

# Predictive Modeling of Metabolomics data to Identify Potential Biomarkers in Renal Cell Carcinoma

Pooja Arora <sup>1\*</sup>, Ambati Sekhar Reddy <sup>1</sup>, Veena Puri <sup>2</sup>, Prasad V. Bharatam<sup>3</sup>

<sup>1</sup> Department of Pharmacoinformatic, National Institute of Pharmaceutical Education and Research (NIPER), Sector-67, S. A. S. Nagar (Mohali)

<sup>2</sup> Center for Systems Biology and Bioinformatics, Panjab University, Sector 14, Chandigarh

<sup>3</sup> Department of Medicinal Chemistry, National Institute of Pharmaceutical Education and Research (NIPER), Sector-67, S. A. S. Nagar (Mohali),

\*Corresponding Author: Pooja Arora, 70022, Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), Panjab University, Sector 14, Chandigarh - 160014, India.

Received date: July 27, 2023; Accepted date: August 14, 2023; Published date: August 17, 2023

**Citation:** Pooja Arora, Ambati Sekhar Reddy, Veena Puri, Prasad V. Bharatam, (2023), Predictive Modeling of Metabolomics data to Identify Potential Biomarkers in Renal Cell Carcinoma, *Clinical Endocrinology and Metabolism*, 2(4) DOI:10.31579/2834-8761/27

**Copyright:** © 2023, Pooja Arora. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

Renal cell carcinoma (RCC) is a rare human cancer whose prevalence is rapidly rising, an early recognition is essential for the effective disease treatment. A statistical approach for biomarkers identification using metabolomics data has been attempted in this work. The metabolomics data for RCC is extracted from metabolomics workbench database with study ID ST001706. The study consisted 50 metabolites and 256 patients; in that 174 are normal individuals and 82 are RCC patients. From 50 Metabolites, top metabolites (as biomarkers) are identified using advanced statistical techniques such as the t-test, principal component analysis (PCA) and partial least square analysis (PLS). These statistical results identified eight biomarkers i.e. Trigonelline, Hippuric acid, 4-hydroxyhippuric acid, 4-amino hippuric acid, Mannitol, Pyruvic acid, Scyllo-Inositol and Deoxycholic acid. The Gaussian software was used to obtain the 3D Structures of the metabolites and to calculate their electronic parameters. Relative quantification of these biomarkers was done using Heatmap. ROC Curve Analysis has been performed to characterize biomarkers early in RCC. The biological significance of identified top metabolites has been evaluated by identifying the metabolic pathways in which the metabolites are involved.

**Keywords:** renal cell carcinoma, biomarkers, metabolomics

## Introduction

Metabolomics is a relatively new approach of evaluating the composition of a biofluid (plasma or urine) or tissue in which small molecule metabolites are studied. This collection of small molecules, when analyzed and interpreted, can provide a distinct signature that can be used for diagnosis as well as to determine gross metabolic differences between a normal and a diseased state.<sup>1</sup> Renal cell carcinoma (RCC) is another name for hypernephroma, renal cancer, or kidney cancer. It is the third most common urological oncology, accounting for 2-3% of all malignancies.<sup>2</sup> RCC is a fast-growing cancer that frequently spreads to other organs such as the lungs.<sup>2</sup> Among urogenital cancers, RCC has the highest mortality rate, and its prevalence has steadily increased. RCC is curable with surgery if detected in time, although a minority is at risk of recurrence.<sup>3</sup> Symptoms of RCC are usually absent in the early stages. As the disease progresses, the patient may develop symptoms such as a lump on the back, hematuria, lower back pain, unexplained weight loss, fatigue, anemia, and hypertension. Some risk factors for the disease include a family history of RCC, dialysis treatment, hypertension, obesity and cigarette smoking. RCC is treated with surgery, radiation therapy, chemotherapy, immunotherapy, and targeted therapy.<sup>4</sup> In metabolomic studies data-driven technology provides numerous

insights into metabolic modelling and tends to help with pharmaceutical research, nutrition, and toxicity.<sup>5</sup> The metabolomics approach in addition to biomarker discovery can identify new druggable targets because understanding the metabolic disorder and altered biochemical pathways that occur with disease progression can provide insight into possible new treatments for that disease by identifying inhibitors of altered pathways among new and already existing drugs. As a result, metabolomics lends itself to a two-pronged approach to the clinical problem, addressing both disease symptoms and providing novel treatment methods.<sup>6</sup> Biaufray et al. (2021), reported Machine Learning-Enabled Renal Cell Carcinoma Status Prediction Using Multiplatform Urine-Based Metabolomics: The study cohort consisted of 105 RCC patients and 179 controls separated into two sub cohorts: the model cohort and the test cohort. The model cohort was used to choose discriminating features using univariate, wrapper, and embedded techniques.<sup>7</sup>

Allegan et al. (2017), Preoperative fasting urine and serum samples were collected from patients with clinical renal masses and metabolomics and multivariate statistical analysis were performed using <sup>1</sup>H NMR and GCMS (gas chromatography-mass spectrometry). RCC had higher levels

of glycolytic and tricarboxylic acid (TCA) cycle intermediates compared to benign masses.<sup>8</sup>

Zheng et al. (2016), used serum metabolome data from 104 participants, including healthy individuals and early-stage RCC patients, to train and validate the SOM model. For the early detection of RCC, a biomarker cluster of seven metabolites (alanine, creatine, choline, isoleucine, lactate, leucine, and valine) was identified. Using a biomarker cluster, the trained SOM model was able to classify 22 test subjects into the appropriate categories.<sup>9</sup>

Bowie Xi et al. (2014), published a chapter on statistical analysis and modelling 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.<sup>10</sup>

Ska et al. (2012) conducted a classification study on diagnostic statistics double-check validation.<sup>11</sup>

Kim et al. (2011), found that quinolinone, 4-hydroxybenzoate, and gentisate are differentially expressed with a false discovery rate of 0.26, and these metabolites are implicated in common amino acid and oxidative metabolism pathways, which is consistent with high tumour protein breakdown and utilisation and the Warburg effect.<sup>12</sup>

Because targeted therapy for RCC has adverse effects, it is critical to identify potential targets for early diagnosis and treatment. Because metabolomics data is complex, there is a need for metabolomics research using statistical modelling to predict meaningful insights.

This research used statistical methods and metabolomics data to find potential biomarkers for the early detection of Renal Cell Carcinoma. The following sections provide more information.

## Methodology:

### Data collection

Olatomiwa, OB.; David, AG.; Machine Learning-Enabled Renal Cell Carcinoma Status Prediction Using Multiplatform Urine-Based Metabolomics, J. Proteome Res. 2021, 20, 7, 3629–3641 - Olatomiwa O. Biafran et.al., Collected Data from liquid chromatography–mass spectrometry (LC–MS) and nuclear magnetic resonance (NMR) and potential metabolomic panels for RCC were discovered using machine learning (ML). The study cohort consisted of 82 RCC patients and 174 controls. The data for this study is available in Metabolomics workbench database (<https://www.metabolomicsworkbench.org/>) with Study IDs: ST001705 and ST001706.<sup>6</sup>

### Data analysis

Various statistical techniques like median normalization, t-test, PCA, PLS were used to analyze the data.

The Median Normalization is based on the assumption that the samples of a data set are separated by a constant. It scales the sample values on a common scale to have same median. By choosing the median instead of the mean, it helps to remove some of the outliers in the data.<sup>13</sup>

The t-test is used to evaluate if a method has an effect on both samples or if the groups are different from each other. An unpaired t-test compares the averages/means of two independent or unrelated groups to see whether there is a statistically significant difference between them.<sup>14</sup>

Principal Component Analysis is a dimensionality reduction method for extracting important variables (in the form of Principal components) from a large number of variables in a data set.<sup>15</sup>

Instead of using the original data, partial least squares (PLS) reduce the data to a smaller set of uncorrelated components and performs least squares regression on these components. It solves the multicollinearity problem by constructing latent vectors that explain both the independent and dependent variables. When more than one dependent variable needs to be predicted, this method is utilized.<sup>16</sup>

The dataset obtained had a wider range of values initially the data was preprocessed using median normalization. The significant metabolites were then identified by applying unpaired t-test on the Normalized Data. Top metabolites were selected as those with a significance value (P value) of less than 0.05 and were employed in further analysis. Following the t test; the top significant metabolites were further analyzed using the PCA method. The principal components were identified, and top 20 metabolites are extracted based on Variable Importance Number using statistica Software, which were then validated and considered as RCC biomarkers. Following PCA, the top metabolites from the t test were validated using the PLS method.

The top metabolites reported after PCA and PLS were compared to the literature, and the common metabolites found in all three were considered as potential biomarkers, which were subsequently studied to learn more about their toxic effects in Renal Cell Carcinoma.

Quantum chemical studies were performed on Identified Biomarkers to Study their Toxicity using Gaussian 09 Software. All the geometry optimizations were carried out using Density Functional Theory (DFT) employing 6-31+ G(d) basis set and the B3LYP functional.<sup>17</sup> Then the global electrophilicity index ( $\chi$ ) of identified biomarkers was calculated, which measures the energy of stabilization when an optimal electronic charge transfer from the environment to the system occurs. In order to prove the toxicity, Toxtree Tool<sup>18</sup> was used and Biological Significance of Biomarkers associated with Renal Cell Carcinoma was studied by identifying the molecular pathways in which metabolites are involved.

## Results and Discussion:

### 1. Identifying biomarkers using Predictive Statistical modelling:

#### 1.1 Data preprocessing by Median Normalization:

The data preprocessing was done by imputing the missing values and Median Normalization. The missing values were identified and they were replaced with median values. The Median Normalization was carried out for each metabolite. Those Normalized Values were used in further Analysis to identify Biomarkers.

#### 1.2. T- test:

For the Normalized dataset unpaired t test was carried out using Statistical Analysis Module of Emotionalist Tool<sup>19</sup> (<https://www.metaboanalyst.ca/>). The metabolites having the p-value less than 0.05 are considered as the significant metabolites which have a significant effect on Renal Cell Carcinoma. Based on the p-value 33 metabolites are identified significant metabolites and those were taken for further processing (Table1)

S.no	Metabolite Name	P. Value
1	Hippurate_2	7.72E-23

2	Hippurate_3	6.13E-22
3	Hippurate_4	9.38E-21
4	4-hydroxyphenylacetate_1	7.51E-19
5	Acetoacetate_2	8.25E-19
6	Aminohippurate	2.66E-14
7	Trigonellinamide_3	2.91E-13
8	Trigonellinamide_2	5.53E-12
9	Trigonellinamide_1	3.91E-11
10	Pyruvate	4.86E-11
11	Scylloinositol	7.49E-10
12	Glycine	1.89E-09
13	Acetoacetate_1	5.25E-09
14	Acetone	7.27E-08
15	Trigonelline_2	7.90E-08
16	Lactate	2.51E-07
17	Trigonelline_1	7.67E-06
18	Acetate	6.33E-05
19	Methyl guanidine	0.0001217
20	Hydroxyhippurate	0.00015553
21	Hippurate_1	0.00015557
22	UNK_2	0.00020524
23	Bile acid_1	0.0006775
24	Dimethyl amine	0.00070915
25	Indoxylsulfate	0.00088526
26	Citrate	0.00093854
27	Mannitol_1	0.0016132
28	UNK_1	0.0020896
29	Mannitol_2	0.0061651
30	UNK_8	0.011314
31	Hypoxanthine_1	0.019981
32	3-Hydroxyisovaleric acid	0.035289
33	Bile acid_2	0.049154

**Table 1:** t-test results Showing 33 Significant Metabolites

### 1.3. Principal Component Analysis (PCA)

The principal component analysis was performed using statistical package statistica 13.3 27 (TIBCO Software Inc) on top 33 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 2). The top 20 ranking metabolites were considered for further analysis

Variable	Power	Variable Importance
Hippurate_2	0.983053	1
Hippurate_1	0.981645	2
Hydroxy Hippurate	0.981619	3
Hippurate_3	0.978545	4
Hippurate_4	0.964275	5
Trigonellinamide_1	0.953488	6
Trigonellinamide_2	0.953253	7
4-hydroxyphenylacetate_1	0.936599	8
Acetoacetate_2	0.936507	9
Trigonellinamide_3	0.936067	10
Acetoacetate_1	0.908971	11
Scylloinositol	0.865268	12
Trigonelline_2	0.848829	13
Mannitol_1	0.845410	14
Trigonelline_1	0.840769	15
Mannitol_2	0.826084	16
Bile acid_1	0.808277	17
Bile acid_2	0.799725	18
Pyruvate	0.780170	19
Amino Hippurate	0.767515	20

**Table 2:** PCA Results Showing Top 20 Metabolites

#### 1.4. PLS Results

PLS analysis was performed to validate the results from PCA. The PLS was also Performed by statistical software and the variable importance of

each metabolite 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 3). The metabolites with top 20 ranks were selected as significant metabolites for identification of biomarkers.

Variable	VIP	Importance
UNK_8	2.180183	1
UNK_2	1.735567	2
Trigonelline_2	1.643086	3
Trigonelline_1	1.497001	4
Hippurate_1	1.446959	5
Hydroxy Hippurate	1.446872	6
amino Hippurate	1.399427	7
Glycine	1.287609	8
Indoxyl sulfate	1.218203	9
Scylloinositol	1.190392	10
Hippurate_2	1.165740	11
Hippurate_3	1.073573	12
Mannitol_1	0.983081	13
Hippurate_4	0.918059	14
Bile acid_2	0.887998	15
3-Hydroxyisovaleric acid	0.755359	16
Pyruvate	0.718033	17
Acetone	0.662696	18
Methyl guanidine	0.651489	19
Dimethyl amine	0.647092	20

**Table 3:** PLS Results Showing Top 20 Metabolites

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 8 metabolites which were common those are Trigonelline, Hippuric acid, 4-hydroxy hippuric acid, 4-amino hippuric acid, Mannitol, Pyruvic acid, Scylla- Inositol and Deoxycholic acid and these were further explored to know their toxic effects in Renal Cell Carcinoma.

#### 2. Electronic structure analysis of identified biomarkers

Quantum Chemical Calculations was performed using Gaussian 09 Software on 8 Biomarkers. All the Geometry Optimizations were carried out using Density Functional Theory (DFT) employing 6-31+G (d,p) basis set and the B3LYP Functional to get their HOMO and LUMO values which were further used for calculating global electrophilicity index (Table 4).

Biomarker name	EHOMO	ELUMO	$\mu$	IP	EA	$\eta$	$\phi$
Trigonelline	-0.27845	-0.00969	-0.14407	0.27845	0.00969	0.13438	2.10
Hippuric acid	-0.26858	-0.0116	-0.14009	0.26858	0.0116	0.12849	2.07
4-Hydroxy hippurate	-0.34151	-0.03411	-0.18781	0.34151	0.03411	0.1537	3.12
Amino hippurate	-0.33318	-0.02479	-0.178985	0.33318	0.02479	0.154195	2.82
Mannitol	-0.26769	-0.01304	-0.140365	0.26769	0.01304	0.127325	2.10
Pyruvic acid	-0.4292	-0.01653	-0.222865	0.4292	0.01653	0.206335	3.27
Scyllo-Inositol	-0.28024	-0.00886	-0.14455	0.28024	0.00886	0.13569	2.09
Deoxycholic acid	-0.20717	-0.03622	-0.121695	0.20717	0.03622	0.085475	2.35

**Table 4:** Global Electrophilicity index values of Biomarkers

The electrophilicity Index values of these biomarkers are High (>2), which denotes that these all are toxic metabolites and highly Responsible for the Disease. Hence the electrophilicity is responsible for the observed RCC.

#### 3. Toxfree Tool

Based on the decision tree approach by applying Cramer rule to the SMILES notation of Biomarkers, Toxfree software gives the Class of Toxicity. The software makes the decision based on the information available in the literature and classify the metabolites as High toxic, Intermediate Toxic and Low toxic. 18 The toxicity of top 8 metabolites according to Toxfree software is shown in Table 5

Metabolite name	Class of Toxicity	Toxicity Level
Trigonelline	Class III	High toxic
Hippuric acid	Class I	Low toxic
4-Hydroxy Hippuric acid	Class I	Low toxic

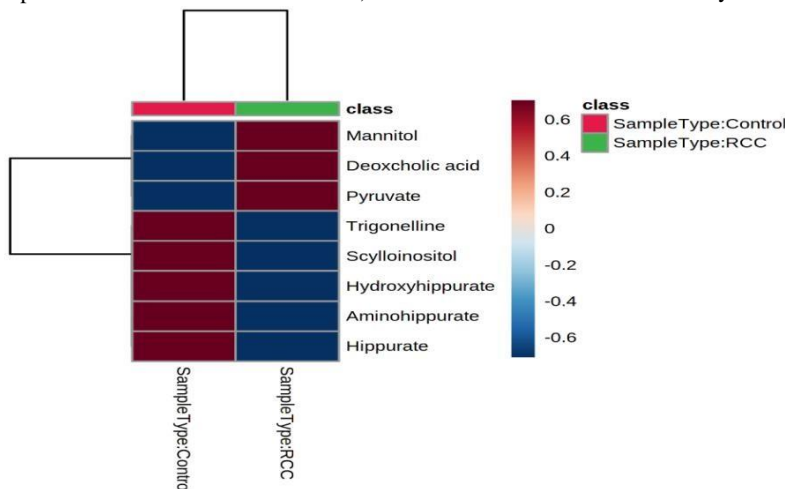
Amino Hippuric acid	Class I	Low toxic
Mannitol	Class I	Low toxic
Pyruvic acid	Class I	Low toxic
Scyllo-Insitol	Class I	Low toxic
Deoxycholic acid	Class III	High toxic

**Table 5:** Toxicity of top metabolites according to Toxfree software

#### 4. Relative Quantification of metabolites by Heatmap

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.<sup>19</sup> Heat Map of Renal Cell Carcinoma Biomarkers was constructed using Emotionalist 5.0 Tool (Figure 1) up regulated and down regulated metabolites can be seen clearly in the Table 6



**Figure 1:** Heat Map of RCC Biomarkers

Upregulated Metabolites	Downregulated Metabolites
Mannitol	Trigonelline
Deoxycholic acid	Scyllo-Insitol
Pyruvic acid	4 - hydroxy hippuric Acid
	Amino hippuric Acid
	Hippuric Acid

**Table 6:** Up regulated and down regulated metabolites

The concentration of up regulated Metabolites like Mannitol, Deoxycholic Acid and Pyruvic acid is increased in RCC patients when compared to Normal Healthy Individuals. Similarly for Down Regulated Metabolites like Trigonelline, Scyllo-Insitol, 4 - hydroxy hippuric acid, amino hippuric acid and hippuric acid, the concentration decreased in RCC patients when compared to Normal Healthy Individuals. From these variations in the concentrations of the metabolites in diseased patients in comparison to normal healthy individuals, it can be concluded that a set of metabolites are down regulated, while another set were unregulated significantly in diseased individuals which can help in prognosis of the Renal cell carcinoma. These results were consistent with the existing literature.

#### 5. ROC Curve Analysis

A receiver operating characteristic (ROC) analysis was performed using Emotionalist 5.0 on these 8 biomarkers to assess their diagnostic accuracy

and characterize them in early stage of RCC.<sup>20</sup> The AUC, sensitivity, specificity, and 95 % confidence intervals of the eight identified Potential urinary biomarkers for RCC early diagnosis are shown in Table 7. To evaluate the diagnostic accuracy of these identified potential urinary biomarkers for RCC, a predictive model for patient classification was constructed using each identified biomarker. Pyruvic acid and Deoxycholic acid showed a AUC Value of 0.854 (Figure 2) and 0.807 (Figure 3) respectively and regarded as best biomarkers having high predictive accuracy and better distinguish the Control and RCC group when compared to other 6 biomarkers. Overall, 8 metabolites showed clinical potential diagnostic value, with an AUC of 0.923 (Figure 4) and an 81 % Predictive Accuracy (Figure 5). It has been observed that combination of biomarkers is more helpful than single biomarkers for early diagnosis of the disease.

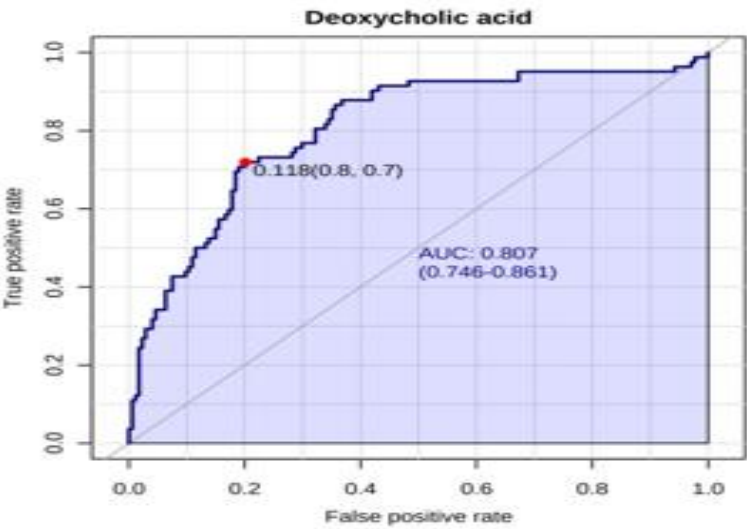


Figure 2: AUC – ROC Curve of Pyruvic acid

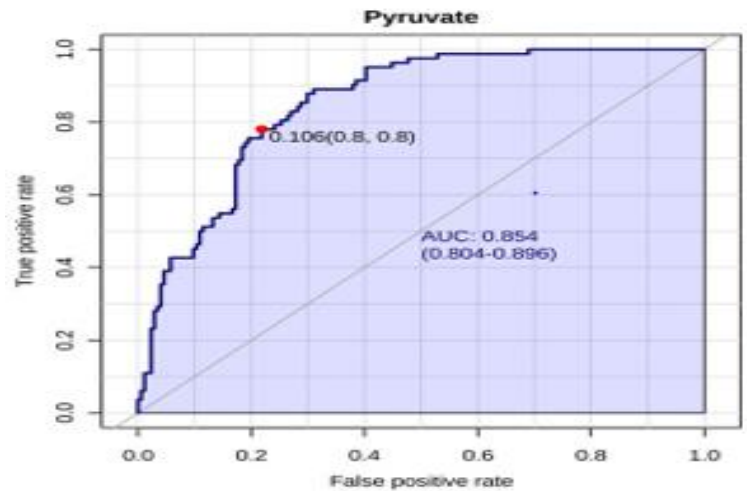


Figure 3: AUC – ROC Curve of Deoxycholic acid

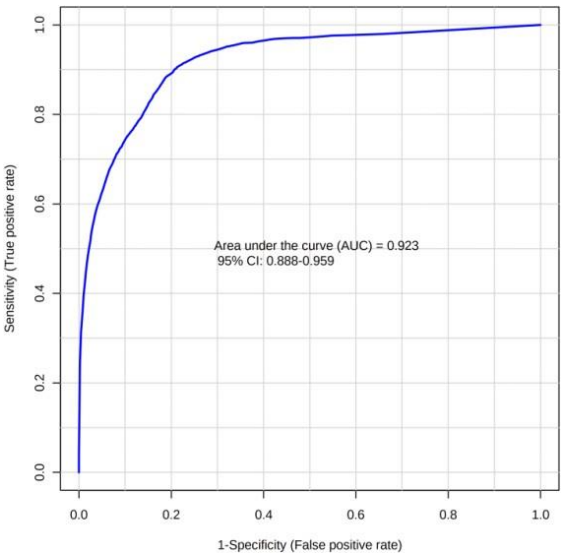
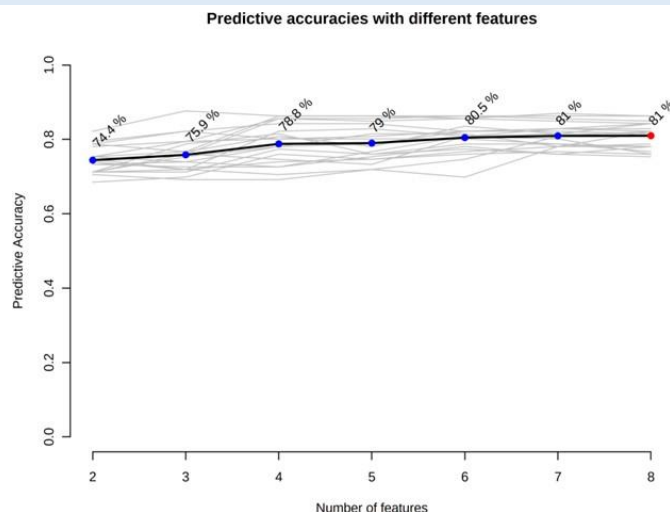


Figure 4: AUC – ROC Curve of All Metabolite Panel





**Figure 5:** Predictive Accuracy with All Metabolite Panel

Biomarker	AUC	C.I	Sensitivity	Specificity	Std.error
Pyruvic acid	0.854	0.804 – 0.894	80 %	80 %	0.106
Deoxycholic acid	0.807	0.746 – 0.861	80 %	70 %	0.118
Aminohippuric acid	0.735	0.675 – 0.804	70 %	80 %	-0.0194
Scyllo-inositol	0.658	0.591- 0.723	60%	70%	-0.00854
Mannitol	0.634	0.559 – 0.697	50%	80%	-0.0425
Hippuric acid	0.516	0.453 – 0.582	50%	70%	0.0851
hydroxyhippuric acid	0.515	0.447 – 0.588	60%	50%	0.0851
Trigonelline	0.501	0.414 – 0.577	60%	50%	-0.0187

**Table 7:** ROC Curve Analysis

## 6. The metabolic pathways in which identified metabolites are involved

### 6.1. Trigonelline:

Márcia S. Monteiro et al (2016) in “Nuclear Magnetic Resonance metabolomics reveals an excretory metabolic signature of renal cell carcinoma” Reported that Trigonelline could be related to certain foods (e.g. coffee), but it can also be produced by endogenous niacin methylation. Reduced excretion of Trigonelline was found to be reported in patients with liver cancer, ovarian cancer, pancreatic ductal adenocarcinoma, and lung cancer. This change, together with the decreasing tendency of trigonelline shows that nicotinate and nicotinamide metabolism is disturbed. Putative interpretation of the identified metabolites changing in RCC compared to controls revealed possible unspecific effects involving Hippurate, trigonelline, and trigonelline, emphasizing the importance of diet and gut microflora, as well as nicotinate and nicotinamide metabolism and anti-oxidative mechanisms as less specific systemic cancer effects.<sup>20</sup> The Nicotinate and Nicotinamide Pathway in which Trigonelline metabolite is collected from KEGG Pathway Database

### 6.2 Pyruvic acid:

Márcia S. Monteiro et al (2016) in “Nuclear Magnetic Resonance metabolomics reveals an excretory metabolic signature of renal cell

carcinoma” Reported that the increased levels of excreted pyruvic acid imply enhanced glycolysis activity. Increased glycolytic flux and altered TCA cycle function are well-known cancer hallmarks, affecting not only cellular energetic efficiency but also anabolic/biosynthetic efficiency, because intermediates in these pathways are diverted to the synthesis of proteins, nucleic acids, lipids, and cholesterol, and generally aid in the maintenance of cellular redox, genetic, and epigenetic status required for cancer cell proliferation.<sup>20</sup> The Glycolysis Pathway in which pyruvic acid metabolite is collected from KEGG Pathway Database

### 6.3. Deoxycholic acid:

Márcia S. Monteiro et al (2016) in “Nuclear Magnetic Resonance metabolomics reveals an excretory metabolic signature of renal cell carcinoma” Reported that the presence of higher levels of Deoxycholic acid (Bile Acid) in both blood and urine of hepatocellular carcinoma patients as bile acid resonances implies that RCC has an influence on endogenous cholesterol metabolism. Bile acids are considered to be a significant indication of Liver injury. However, however at point, the exact connection between these compounds and RCC is unclear.<sup>20</sup> The Secondary Bile Acid Biosynthesis Pathway in which Deoxycholic acid metabolite is collected from KEGG Pathway Database

### 6.4. Scyllo-Insitol / Myo-inositol:

Piotr Popławski et.al, (2017) in “Integrated transcriptomic and metabolomic analysis shows that disturbances in metabolism of tumor cells contribute to poor survival of RCC patients” reported that in RCC tumors Myo-inositol level was reduced by 16-fold and was associated with decreased MIOX. However, given the lower myo-inositol levels in RCC tumours, a decrease in MIOX expression may appear counter-intuitive. Decreased myo- inositol level in RCC tumours may be the result from increased excretion in urine. This study suggested that Myo-inositol and Myo-inositol pathway may offer attractive targets for potential treatment of RCC patients.<sup>21</sup> The Inositol Phosphate Pathway in which Myo-inositol metabolite is collected from KEGG Pathway Database

### 6.5. Mannitol:

Leuthold et al. (2017) conducted a comprehensive metabolomic and lipidomic profiling of human and porcine kidney tissue study using a LC-QTOF-MS analytical approach. PCA analysis showed differentiation in aqueous extracted metabolites from RCC compared to adjacent nontumor tissue. The metabolite level of metabolite Mannitol is increased in the RCC tissue. This Study concluded that Disruption in fructose and mannose metabolism pathway of Mannitol is responsible for RCC.<sup>22</sup> The Fructose and Mannose metabolism pathway in which Mannitol metabolite is collected from KEGG Pathway Database

### 6.6. Hippuric acid, 4-hydroxyhippuric acid and 4-aminohippuric acid:

Daniela Rodrigues et.al, (2017) Márcia, Monteiro.; Carmen, Jerónimo.; Rui, Henrique.; Luís, Belo.; Maria de Lourdes Bastos, Paula Guedes de Pinho, Márcia, Carvalho.; Renal cell carcinoma: a critical analysis of metabolomic biomarkers emerging from current model systems, Translational Research, 2017, 180, 1-11 Hippuric acid, 4-hydroxyhippuric acid and 4-aminohippuric acid, phenylalanine downstream metabolites showed potential as RCC biomarkers. It was found diminished in urine of RCC patients, suggesting that cancer cells rapidly metabolize it, which might cause impaired secretion into renal tubules. Hippuric acid was also found decreased in other renal diseases, emphasizing its poor specificity as an individual biomarker for RCC detection.<sup>23</sup> The Phenylalanine Pathway in which Hippuric acid, 4-hydroxyhippuric acid and 4-aminohippuric acid metabolites is collected from KEGG Pathway Database

The above-mentioned metabolic pathways have been associated with Renal Cell Carcinoma in the literature, indicating that alterations in these pathways can cause Renal Cell Carcinoma.

## Conclusion

Renal cell carcinoma (RCC) is a heterogeneous disease that is usually asymptomatic until late stage. There is an urgent need for RCC specific biomarkers identification that may be exploited clinically for diagnostic and prognostic purposes. In this study, metabolomic data was statistically explored for the identification of biomarkers in Renal Cell Carcinoma.

The raw dataset obtained from Metabolomics Society was preprocessed using Emotionalist 5.0, a web based interface used for Metabolomics Data Analysis, and further statistical methods such as t-tests, PCAs, and PLS were applied using Statistical Software to identify Significant Metabolites. The top ranking eight metabolites obtained after both PCA and PLS were compared with the literature and the common metabolites present in all the three were considered as the potential biomarkers. Trigonelline, Hippuric acid, 4-hydroxy hippuric acid, 4-amino hippuric acid, Mannitol, Pyruvic acid, Scyllo-Insitol and Deoxycholic acid are

considered as the Biomarkers of Renal Cell Carcinoma. These metabolites were further explored to understand their effect on the body. The quantum chemical studies were performed using Gaussian Software to evaluate their Electronic Parameters and Toxicity. A Toxicity Prediction tool called Toxfree was used to classify the biomarkers on the basis of toxicity, which showed 2 metabolites i.e. Trigonelline, Deoxycholic acid in the high toxic class (Class III) and 6 metabolites i.e. Hippuric acid, 4-hydroxy hippuric acid, 4-amino hippuric acid, Mannitol, Pyruvic acid and Scyllo-Insitol in Low Toxic Class (Class I). Relative Quantification using Heatmap showed that Mannitol, Deoxycholic Acid and Pyruvic acid are upregulated metabolites and Trigonelline, Scyllo-Insitol, 4 - hydroxy hippuric acid, 4- amino hippuric acid and hippuric acid are downregulated metabolites. From ROC curve analysis it has been observed that combination of biomarkers is more helpful than single biomarkers for early diagnosis of the disease. The toxic metabolites were further explored for their metabolic pathways and their association with Renal Cell Carcinoma using the literature. It can be predicted that the disturbance in the identified metabolic pathways due to these eight metabolites may be the cause of observed Renal Cell Carcinoma.

## Abbreviations

RCC	Renal Cell Carcinoma
MIOX	Myo-Inositol Oxygenase
SMILES	Simplified Molecular Input Line Entry System
HOMO	Highest Occupied Molecular Orbital
LUMO	Lowest Unoccupied Molecular Orbital
AUROC	Area Under Receiver Operating Characteristics
LC-MS	Liquid Chromatography–Mass Spectrometry
AUC	Area Under Curve
ROC	Receiver Operator Characteristic
PCA	Principal Component Analysis
PLS	Partial Least Square
DFT	Density Functional Theory

## Funding

Thanks to Department of Biotechnology, Ministry of Science and Technology, Government of India under DBT project number - BT/PR40164/BTIS/137/17/2021.

## References

1. Hsieh, J, Purdue, M, Signoretti, S, (2016). Renal cell carcinoma. *Nat Rev Dis Primers*. 2017 ,3, 17009
2. Jun, G.; Manuel, C M.; Nazli, D.; Ameish, G.; Sumanta, K. P.; Metastasis in renal cell carcinoma: *Biology and implications for therapy*. *Asian Journal of Urology*. 3 ,286-292.
3. Brian, I Rini.; Steven., C Campbell, Bernard, Escudero.; (2009). Renal cell carcinoma, *The Lancet*.,373,1119-1132
4. Johnson, C.H., Ivanisevic, J. and Siuzdak, G., (2016). Metabolomics: beyond biomarkers and towards mechanisms. *Nature reviews Molecular cell biology*. 17, 451-459.
5. Jeffrey R. Idle, Frank J. (2007). Gonzalez, Metabolomics, *Cell Metabolism*,6.



6. Mamas, M.; Dunn, W. B.; Neryses, L.; Goodacre, R. (2011). The role of metabolites and metabolomics in clinically applicable biomarkers of disease. *Arch. Toxicol.*, 85, 5-17.
7. Olatomiwa O. Biafran.; David A. Gaul.; Samyukta Sah, Rebecca S. Arnold, et al., (2021). ; Machine Learning-Enabled Renal Cell Carcinoma Status Prediction Using Multiplatform Urine-Based Metabolomics. *J. Proteome Res.* 20, 3629–3641.
8. Falegan, O.S.; Ball, M.W.; Shaykhutdinov, R.A.; Pieroraio, et al., (2017). M.E. Urine and Serum Metabolomics Analyses May Distinguish between Stages of Renal Cell Carcinoma. *Metabolites*, 7, 6.
9. Zheng H, Ji J, Zhao L, Chen M, Shi A, et al., (2016). Prediction and diagnosis of renal cell carcinoma using nuclear magnetic resonance-based serum metabolomics and self-organizing maps. *Oncotarget*. 7, 59189-59198.
10. Xi B.; Gu H.; Baniyadi H.; Raftery D.; (2014). Statistical Analysis and Modeling of Mass Spectrometry-Based Metabolomics Data. In: Raftery D. (eds) *Mass Spectrometry in Metabolomics. Methods Mol. Bio.*, 1198.
11. Szymańska, E.; Saccenti, E.; Smilde, A.K.; (2012). Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*, 8, 3– 16.
12. Kyoungmi Kim, Sandra L. Taylor, Sheila Ganti, Lining Guo.; et al., (2011). Urine Metabolomic Analysis Identifies Potential Biomarkers and Pathogenic Pathways in Kidney Cancer. *OMICS: J. Integr. Biol.* 293-303.
13. Alysha M. De Livera.; Daniel A. Dias.; David De Souza.; Thusitha Rupasinghe.; James Pyke.; et al., (2012). *Normalizing and Integrating Metabolomics Data. Anal. Chem.* 84, 10768-10776.
14. Pierre Baldi; Anthony D. Long, (2001). A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. *Bioinformatics*, 17, 509–519.
15. Bartel, J; Krumsiek, J, Theis, F.J., (2013). Statistical methods for the analysis of high- throughput metabolomics data. *Comput. Struct. Biotechnology. J.* 4.
16. Szymańska, E; Saccenti, E, Smilde, A.K.; Westerhuis, J.A.; (2012). Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics*, 8 ,3-16.
17. Parr, R. G.; Szentpály, L. v.; Liu, S. (1999). Electrophilicity Index. *J. Am. Chem. Soc.* 121, 1922–1924.
18. Patlewicz, G.; Jeliaskova, N.; Safford, R.J.; Worth, A.P.; Aleksiev, B.; (2008). An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR QSAR Environ. Res.* 19, 495-524.
19. Jianguo, Xia; Nick, Psychogios.; Nelson, Young.; David, S. Wishart.; (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research*, 37, 652–660
20. Soreide, K.; Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research. *J. Clin. Pathol.*, , 62, 1-5
21. Monteiro, M.; Barros, A.; Pinto, J.; Carvalho M.; Ana S. Et al., (2016). Nuclear Magnetic Resonance metabolomics reveals an excretory metabolic signature of renal cell carcinoma. *Sci Rep*, 6, 37275.
22. Piotr, Popławski.; Takayuki, Tohge.; Joanna, Bogusławska.; Beata, Rybicka.; Zbigniew, Tański.; et al., (2017). Integrated transcriptomic and metabolomic analysis shows that disturbances in metabolism of tumor cells contribute to poor survival of RCC patients. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 3,744-752
23. Leuthold P, Schaeffeler E, Winter S, Büttner F, Hofmann U, et al., (2017). Comprehensive metabolomic and lipidomic profiling of human kidney tissue: a platform comparison. *J. Proteome Res* 16, 933-944.
24. Daniela,Rodrigues.; Márcia, Monteiro.; Carmen, Jerónimo.; Rui, Henrique.; Luís, Belo et al., (2017,).a critical analysis of metabolomic biomarkers emerging from current model systems, *Translational Research*, 180,1-11

**Ready to submit your research? Choose ClinicSearch and benefit from:**

- fast, convenient online submission
- rigorous peer review by experienced research in your field
- rapid publication on acceptance
- authors retain copyrights
- unique DOI for all articles
- immediate, unrestricted online access

**At ClinicSearch, research is always in progress.**

Learn more <https://clinicsearchonline.org/journals/clinical-endocrinology-and-metabolism>



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.