinnia F, Rajendiran TM, Karnovsky A, et al. Lipidomic Signature of Progression of Chronic Kidney Disease in the Chronic Renal Insufficiency Cohort [published correction appears in *Kidney Int Rep.* 2017 Sep 18; 2(6):1265]. *Kidney Int Rep.* 2016;1(4):256-268.
9. Tsuruya K, Yoshida H, Nagata M, et al. Association of the triglycerides to high-density lipoprotein cholesterol ratio with the risk of chronic kidney disease: analysis in a large Japanese population. *Atherosclerosis.* 2014;233(1):260-267.
10. Xi B, Gu H, Baniasadi H, Raftery D. Statistical analysis and modeling of mass spectrometry-based metabolomics data. *Methods Mol Biol.* 2014;1198:333-353.
11. Szymańska E, Saccenti E, Smilde AK, Westerhuis JA. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics.* 2012; 8(Suppl 1):3-16.
12. Almasoud, M, Ward, T.E. Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors. *International Journal of Advanced Computer Science and Applications.* 2019;10(8)
13. Cheadle C, Vawter MP, Freed WJ, Becker KG. Analysis of microarray data using Z score transformation. *J Mol Diagn.* 2003;5(2):73-81.
14. Baldi P, Long AD. A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. *Bioinformatics.* 2001;17(6):509-519.
15. Bartel J, Krumsiek J, Theis FJ. Statistical methods for the analysis of high-throughput metabolomics data. *Comput Struct Biotechnol J.* 2013;4:e201301009. Published 2013 Mar 22.
16. Szymańska E, Saccenti E, Smilde AK, Westerhuis JA. Double-check: validation of diagnostic statistics for PLS-DA models in

- metabolomics studies. *Metabolomics*. 2012; 8(Suppl 1):3-16.
17. Padmanabhan J, Parthasarathi R, Subramanian V, Chattaraj PK. Electrophilicity-based charge transfer descriptor. *J Phys Chem A*. 2007;111(7):1358-1361.
18. Patlewicz G, Jeliaskova N, Safford RJ, Worth AP, Aleksiev B. An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR QSAR Environ Res*. 2008;19(5-6):495-524.
19. Xia J, Psychogios N, Young N, Wishart DS. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res*. 2009;37(Web Server issue):W652-W660.
20. Dean ED, Li M, Prasad N, et al. Interrupted Glucagon Signaling Reveals Hepatic α Cell Axis and Role for L-Glutamine in α Cell Proliferation. *Cell Metab*. 2017;25(6):1362-1373.e5.
21. Daniel E, Anitha, J. Gnanaraj J. Optimum laplacian wavelet mask based medical image using hybrid cuckoo search – grey wolf optimization algorithm. *Knowledge-Based Systems*. 2017; 131:58-69.
22. Ma J, Liang P, Yu W, Chen C, Guo X, Wu J, Jiang J. Infrared and visible image fusion via detail preserving adversarial learning. *Inf. Fusion*. 2020; 54: 85-98.
23. Zhang G, Darshi M, Sharma K. The Warburg Effect in Diabetic Kidney Disease. *Semin Nephrol*. 2018;38(2):111-120.