



OPEN

# Urine metabolomics based prediction model approach for radiation exposure

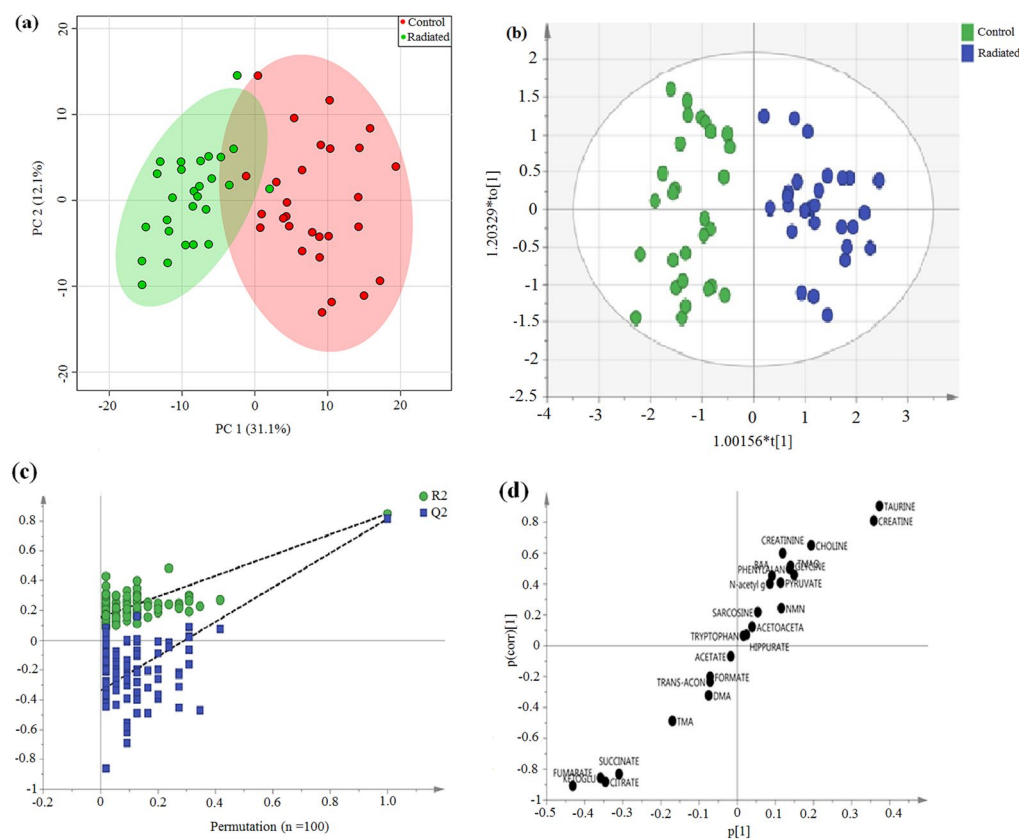
Ritu Tyagi<sup>1,3</sup>, Kiran Maan<sup>1,3</sup>, Subash Khushu<sup>2</sup> & Poonam Rana<sup>1</sup>✉

The radiological incidents and terrorism have demanded the need for the development of rapid, precise, and non-invasive technique for detection and quantification of exposed dose of radiation. Though radiation induced metabolic markers have been thoroughly investigated, but reproducibility still needs to be elucidated. The present study aims at assessing the reliability and reproducibility of markers using nuclear magnetic resonance (NMR) spectroscopy and further deriving a logistic regression model based on these markers. C57BL/6 male mice (8–10 weeks) whole body  $\gamma$ -irradiated and sham irradiated controls were used. Urine samples collected at 24 h post dose were investigated using high resolution NMR spectroscopy and the datasets were analyzed using multivariate analysis. Fifteen distinguishable metabolites and 3 metabolic pathways (TCA cycle, taurine and hypotaurine metabolism, primary bile acid biosynthesis) were found to be amended. ROC curve and logistic regression was used to establish a diagnostic model as  $\text{Logit}(p) = \log(p/1-p) = -0.498 + 13.771(\text{tau}) - 3.412(\text{citrate}) - 34.461(\alpha\text{-KG}) + 515.183(\text{fumarate})$  with a sensitivity and specificity of 1.00 and 0.964 respectively. The findings demonstrate the proof of concept and the potential of NMR based metabolomics to establish a prediction model that can be implemented as a promising mass screening tool during triage.

The increasing burden of natural background radiation and terrestrial radionuclides is a major concern for exposure to radiation to the mass population. According to U.S. commission report on prevention of munitions for terrorism and mass destruction<sup>1</sup>, accidental radiation exposure is a matter of global worry that needs emergency preparedness, policy makers to prevent undesirable exposures as well as to treat and manage exposed individuals. Additionally, the accidental nuclear disasters of Fukushima and the growing risks of radiological terrorism demanded the development of biomarkers of ionizing radiation and countermeasures for rapid and accurate measurement of absorbed radiation dose for mass screening of exposed individuals<sup>2–4</sup>. There is a clear need for high throughput biomonitoring and diagnostic platform that is easily deployable, rapid, reproducible, reliable, and appropriate to triage and permit rapid assessment of possibly massive radiation exposure casualties. Recognition of predictive biomarker(s) that have the potential to detect radiation induced syndromes at the earliest preceding the onset of organ specific damage is the need of the hour<sup>2,5,6</sup>.

Metabolomics is one such high-throughput technology that has great potential for radiation bio-dosimetry that can play an imperative role during the initial triage of radiological disasters to characterize the metabolic status of an individual before the onset of symptoms. It is about a decay old omics technique and in the last few years, its application has been extended towards qualitative and quantitative assessment of ionization radiation exposure induced response mainly in biological fluids, serum and urine<sup>7–14</sup>. Liquid chromatography–mass spectrometry (LC–MS) based studies have elucidated many urinary metabolic markers for radiation exposure pertaining to oxidative stress, energy metabolism, DNA damage, inflammation, and tissue damage<sup>4,15,16</sup>. NMR spectroscopy based metabolomics due to its ease and reproducibility over LC–MS has allowed a time and dose responsive comprehensive coverage of urinary metabolites in toxicological and clinical research. However, NMR based metabolomics has been less explored in the case of radiation research. Metabolomics reveals the pathophysiological perturbations to metabolic pathways that could serve as predictive markers for possible health outcomes and therefore, may serve as targets for therapeutic intervention post radiation exposure. Performance

<sup>1</sup>Metabolomics Research Facility, Institute of Nuclear Medicine and Allied Sciences (INMAS), DRDO, S. K Mazumdar Road, Timarpur, Delhi 110054, India. <sup>2</sup>NMR Research Centre, Institute of Nuclear Medicine and Allied Sciences (INMAS), DRDO, Delhi, India. <sup>3</sup>These authors contributed equally: Ritu Tyagi and Kiran Maan. ✉email: poonam@inmas.drdo.in



**Figure 1.** Multivariate analysis derived from  $^1\text{H}$  NMR spectra of urine sample from control and irradiated animals. (a) PCA score plot, (b) OPLS-DA score plot, (c) OPLS-DA validation plot (permutation times  $n = 100$ ) and (d) The OPLS-DA derived corresponding S-plot.

evaluation and final predictive model testing of biomarkers are absolutely required for translational research or for the development of a high throughput bio-dosimetry device for mass screening. Primarily, role of metabolomics in the development of biomarker is to focus on building a predictive model that could be used to classify unidentified sample(s) into specific groups with the best specificity and sensitivity<sup>17</sup>.

Our previous published study has demonstrated radiation induced patho-physiological perturbations in urine using  $^1\text{H}$  NMR based metabolomics wherein, metabolite markers were found to be associated with energy metabolism, amino acids, and gut flora metabolism<sup>18</sup>. However, reproducibility, validation, and estimation of sensitivity and specificity of identified markers are the prerequisite step towards the translational approach. Therefore, the present study has been conceived to provide a proof of concept and establish a model for identification of radiation exposure based on metabolite markers identified through NMR spectroscopy. To the best of our knowledge, this is the first study that looks into the potentiality of the prediction model approach for the identification of NMR spectroscopy based radiation markers using a logistic regression model.

## Results

**Metabolomic analysis of urine.** Representative  $^1\text{H}$  NMR spectra with identified metabolites are shown in supplementary Fig. 1. The competency of  $^1\text{H}$  NMR based metabolomics approach to differentiate irradiated from non-irradiated controls was assessed using multivariate analysis in our study. Principal component analysis (PCA) score plot displayed clear separation between the irradiated and control group (Fig. 1a). From the corresponding loading plots of PCA, the chemical shifts values that lied far away from origin has the greatest contribution and substantially responsible for demarcation between irradiated and control group and therefore may be regarded as the discriminating metabolites for the irradiated group. Based on PCA, 24 metabolites were identified (supplementary Fig. 2), which were then used to perform supervised clustering, orthogonal partial least squares discriminant analysis (OPLS-DA) (Fig. 1b). The analysis produced a strong model with high validated predictability ( $Q^2$  cum = 0.813) and goodness of fit value ( $R^2$  cum = 0.85). The values greater than 0.5 indicates that model possessed a reasonable fit with good predictive power<sup>19,20</sup>. The permutation plot with  $n = 100$  permutations further confirmed the constructed OPLS-DA model was positive and valid as all permuted values were lower than the original values at the right and the line plot intercepts Y axis below zero (Fig. 1c). The corresponding S-plot shows the potential 6 metabolites responsible for demarcation between the control and irradiated group (Fig. 1d). Overfitting of the OPLS-DA model was further confirmed by cross validated ANOVA (CV ANOVA) and highly significant  $p$  value ( $1 \times e^{-17}$ ) of CV ANOVA score suggests the strengthening and robustness of the model since score can be used as a guide for the optimal fitting of a model<sup>20</sup>.

Metabolites	HMDB ID	<i>p</i> value <sup>a</sup>	VIP <sup>b</sup>	R <sup>c</sup>	Fold changed <sup>d</sup>
Taurine	HMDB0000251	$3.49 \times 10^{-15}$	1.70	0.85	0.31
Citrate	HMDB0000094	$1.74 \times 10^{-13}$	1.58	-0.83	2.54
Creatine	HMDB0000064	$5.08 \times 10^{-12}$	1.69	0.77	0.29
$\alpha$ KG	HMDB0000208	$1.63 \times 10^{-13}$	1.66	-0.75	3.49
Fumarate	HMDB0000134	$6.37 \times 10^{-15}$	2.00	-0.70	4.69
Succinate	HMDB0000254	$4.62 \times 10^{-12}$	1.47	-0.69	2.40
Choline	HMDB0000097	$1.84 \times 10^{-08}$	0.86	0.69	0.62
Creatinine	HMDB0000562	$6.13 \times 10^{-05}$	0.55	0.54	0.84
Glycine	HMDB0000123	$1.41 \times 10^{-04}$	0.66	0.52	0.71
Phenylalanine	HMDB0000159	$1.35 \times 10^{-03}$	0.65	0.48	0.75
Pyruvate	HMDB0000243	$5.22 \times 10^{-03}$	0.74	0.45	0.80
TMAO	HMDB0000925	$1.42 \times 10^{-03}$	0.75	0.42	0.71
Branched amino acids (BAA)	HMDB0000687 (L-Leucine), HMDB0000172 (L-Isoleucine), HMDB0000883 (L-Valine)	$2.10 \times 10^{-03}$	0.49	0.40	0.86
TMA	HMDB0000906	$2.15 \times 10^{-03}$	0.82	-0.40	1.53
N-Acetyl glycoprotein	HMDB0000215	$1.38 \times 10^{-02}$	0.40	0.40	0.88

**Table 1.** List of 15 Key metabolites responsible for discriminating irradiated and control group. <sup>a</sup>*p* values were derived from two-tailed Student's *t* test. <sup>b</sup>Variable Importance in the projection (VIP) was obtained from OPLS DA with a threshold of 1.0 <sup>c</sup>Correlation coefficient was obtained from OPLS DA with a threshold of 1.0 <sup>d</sup>Positive values indicate higher levels in irradiated group and negative values indicate lower levels in irradiated group.

**Identification of important differential metabolites.** PCA analysis identified 24 metabolites, however, based on the correlation coefficient with threshold value of ( $|r| > 0.3$ ) ( $r > 0.3$  and  $r < -0.3$ ), out of the 24 metabolites, 15 differential metabolites were identified (Table 1). Further, supervised analysis was able to detect 6 metabolites out of 15 metabolites on the basis of VIP score ( $> 1$ ) and visualized S plot. These 6 differential metabolites i.e. taurine, creatine, succinate, citrate,  $\alpha$ -ketoglutarate ( $\alpha$ -KG) and fumarate were then used to perform predictive modeling using stepwise logistic regression analysis. The result ascertained that among these 6 metabolites, only 4 metabolites i.e. taurine, citrate,  $\alpha$ -ketoglutarate ( $\alpha$ -KG) and fumarate had the greatest and significant contribution to the combined model between irradiated and control groups (Table 2a) as the model using all 6 differential metabolites could not be converged. Therefore the final model was built using 4 metabolites only. The diagnostic performance of the model and individual metabolites was quantified through the Receiver operating characteristic (ROC) curve. All six metabolites had an area under the curve (AUC)  $> 0.90$  depicting high range of sensitivity (77.78–100%) and specificity (89.29–100%) (Table 2b and supplementary Fig. 3). However, the combined metabolites were the better discriminator than each metabolite individually as depicted in Table 2a and supplementary Fig. 3 indicating strong diagnostic performance of combined biomarker panel. The final diagnostic model was expressed as:  $\text{Logit}(p) = \log(p/1-p) = -0.498 + 13.771(\text{tau}) - 3.412(\text{citrate}) - 34.461(\alpha\text{-KG}) + 515.183(\text{fumarate})$ . The corresponding ROC curve had an AUC of 0.999 (95% confidence interval (CI) 0.933 to 1.00) with a Youden index *J* of 0.964 and sensitivity and specificity of 100 and 96.4 respectively (Fig. 2). The training/discovery set displayed a sensitivity of 0.947 and specificity of 1.00 with an AUC of 0.998. The model was also validated with tenfold cross validation which had a sensitivity, specificity, and AUC of 0.926, 0.893, and 0.889 respectively (supplementary Fig. 4). This further verifies the robustness of the diagnostic model for the distinctive classification of irradiated and control groups.

**Identification of metabolic pathways.** To recognize the metabolic pathways that were disturbed due to radiation, metabolic pathway analysis was carried out on 15 most contributory metabolites with  $|r| > 0.3$  and  $p < 0.05$ . Out of the 28 pathways analyzed, 3 pathways with pathway impact  $> 0.05$ ,  $-\log(p) > 35$  and with good hits (5, 1 and 3 for Tri-carboxylic acid (TCA) cycle, taurine and hypotaurine metabolism and primary bile acid biosynthesis respectively) were considered. Of all the 3 pathways analyzed the taurine and hypotaurine pathway was recognized as the most considerably altered pathway between control and irradiated group with highest impact value of 0.43. For the TCA cycle among the total of 20 metabolites present in the pathway, 5 metabolites (succinate, pyruvate,  $\alpha$ -K.G., fumarate and citrate) has been found altered in the present study (Fig. 3).

## Discussion

Nuclear accidents, dispersal of radiological devices and exposure of radioactive materials to mass population or individual demanded the need to develop fast and trustworthy means of biological dosimetry for recognition of radiation exposed individuals. Though cytogenetics is the gold standard for radiation bio-dosimetry however, metabolomics also has the potential for the first level of assessment of exposed individuals. It is nearly impossible to get the human subjects/patients exposed to whole body radiation exposure, as therapeutic radiation cannot mimic whole body radiation exposure scenarios of any radiological incident or accident. Also, the patients undergoing radiation therapy either receive localized or focal radiation and constitute a heterogeneous population

	Prediction models	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>	Youden index (J)	Sensitivity	Specificity
<b>(a)</b>							
P1	$p = [1/1 + e^{-\{-9.026 + 4.319 * \text{taurine}\}}]$	0.989	0.0084	0.916–1.00	0.891	96.3	92.86
P2	$p = [1/1 + e^{-\{-4.161 + 3.697 * \text{taurine} - 0.894 * \text{citrate}\}}]$	0.995	0.0058	0.925–1.00	0.963	96.3	100.00
P3	$p = [1/1 + e^{-\{-5.057 + 6.006 * \text{taurine} - 0.500 * \text{citrate} - 5.335 * \alpha \text{ keto glutarate}\}}]$	0.997	0.0032	0.930–1.00	0.963	96.3	100.00
P4	$p = [1/1 + e^{-\{-0.498 + 13.771 * \text{taurine} - 3.412 * \text{citrate} - 34.461 * \alpha \text{ keto glutarate} + 515.183 * \text{fumarate}\}}]$	0.999	0.0018	0.933–1.00	0.964	100.00	96.4
<b>(b)</b>							
	Metabolites	AUC	SE <sup>a</sup>	95% CI <sup>b</sup>	Youden index (J)	Sensitivity	Specificity
1	Citrate	0.966	0.033	0.878–0.996	0.927	96.3	96.43
2	Taurine	0.989	0.0084	0.916–1.00	0.891	96.3	92.86
3	Fumarate	0.976	0.0192	0.893–0.999	0.927	96.3	96.43
4	$\alpha$ -Ketoglutarate	0.955	0.0298	0.862–0.993	0.892	100.00	89.29
5	Creatine	0.954	0.0238	0.861–0.992	0.777	77.8	100.00
6	Succinate	0.954	0.0256	0.861–0.992	0.817	88.89	92.86

**Table 2.** Prediction models from the logistic regression and the ROC analysis results of the combined (a) and individual metabolites (b). <sup>a</sup>Standard error. <sup>b</sup>Confidence interval.

based on the dose and area of radiation. Moreover, because of comorbidity, these patients cannot become true models/subjects of radiation exposure. Therefore, the present study has been designed in preclinical model to provide a proof of concept to an established prediction model for whole body radiation exposure using NMR based metabolomics. Urinary metabolomics has evolved as one of the most promising biomarker platform as urine can be easily collected and is more stable compared to other bio-fluids. Urinary radiation metabolomics has identified radiation induced metabolic markers however, the reproducibility, reliability, and validation of these markers still need to be elucidated. The determination of sensitivity and specificity of these metabolic markers is very much essential for delineation of early exposure to radiation and thereafter early/prompt intervention.

For disease model, metabolomics have been widely utilized for single biomarker discovery however, multiple biomarkers or panels might be better for discovering candidates with higher specificity and sensitivity. Thus the aim of this study is to assess the utility of multiple biomarkers for radiation exposure using multivariate analysis and logistic regression using stepwise variable selection method.

PCA revealed clear differences between controls and irradiated groups. However, OPLS DA may be an appropriate diagnostic model as it is a better discriminator in categorizing irradiated and control groups. The established prediction model based on four markers, i.e. taurine, citrate,  $\alpha$ -KG and fumarate with high sensitivity, specificity, and AUC suggests that they could be the candidate biomarkers for radiation exposure. It also indicates that individual metabolic biomarkers cannot be ideal for radiation exposure screening. The pathway analysis has shown that these metabolites are closely linked with the citric acid cycle, hypotaurine and taurine metabolism and primary bile acid biosynthesis.

With the use of multivariate statistical analysis and prediction modeling, the findings of our earlier study have been validated and show reproducible results. In the last few years, many LC–MS metabolomics studies have identified many metabolites associated with radiation exposure<sup>16,21</sup>. However, there are few NMR based studies that have exploited the potential of metabolomics for the identification of potent metabolite markers<sup>22–25</sup>. Undoubtedly the sensitivity of LC–MS is way ahead of NMR for profiling but the ease of acquisition, identification of metabolites and reproducibility of information in spectrum makes NMR based metabolomics a suitable and comfortable platform in biomarker identification. Notably, the present, as well as earlier study, have shown several NMR based metabolite markers associated with energy metabolism (citrate,  $\alpha$ -K.G, succinate, fumarate), amino acids (leucine/Isoleucine), creatine, creatinine, hippurate and taurine overlapping with LC–MS studies<sup>6,18,25–27</sup>. Few of the identified metabolite markers in the present study are also in accordance with other radiation metabolomics studies<sup>6</sup>. The present study and the recent literature suggest taurine and citrate amongst the candidate biomarkers for radiation exposure. Therefore, ease and reproducibility of information obtained by NMR spectroscopy make these markers as candidate markers for radiation exposure.

Taurine and hypotaurine metabolism has the highest impact score in our metabolic pathway analysis. Among the known urinary metabolites of radiation exposure taurine is well documented, but its exact cause for its upregulation following radiation is still not known. Its biological function includes bile acid conjugation and it also acts as anti-oxidant to protect the body by inhibiting reactive oxygen species (ROS)<sup>28</sup>. It might play a direct protective role for cell by preventing the ionic and water shifts that lead to cellular damage and death. The increased urinary excretion of taurine post high dose of ionizing radiation might be due to cellular damage (i.e. damage to the plasma membrane) that results in enhanced passive leakage of taurine out of cells<sup>29</sup>. Additionally, tissue injury induced by radiation results in a greater concentration of circulating sulfur-containing amino acids that are excreted in the urine in terms of excess taurine<sup>29</sup>. Another proposed mechanism includes the destruction of circulating lymphocytes post radiation exposure<sup>30</sup>.

The TCA cycle identified in our pathway analysis provides evidence of mitochondrial dysfunction post irradiation. The radiation induces oxidative stress slows down the mitochondrial TCA cycle to minimize the

generation of free radicals<sup>31</sup>. The reduced excretion of energy metabolites (citrate,  $\alpha$ -KG, succinate, and fumarate) further suggest perturbed energy metabolism and are indicative of the radiation induced oxidative stress.

The metabolites creatinine and creatine were also found to be increased post radiation exposure. One of the previous study has reported creatine as a biomarker for 6.5 Gy dose at 24 h post irradiation<sup>16</sup>. Additionally, creatine is also a biomarker of radiation exposure in non-human primates (*Macaca mulatta*)<sup>32</sup>. Creatinine is formed from creatine and creatine phosphate and increased excretion of both (creatinine and creatine) indicates incapability of irradiated muscle to use creatine<sup>33</sup>.

To best of our knowledge, this is the first study to recognize a set of NMR spectroscopy based biomarkers that can differentiate between control and irradiated group with such high sensitivity, specificity and accuracy based on logistic regression based prediction modeling in preclinical setup. In the present study, using a prediction model approach a proof of concept has been provided for whole body radiation exposure, however, this work will be further extended with a range of radiation doses.

## Conclusion

The present findings of our work have provided a strong foundation for the potential application of NMR based metabolomics approach for screening of triage during nuclear accidental scenario. However, limiting the human subjects with radiation exposure mimicking or close to real nuclear accident, the findings need to be scaled up to humans using advanced mathematical modeling and that needs extensive data set generated from a range of radiation dose exposure and would be a straight away expansion of the current study.

## Materials and methods

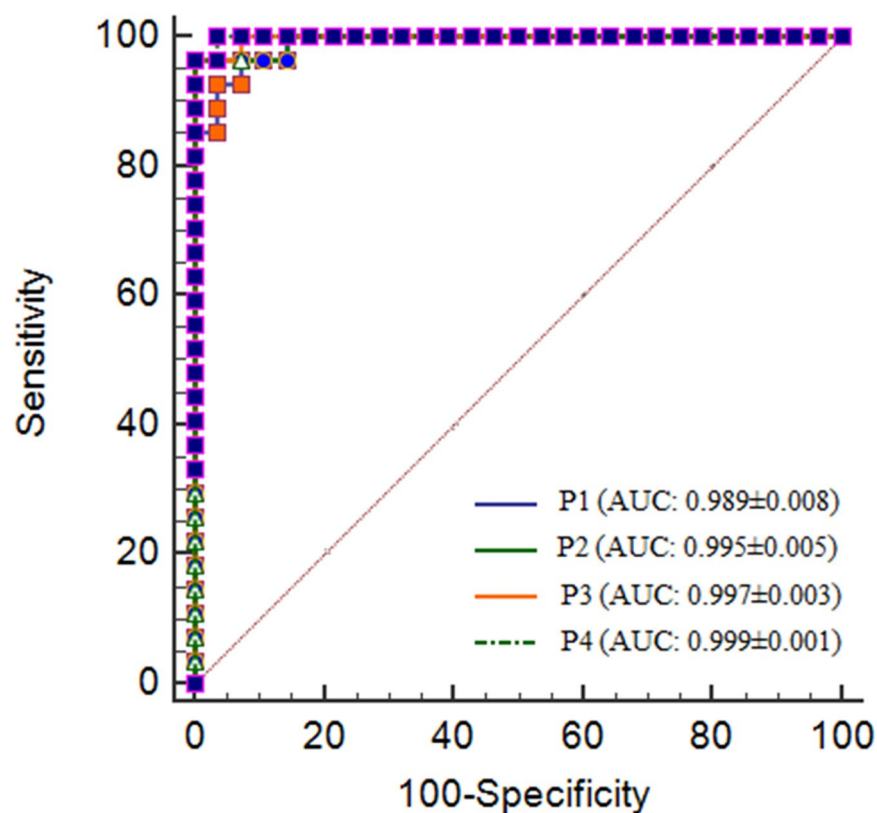
**Chemicals.** All chemicals, trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP), NMR solvents, and deuterium oxide (D<sub>2</sub>O) used during the study were obtained from Sigma-Aldrich (St. Louis, MO, USA).

**Animal handling and radiation exposure.** Sixty 'C57BL/6' male mice, 8–10 weeks of age and 25–35 g of body weight obtained from the animal facility of the institute were used for this study. Out of the 60 mice, sufficient urine was obtained only from 55 mice. Prior to group allocation and whole body radiation exposure, the animals were acclimatized in polypropylene cages for 48 h under standard room temperature of 19–23 °C, humidity in range of 45–65% and fluorescent lighting was provided for 12 h light/12 h dark cycle. Food and water were provided ad libitum. The animals after acclimatization were randomly allocated into two groups. One group with (n = 27) animals were exposed to 7.5 Gy whole body radiation dose through the Tele <sup>60</sup>Co gamma irradiation facility at our institute (Bhabhatron II, Panacea, India) with source operating at 1.568 Gy/min. Mice were exposed to 7.5 Gy of whole body radiation with a field of view of 30 × 30 cm<sup>2</sup> and surface to source distance (SSD) of 80 cm. A whole body radiation dose of 7.5 Gy in mice is approximately equivalent to 3.375 Gy of human and is considered to be associated with a moderate dose of radiation<sup>26,34,35</sup>. Rest of the animals (n = 28) were sham irradiated and served as a control group. None of the animals were anaesthetized during sham or radiation exposure. All experimental protocols and animal handling were approved and according to the strict guidelines of the institutional animal ethical committee at the institute of nuclear medicine and allied sciences (INMAS), DRDO, Delhi-110054 (8/GO/RBI/S/99/CPCSEA/INM/IAEC/2017/09).

**Sample collection.** Prior to urine collection, all the animals were kept in clean metabolic cages for 3 days for acclimatization. Urine samples were collected in ice-cooled tubes containing 1% sodium azide (n = 27) at 24 h for post irradiation group and (n = 28) for the control group. The criteria for choosing this time point was based on our previous findings that have shown maximum metabolic perturbations at 24 h after 8 Gy of WB radiation and also in line with the literature<sup>6,27</sup>. Urine samples were centrifuged and the supernatant of urine so obtained was stored at – 80°C for further NMR spectroscopic analysis.

**NMR analysis.** Three hundred fifty microliter of thawed and centrifuged urine was mixed with 250  $\mu$ l of 0.2 M deuterated phosphate buffer (pH 7.4) with 1 mM TSP and transferred to 5 mm NMR tubes. <sup>1</sup>H NMR spectra of urine samples were recorded at 298 K on a narrow bore NMR spectrometer operating at 700 MHz (Agilent, USA). 1D spectra was acquired with NOESY using a water presaturation pulse. The parameters used for the 1D experiment were as follows: dummy scan = 4, relaxation delay of 5 s, mixing time of 50 ms and acquisition time of 25 s was used. Sixty four FIDs were collected into 32 K data points over a spectral width of 6,410.25 Hz. The FID was weighted by an exponential function with a 0.3 Hz line broadening factor prior to Fourier transformation. The data was pre-processed using topspin software. Individual metabolites were identified using published literature using Human Metabolome Database (HMDB).

**Multivariate statistical data analysis.** The statistical expression of different metabolites in control and radiated group were analyzed using multivariate statistical data analysis. The multivariate analysis was performed in a total of 55 spectra of which 27 are from the radiated group and n = 28 from the control group. All spectra were phased and corrected manually for baseline using TOPSPIN 2.1 (Bruker, Germany). The corrected NMR spectra with spectral range of  $\delta$  0.5–9.5 were imported into AMIX (Bruker, Biospin, Germany) and all the spectra were segmented into 901 binned region of equal width of 0.01 ppm. Prior to any statistical analysis, regions for water (4.5–5.0 ppm) and urea (5–6.0 ppm) were eliminated and the data were pre-processed using scaling and normalization. The pre-processing is done to eliminate the possible bias that could arise either due to sample variability or handling or both. Normalization (by sum) was performed to minimize possible differences in concentration between samples. Following normalization, pareto-scaling (mean-centering and division by the



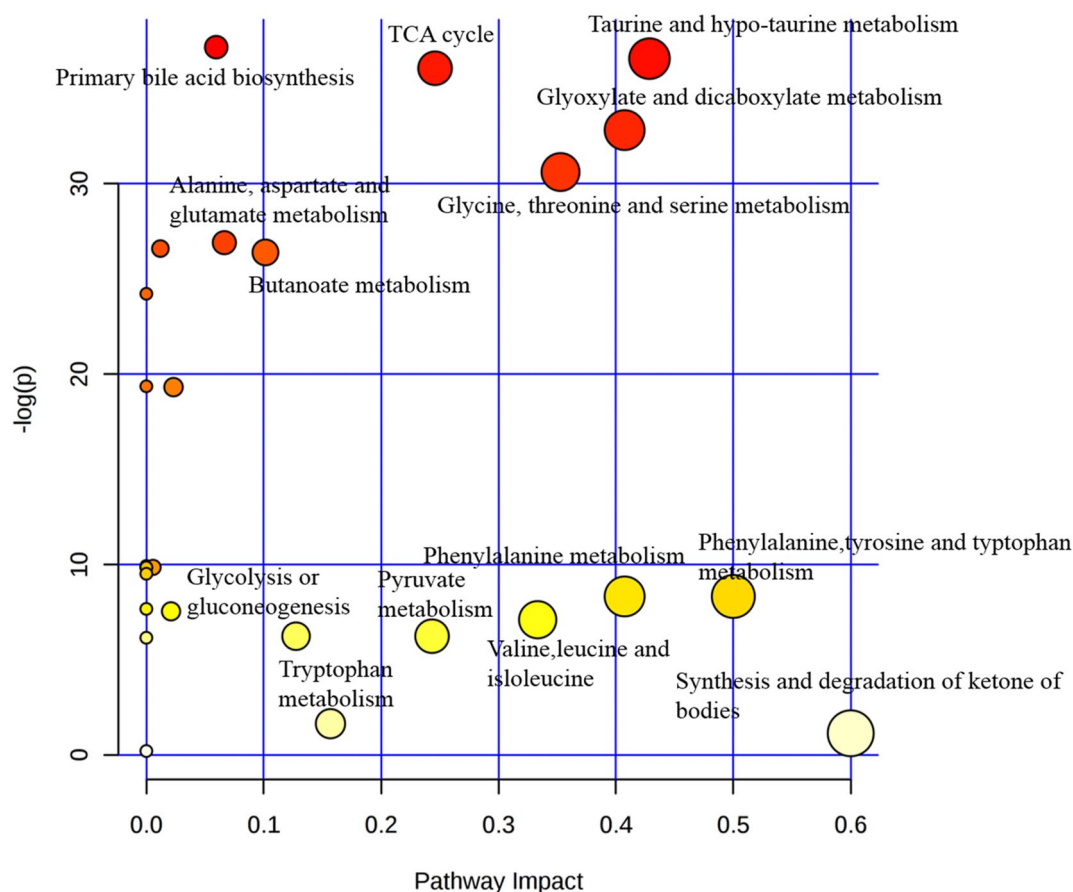
**Figure 2.** The ROC analysis results from the four prediction models calculated from the logistic regression analysis. The diagnostic performance of each biomarker model was assessed by the area under the ROC curve (AUC) and the determination of sensitivity and specificity at the optimal cut-off was determined using the Youden index (J). The optimized model was the P4 model with an AUC of 0.999 (95% CI 0.933–1.00).

square root of the standard deviation of each variable) was done to give all variables equal weight regardless of their absolute value. After data pre-processing, an unsupervised pattern recognition (PR) method PCA was done to detect systemic variation and intrinsic clusters within the data set.

Based on PCA, the spectral region (bins) differentiating between controls and radiated group were identified, integrated and the relative intensity of each of the identified metabolites was calculated. Relative intensities were then further used to perform the univariate and supervised multivariate analysis. A supervised PR method OPLS-DA was done to achieve maximum separation of control and irradiated group. OPLS-DA model was validated using CV ANOVA which provides  $p$  value that indicates the level of significance for group separation in OPLS analysis. A permutation test was used to validate the predictive capability of the computed OPLS-DA model. The most relevant potential markers were identified using the S plot that was constructed from the loading plots of OPLS-DA. Additionally, univariate analysis such as for fold change and correlation analysis was also carried out for the identified metabolites. To facilitate the interpretation of the findings, correlation coefficient of both groups were then calculated. In the present study, a cut-off value of  $|r| > 0.3$  ( $r > 0.3$  and  $r < -0.3$ ) was chosen for correlation coefficient as significant based on the discrimination significance ( $p < 0.05$ ). Finally, based on correlation coefficient, VIP and S plot, the specific metabolite responsible for differentiation of control and radiated group were identified.

**Selection of potential biomarker candidates.** The selected metabolites were then used to construct the prediction model and the accuracy of the model was evaluated using the ROC curve. A stepwise forward logistic regression model was constructed to design the best metabolite combination. ROC curve analysis was carried out to evaluate the combined model following the DeLong method. The importance and demonstration of metabolites as a biomarker model were assessed using sensitivity, specificity, and AUC, at the optimum cut off defined by the Youden index (maximum vertical distance).

**Multivariate power analysis calculation.** The minimum number of animals required to attain statistical significance between the radiated and control groups was determined by the power analysis option of Metaboanalyst 3.0. The input data for sample size estimation consisted of 24 key metabolite variables used for OPLS-DA analysis. With an FDR-adjusted  $p$  value of 0.004, the sample size of 3, 10, 16, 24, 40 and 60 per group were calcu-



**Figure 3.** Metabolic pathway mapping of the impacted metabolic network identified between control and irradiated group. The  $\chi$ -axis represents the pathway impact, and the y-axis represents the  $-\log(p)$ .

lated with predicted test powers of 0.89, 1.00, 1.00, 1.00 and 1.00 (supplementary Fig. 5), and this clearly verify that the sample sizes selected for this study were indeed more than sufficient.

**Metabolic pathway analysis.** To appreciate the inherent and complex network properties of various metabolites, detailed analysis of the most significant metabolic pathways between irradiated and control groups was performed by Metaboanalyst 3.0. For metabolic pathway analysis, Metaboanalyst software makes use of a high-quality KEGG database. The software also uses algorithms and concepts that include Global Test and Global Ancova, as well as pathway topology analysis. The  $p$  value and the pathway impact value are calculated using enrichment and pathway topology analysis respectively. The most impacted metabolic pathway was set as pathway impact  $> 0.05$  and  $-\log(p) > 35$ .

**Statistics.** Multivariate analysis of NMR data and pathway analysis were performed using SIMCA-P version 14.0 (Umetrics AB, Umea, Sweden) and Metaboanalyst 3.0 ([www.metaboanalyst.ca](http://www.metaboanalyst.ca)) respectively. The ROC curve for prediction modelling was performed using MedCalc software version 19.0.7 (Broekstraat, Mariakerke, Belgium). The measurement data are expressed as the mean  $\pm$  SE (standard error). For univariate analysis, the differences between the two groups were analyzed using Student's  $t$  test. The level of significance was set at  $p < 0.05$ .

Received: 6 January 2020; Accepted: 13 August 2020  
Published online: 30 September 2020

## References

- Graham, B. et al. *The Report of the Commission on the Prevention of WMD Proliferation and Terrorism* (Vintage Books, New York, 2008). World at Risk.
- Blakely, W. F., Salter, C. A. & Prasanna, P. G. Early-response biological dosimetry-recommended countermeasure enhancements for mass-casualty radiological incidents and terrorism. *Health Phys.* **89**, 494–504 (2005).
- Pellmar, T. C. & Rockwell, S. Priority list of research areas for radiological nuclear threat countermeasures. *Radiat. Res.* **163**, 115–123 (2005).

4. Menon, S. S. *et al.* Radiation metabolomics: current status and future directions. *Front. Oncol.* **6**, 1–10 (2016).
5. Riecke, A., Ruf, C. G. & Meineke, V. Assessment of radiation damage—the need for a multiparametric and integrative approach with the help of both clinical and biological dosimetry. *Health Phys.* **98**, 160–167 (2010).
6. Pannkuk, E. L., Fornace, A. J. Jr. & Laiakis, E. C. Metabolomic applications in radiation biodosimetry: exploring radiation effects through small molecules. *Int. J. Radiat. Biol.* **93**, 1151–1176 (2017).
7. Jones, J. W. *et al.* Identification and quantitation of biomarkers for radiation-induced injury via mass spectrometry. *Health Phys.* **106**, 106–119 (2014).
8. Laiakis, E. C. *et al.* Salivary metabolomics of total body irradiated nonhuman primates reveals long-term normal tissue responses to radiation. *Int. J. Radiat. Oncol. Biol. Phys.* **105**, 843–851 (2019).
9. Pannkuk, E. L. *et al.* Temporal effects on radiation responses in nonhuman primates: identification of biofluid small molecule signatures by gas chromatography–mass spectrometry metabolomics. *Metabolites* **9**, 1–14 (2019).
10. Pannkuk, E. L. *et al.* Liquid chromatography–mass spectrometry-based metabolomics of nonhuman primates after 4 Gy total body radiation exposure: global effects and targeted panels. *J. Proteome Res.* **18**, 2260–2269 (2019).
11. Pannkuk, E. L., Laiakis, E. C., Garcia, M., Fornace, A. J. Jr. & Singh, V. K. Nonhuman primates with acute radiation syndrome: results from a global serum metabolomics study after 7.2 Gy total-body irradiation. *Radiat. Res.* **190**, 576–583 (2018).
12. Pannkuk, E. L., Laiakis, E. C., Fornace, A. J. Jr., Fatanmi, O. O. & Singh, V. K. A metabolomic serum signature from nonhuman primates treated with a radiation countermeasure, gamma-tocotrienol, and exposed to ionizing radiation. *Health Phys.* **115**, 3–11 (2018).
13. Chen, Z. *et al.* Differential mobility spectrometry–mass spectrometry (DMS–MS) in radiation biodosimetry: rapid and high-throughput quantitation of multiple radiation biomarkers in nonhuman primate urine. *J. Am. Soc. Mass Spectrom.* **29**, 1650–1664 (2018).
14. Kultova, G., Tichy, A., Rehulkova, H. & Myslivcova-Fucikova, A. The hunt for radiation biomarkers: current situation. *Int. J. Radiat. Biol.* **96**, 370–382 (2020).
15. Johnson, C. H. *et al.* Radiation metabolomics. 4. UPLC-ESI-QTOFMS-based metabolomics for urinary biomarker discovery in gamma-irradiated rats. *Radiat. Res.* **175**, 473–484 (2011).
16. Johnson, C. H. *et al.* Radiation metabolomics. 5. Identification of urinary biomarkers of ionizing radiation exposure in nonhuman primates by mass spectrometry-based metabolomics. *Radiat. Res.* **178**, 328–340 (2012).
17. Xia, J., Broadhurst, D. I., Wilson, M. & Wishart, D. S. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* **9**, 280–299 (2013).
18. Khan, A. R. *et al.* NMR spectroscopy based metabolic profiling of urine and serum for investigation of physiological perturbations during radiation sickness. *Metabolomics* **7**, 583–592 (2011).
19. Triba, M. N. *et al.* PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross validation quality parameters. *Mol. Biosyst.* **11**, 13–19 (2015).
20. Wheelock, A.M. & Wheelock, C. E. Trials and tribulations of ‘omics data analysis: assessing quality of SIMCA-based multivariate models using examples from pulmonary medicine. *Mol. Biosyst.* **9**, 2589–2596 (2013).
21. Lanz, C. *et al.* Radiation metabolomics. 3. Biomarker discovery in the urine of gamma-irradiated rats using a simplified metabolomics protocol of gas chromatography–mass spectrometry combined with random forests machine learning algorithm. *Radiat. Res.* **172**, 198–212 (2009).
22. Rana, P. *et al.* NMR based metabolomics reveals acute hippocampal metabolic fluctuations during cranial irradiation in murine model. *Neurochem. Int.* **74**, 1–7 (2014).
23. Xiao, X., Hu, M., Liu, M. & Hu, J. Z. <sup>1</sup>H NMR metabolomics study of Spleen from C57BL/6 Mice exposed to gamma radiation. *Metabolomics (Los Angel)* **6**, 1–11 (2016).
24. Xiao, X., Hu, M., Zhang, X. & Hu, J. Z. NMR-based metabolomics analysis of liver from C57BL/6 mouse exposed to ionizing radiation. *Radiat. Res.* **188**, 44–55 (2017).
25. Chen, C., Brenner, D. J. & Brown, T. R. Identification of urinary biomarkers from X-irradiated mice using NMR spectroscopy. *Radiat. Res.* **175**, 622–630 (2011).
26. Coy, S. L. *et al.* Radiation metabolomics and its potential in biodosimetry. *Int. J. Radiat. Biol.* **87**, 802–823 (2011).
27. Tyburski, J. B. *et al.* Radiation metabolomics. 1. Identification of minimally invasive urine biomarkers for gamma-radiation exposure in mice. *Radiat. Res.* **170**, 1–14 (2008).
28. Chesney, R. W. Taurine: its biological role and clinical implications. *Adv. Pediatr.* **32**, 1–42 (1985).
29. Christophersen, O. A. Radiation protection following nuclear power accidents: a survey of putative mechanisms involved in the radioprotective actions of taurine during and after radiation exposure. *Microb. Ecol. Health Dis.* **1**, 23 (2012).
30. Dilley, J. V. The origin of urinary taurine excretion during chronic radiation injury. *Radiat. Res.* **50**, 191–196 (1972).
31. Singh, K. K. (ed.) *Oxidative Stress, Disease and Cancer* 825–850 (Imperial College Press, London, 2006).
32. Krise, G. M., Williams, C. M. & Anderson, D. R. Post irradiation creatinuria in *Macaca mulatta*. *Proc. Soc. Exp. Biol. Med.* **95**, 764–766 (1957).
33. Anderson, D. R., Joseph, B. J., Krise, G. M. & Williams, C. M. Effect of radiation on the metabolism of creatine. *Am. J. Physiol.* **192**, 247–252 (1958).
34. Hall, E. J. & Giaccia, A. J. *Radiobiology for the Radiologist*. 6 (Lippincott Williams & Wilkins, Philadelphia, 2006).
35. Coleman, C. N. *et al.* Molecular and cellular biology of moderate-dose (1–10 Gy) radiation and potential mechanisms of radiation protection: report of a workshop at Bethesda, Maryland, December 17–18, 2001. *Radiat. Res.* **159**, 812–834 (2003).

## Acknowledgements

The authors would like to acknowledge Dr. NR Jagannathan and Dr. Uma Sharma, Department of NMR, All India Institute of Medical Sciences (AIIMS) for their support and assistance to carry out sample analysis at 700 MHz NMR Spectrometer (Agilent, USA). Authors would like to acknowledge Defence Research and Development Organisation for providing financial assistance for the Project (DRDO/TD/INM-313) and university grant commission (UGC), New Delhi, India for awarding junior research fellowship to Kiran Maan.

## Author contributions

P.R. and S.K. conceived the project and designed the study, P.R. supervised NMR the experiments and data analysis. R.T. and K.M. involved in experimentation, analysis, and interpretation of data and writing of the manuscript. P.R. and S.K. evaluated the manuscript critically and all the authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.



### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-72426-4>.

**Correspondence** and requests for materials should be addressed to P.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**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 Author(s) 2020