Supplementary Information

Characterizing Metabolic Alteration in Early-Stage Chronic Kidney Disease for Diagnostic Insight.

Upasna Gupta[1,2], Amrita Sahu[1,2], Dharmendra Singh Bhadauria[3]*, Bikash Baishya[1,2], Neeraj Sinha[1, 2]*.

[1]Centre of Biomedical Research, Sanjay Gandhi Post Institute of Medical Sciences Campus, Raebareli Rd, Lucknow, Uttar Pradesh, 226014, India.
[2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad- 201002, India.
[3]Department of Nephrology and Renal Transplantation, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, 226014, India.

***Corresponding author:** neerajcbmr@gmail.com; neeraj.sinha@cbmr.res.in (Neeraj Sinha)
docdharm10@gmail.com (Dharmendra Singh Bhadauria)

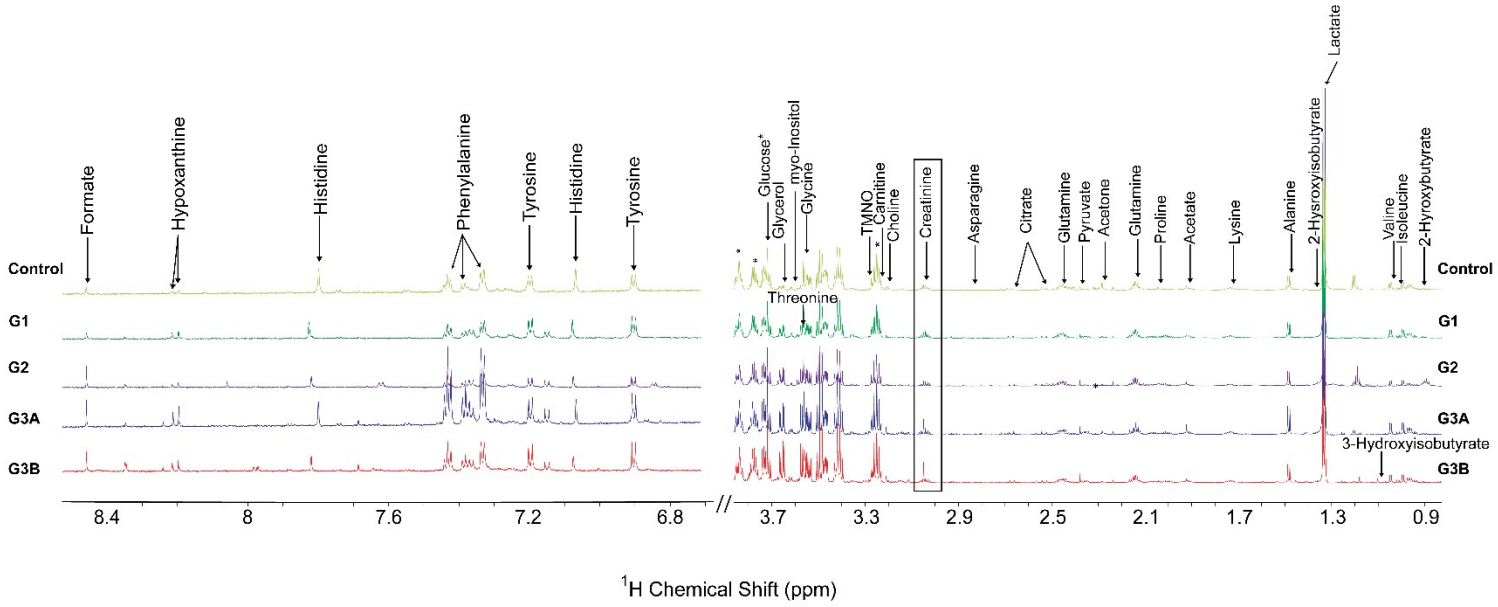

Figure S1: A cumulative 1D ^{1}H CPMG NMR spectra stack plot, covering δ0.9–δ8.4 ppm, was produced using serum samples of early-stage CKD G1 (dark green), G2 (violet), G3A (dark blue), and G3B (red). The control group was represented by light green. The associated metabolic assignments are labelled on the spectral peaks.

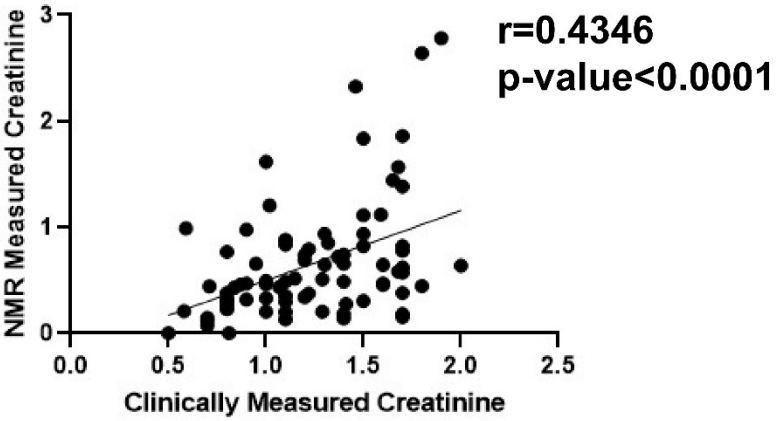

Figure S2: Correlation analysis between NMR-measured creatinine and clinically measured Creatinine.

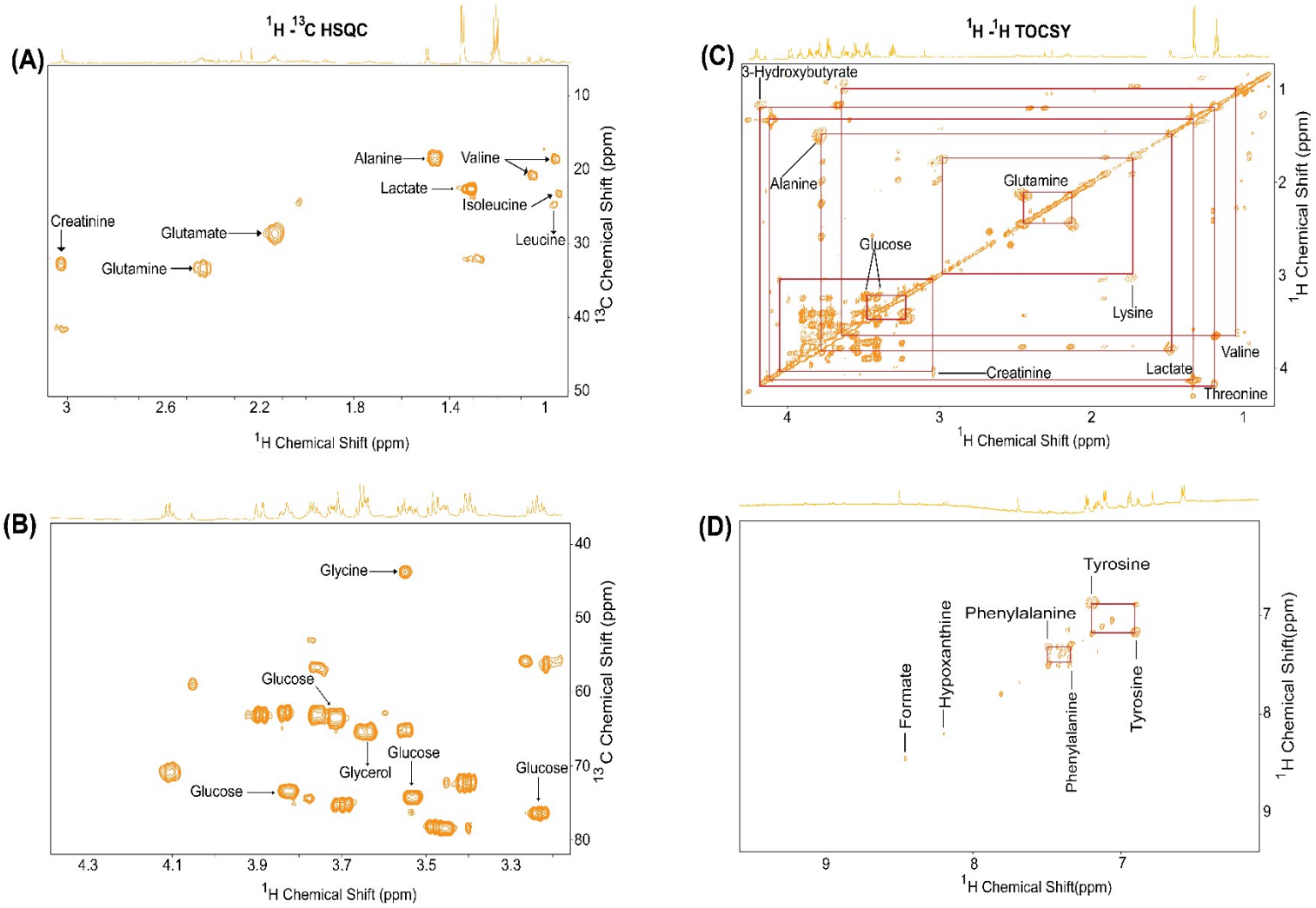

Figure S3: The two-dimensional spectra of Heteronuclear Single Quantum Correlation (HSQC) (**A** and **B**) and Total Proton-Proton Correlation Spectroscopy (TOCSY) (**C** and **D**) show the assigned metabolites in panels a–d.

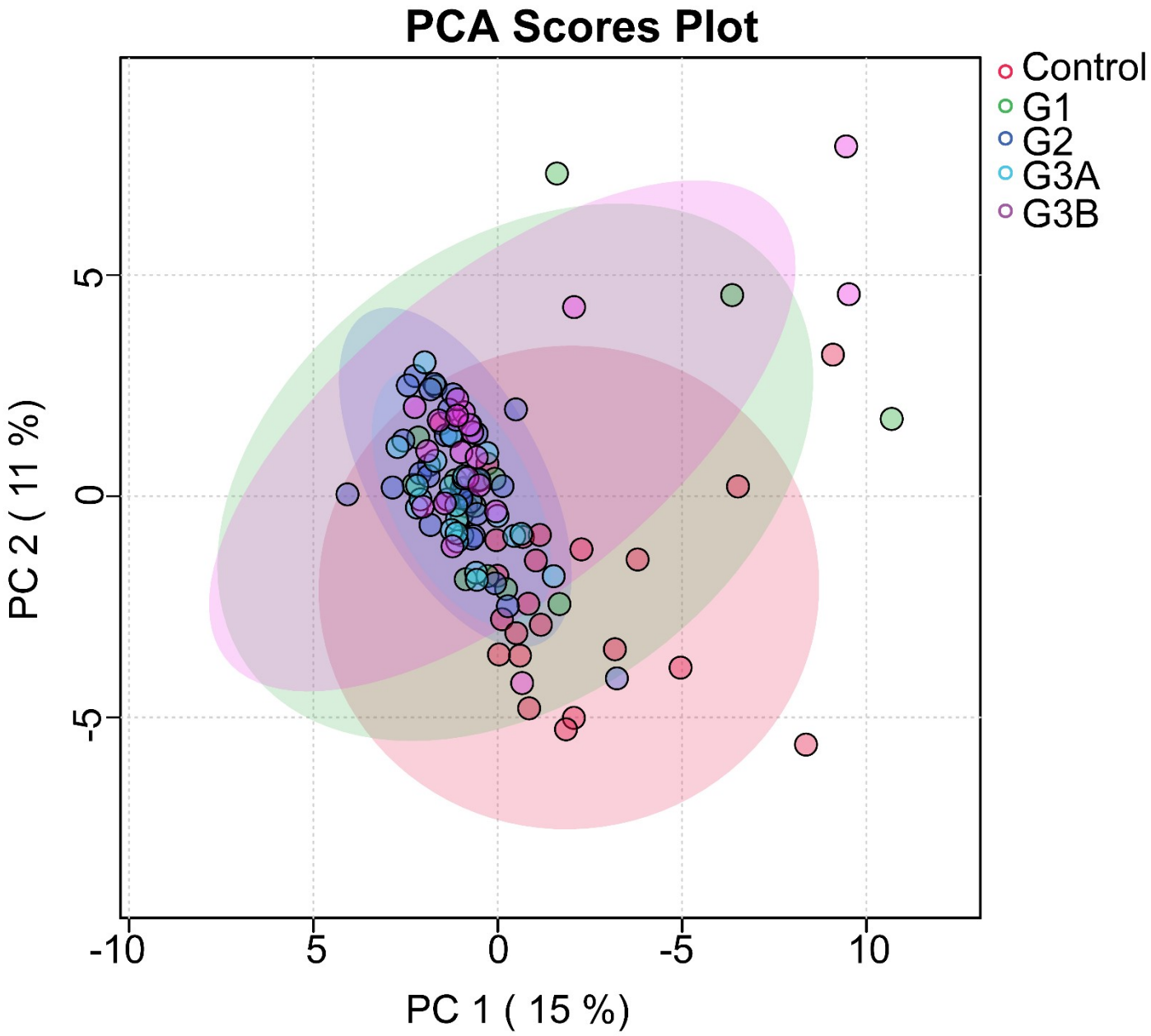

Figure S4: Two-dimensional score plot of principal component analysis (PCA) was performed on serum samples containing early-stage CKD patients G1 (green), G2 (blue), G3A (light blue), G3B (violet), and control subjects (red) utilizing ^{1}H CPMG spectra.

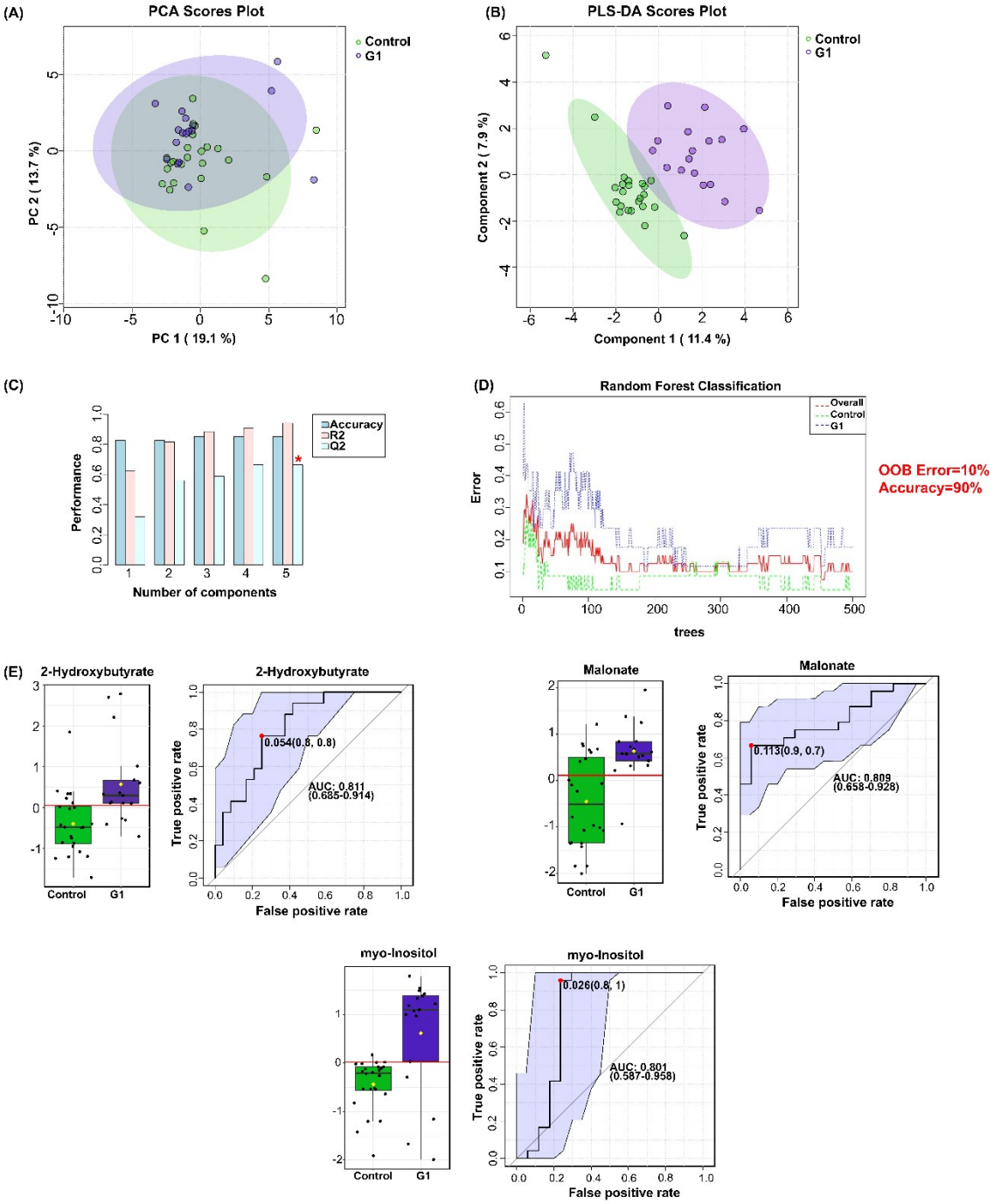

Figure S5: (A) Two-dimensional score plot of principal component analysis (PCA) utilizing ^{1}H CPMG spectra; Control is represented by green, and those in the G1 stage by violet. **(B)** Two-dimensional score plot of the partial least squares discriminant analysis (PLS-DA) utilizing the ^{1}H CPMG spectra. **(C)** Three measures were used to assess the PLS-DA classification performance: prediction accuracy, multiple correlation coefficient (R2), and explained variance in prediction (Q2). An asterisk denotes the model that performs the best. **(D)** The biomarker analysis's major metabolites are displayed with altered metabolites and area under the curve (AUC) plot between the Control and G1 stage of CKD.

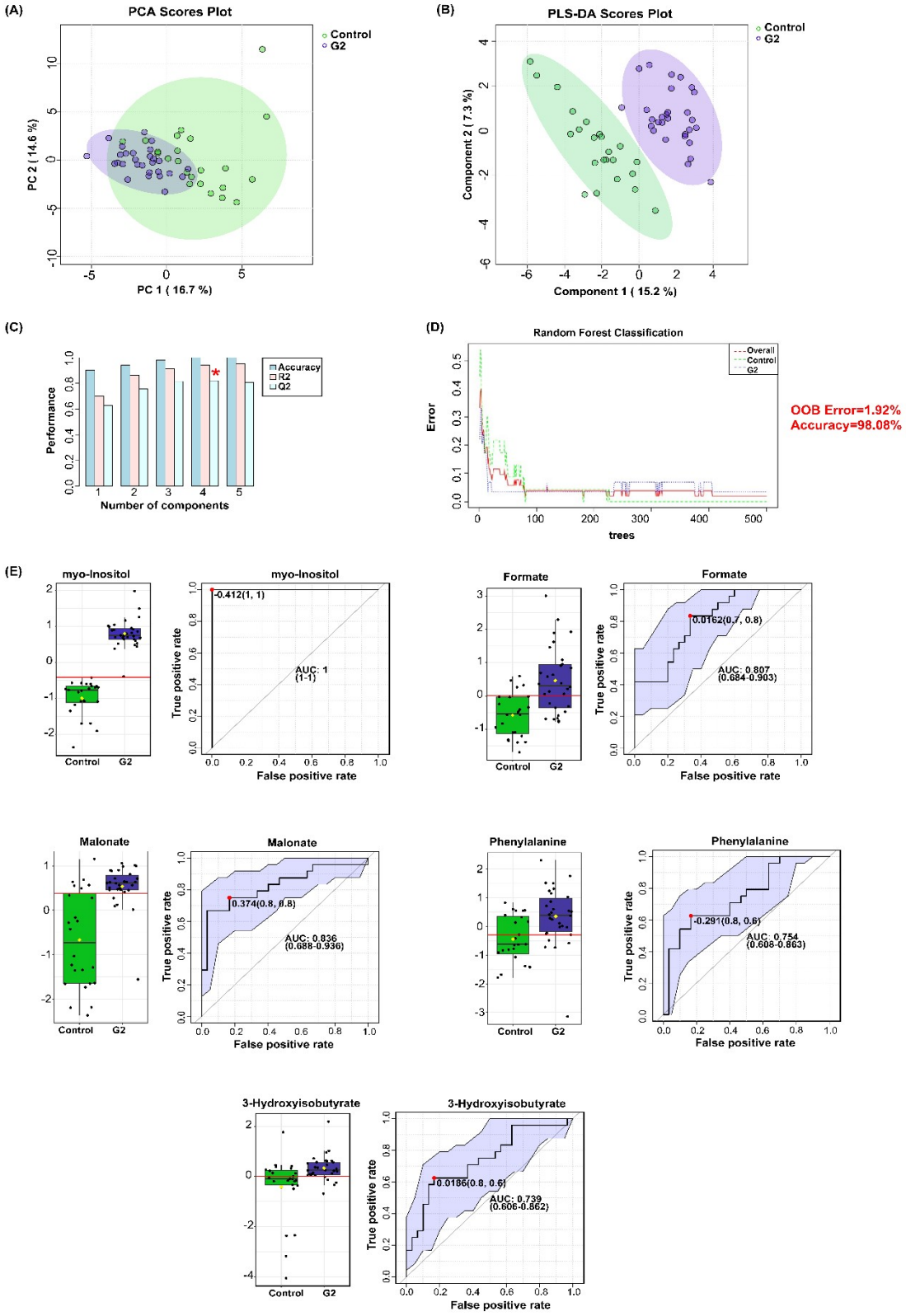

Figure S6: (A) Two-dimensional score plot of principal component analysis (PCA) utilizing ^{1}H CPMG spectra; Control is represented by green, and those in the G2 stage by violet. **(B)**

Two-dimensional score plot of the partial least squares discriminant analysis (PLS-DA) utilizing the ^{1}H CPMG spectra. **(C)** Three measures were used to assess the PLS-DA classification performance: prediction accuracy, multiple correlation coefficient (R2), and explained variance in prediction (Q2). An asterisk denotes the model that performs the best. **(D)** The biomarker analysis's major metabolites are displayed with altered metabolites and area under the curve (AUC) plot between the Control and G2 stage of CKD.

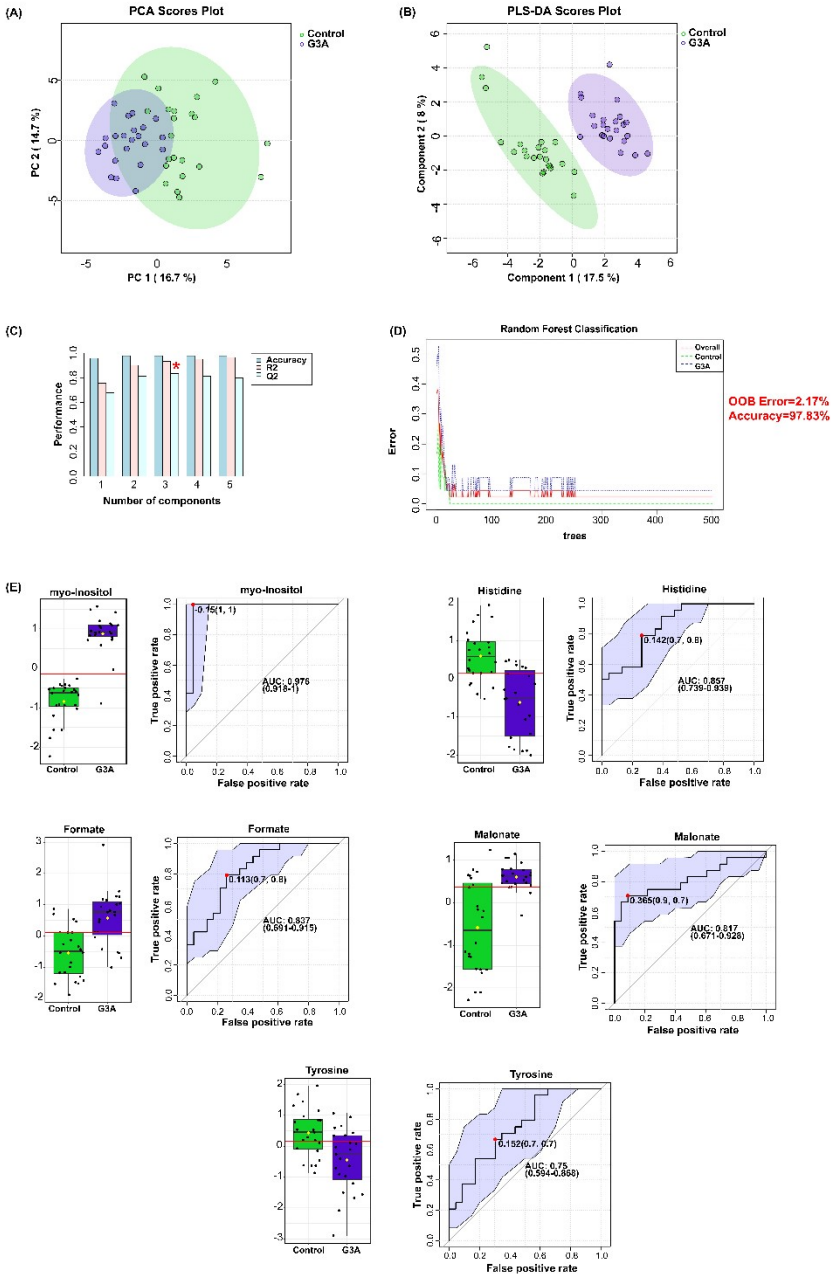

Figure S7: (A) Two-dimensional score plot of principal component analysis (PCA) utilizing ^{1}H CPMG spectra; Control is represented by green, and those in the G3A stage by violet. **(B)** Two-dimensional score plot of the partial least squares discriminant analysis (PLS-DA) utilizing the ^{1}H CPMG spectra. **(C)** Three measures were used to assess the PLS-DA classification performance: prediction accuracy, multiple correlation coefficient (R2), and explained variance in prediction (Q2). An asterisk denotes the model that performs the best. **(D)** The biomarker analysis's major metabolites are displayed with altered metabolites and area under the curve (AUC) plot between the Control and G3A stage of CKD.

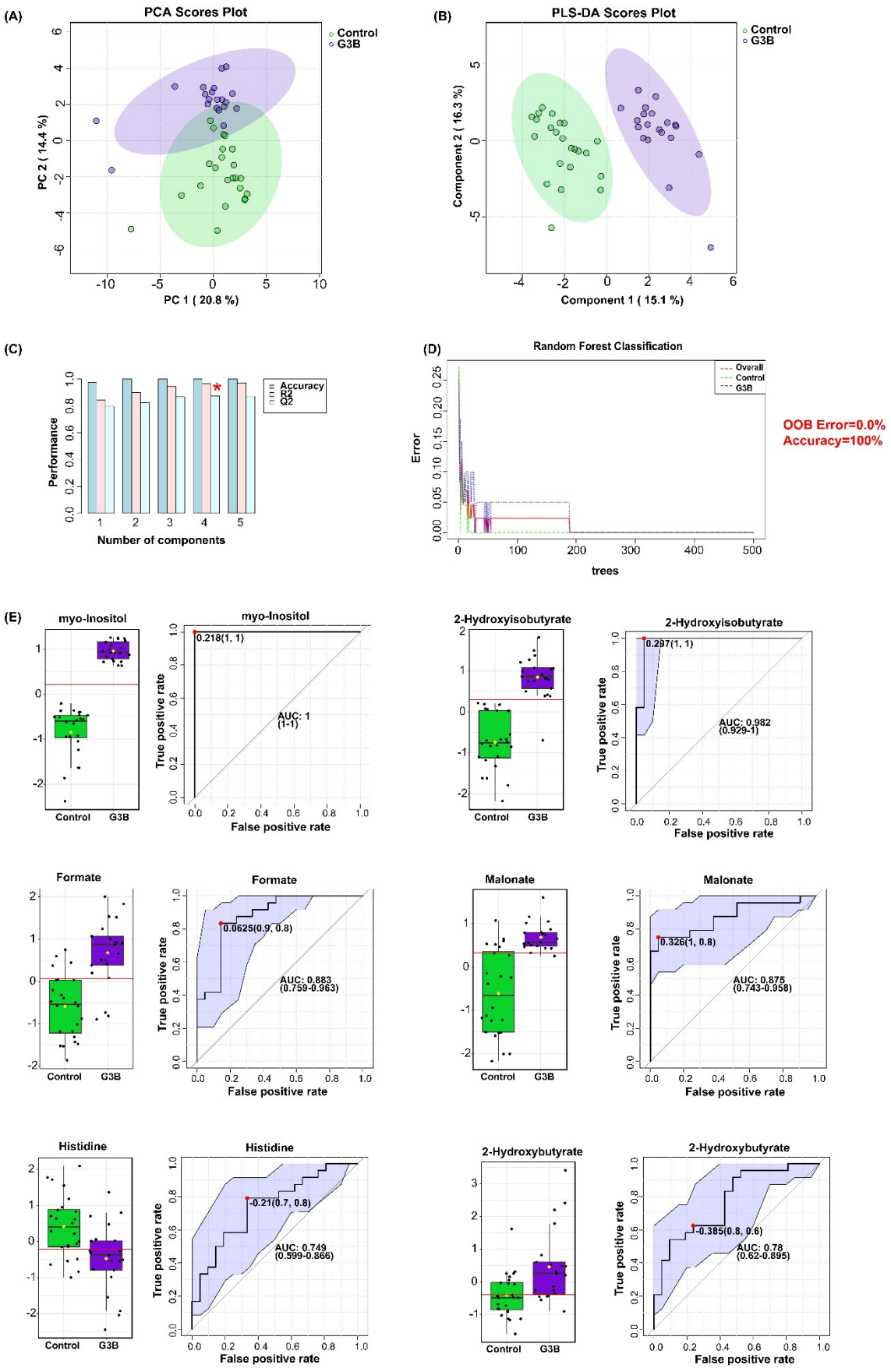

Figure S8: (A) Two-dimensional score plot of principal component analysis (PCA) utilizing

^{1}H CPMG spectra; Control is represented by green, and those in the G3B stage by violet. **(B)** Two-dimensional score plot of the partial least squares discriminant analysis (PLS-DA) utilizing the ^{1}H CPMG spectra. **(C)** Three measures were used to assess the PLS-DA classification performance: prediction accuracy, multiple correlation coefficient (R2), and explained variance in prediction (Q2). An asterisk denotes the model that performs the best. **(D)** The biomarker analysis's major metabolites are displayed with altered metabolites and area under the curve (AUC) plot between the Control and G3B stage of CKD.

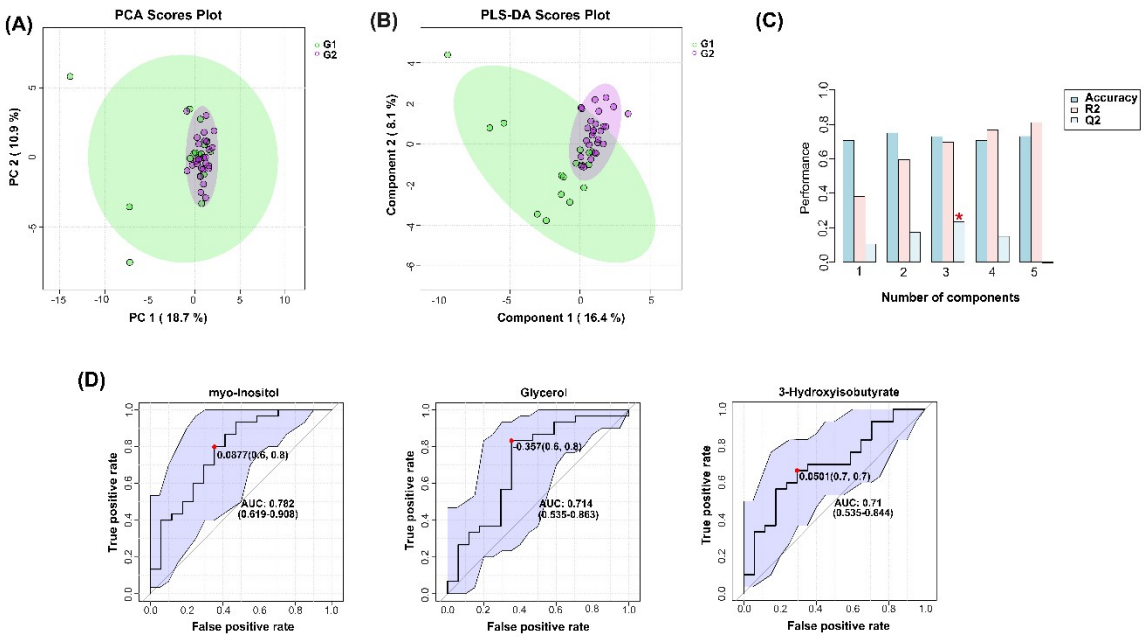

Figure S9: (A) Two-dimensional score plot of principal component analysis (PCA) utilizing ^{1}H CPMG spectra; patients in the G1 stage of CKD are represented by green, and those in the G2 stage by violet. **(B)** Two-dimensional score plot of the partial least squares discriminant analysis (PLS-DA) utilizing the ^{1}H CPMG spectra; patients in the G1 stage of CKD are shown by green and those in the G2 stage by violet. **(C)** Three measures were used to assess the PLS-DA classification performance: prediction accuracy, multiple correlation coefficient (R2), and explained variance in prediction (Q2). An asterisk denotes the model that performs the best. **(D)** The biomarker analysis's major metabolites are displayed in the area under the curve (AUC) plot between the G1 and G2 stages of patients with CKD.

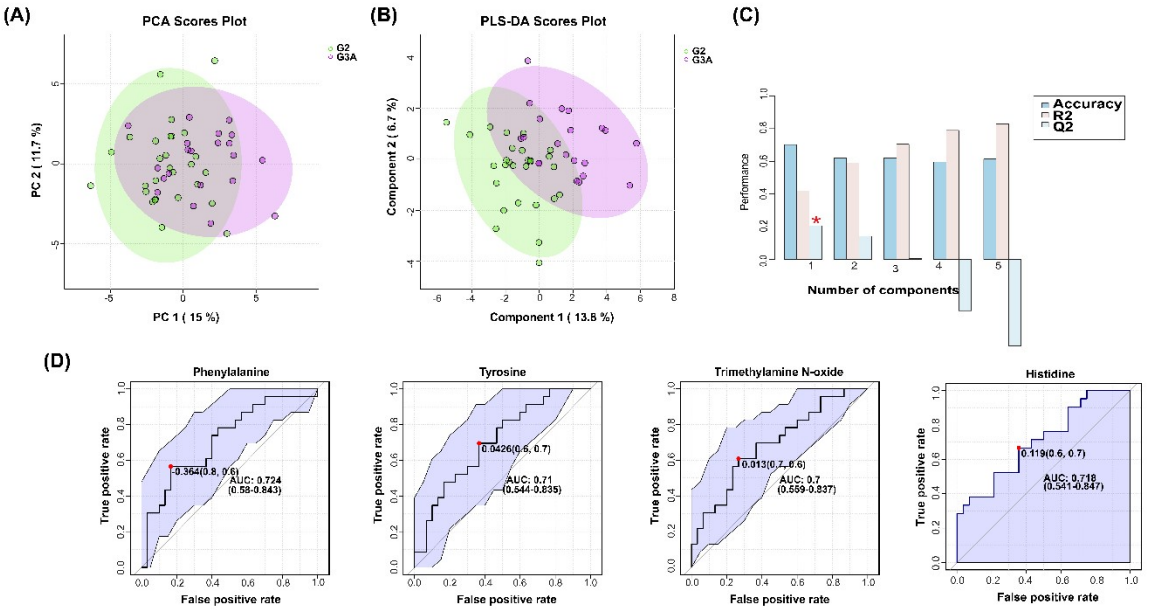

Figure S10: (A) Principal component analysis (PCA) two-dimensional score plot using ^{1}H CPMG spectra; patients with CKD in the G2 stage are shown as green, and those in the G3A stage as violet. **(B)** Two-dimensional score plot of partial least squares discriminant analysis (PLS-DA) using ^{1}H CPMG spectra; green represents G2 stage CKD patients and violet represents G3A stage patients. **(C)** The performance of the PLS-DA classification was evaluated using three metrics: explained variance in prediction (Q2), multiple correlation coefficient (R2), and prediction accuracy. The best-performing model has an asterisk above it. **(D)** The area under the curve (AUC) plot between the G2 and G3A stages of patients' CKD shows the main metabolites identified by the biomarker analysis.

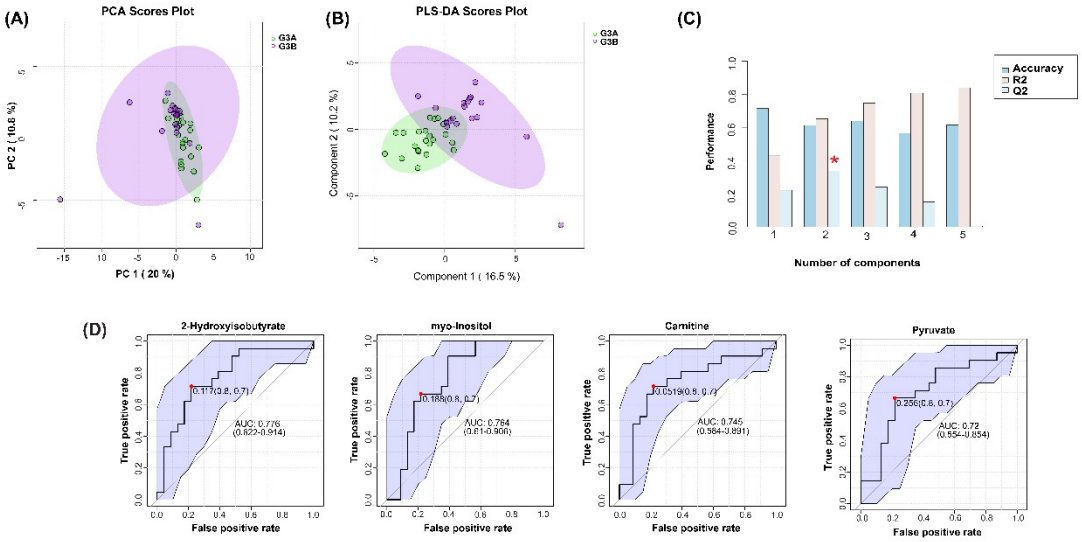

Figure S11: (A) Two-dimensional score plot of principal component analysis (PCA) using ^{1}H CPMG spectra; green indicates patients with G3A CKD and violet indicates those with G3B CKD. **(B)** A two-dimensional score plot of the partial least squares discriminant analysis (PLS-DA) using the ^{1}H CPMG spectra; green represents patients with CKD in the G3A stage and violet represents patients in the G3B stage. **(C)** Prediction accuracy, multiple correlation coefficient (R2), and explained variance in prediction (Q2) were the three metrics utilized to evaluate the PLS-DA classification performance. The best-performing model is indicated by an asterisk. **(D)** The key metabolites identified by the biomarker analysis are shown in the area under the curve (AUC) plot between the G3A and G3B stages of patients' CKD

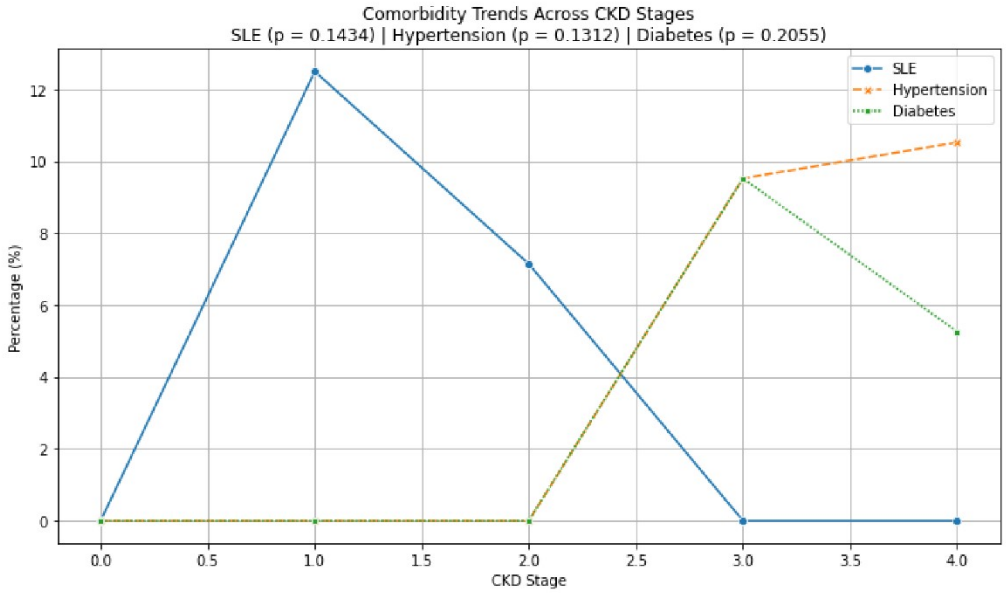

Figure S12 : Trend analysis of comorbidity distribution across CKD stages.

Table S1: Serum metabolic entities exhibiting statistically significant difference between the study groups evaluated using ANOVA statistics based on Fisher's least significant difference (LSD, a commonly used post-hoc test). The analysis has been performed using ANOVA Statistical analysis module of Metaboanalyst 6.0.

Metabolic Entity	f.value	p.value	-log[p]	FDR	Fisher's LSD
myo-Inositol	47.711	3.33E-23	22.477	1.43E-21	G1 - Control; G2 - Control; G3A - Control; G3B - Control; G2 - G1; G3A - G1; G3B - G1
2-Hydroxyisobutyrate	25.806	4.28E-15	14.368	9.21E-14	G1 - Control; G2 - Control; G3A - Control; G3B - Control; G3B - G2
Malonate	18.469	1.23E-11	10.909	1.77E-10	G1 - Control; G2 - Control; G3A - Control; G3B - Control
Formate	7.4606	2.35E-05	4.6285	0.000253	G1 - Control; G2 - Control; G3A - Control; G3B - Control
Creatinine	6.4532	0.000106	3.9762	0.000908	G2 - Control; G3A - Control; G3B - Control; G3B - G1; G3B - G2
Histidine	5.5104	0.000442	3.3547	0.003167	Control - G1; Control - G2; Control - G3A; Control - G3B; G1 - G3A; G2 - G3A
3-Hydroxybutyrate	3.7255	0.006981	2.1561	0.036938	G1 - Control; G2 - Control; G3A - Control
Tyrosine	3.634	0.008049	2.0943	0.036938	Control - G3A; G2 - G3A; G2 - G3B
2-Hydroxybutyrate	3.4917	0.010044	1.9981	0.036938	G1 - Control; G3B - Control; G3B - G3A
3-Hydroxyisobutyrate	3.4025	0.01154	1.9378	0.036938	G2 - Control; G3A - Control; G3B - Control; G2 - G1

Table S2: Control vs G1, G2, G3A and G3B, PLS-DA cross-validation values assessed by accuracy, R2, and Q2 with the 2nd component best classifying the model.

Measures	**1 comps**	**2 comps**	**3 comps**	**4 comps**	**5 comps**
Accuracy	0.44334	0.46142	0.47021	0.5043	0.49481
R2	0.57167	0.67708	0.70633	0.72463	0.73478
Q2	0.47851	0.57922	0.57585	0.49515	0.35078

Table S3: Control vs G1, PLS-DA cross-validation values assessed by accuracy, R2, and Q2 with the **5th** component best classifying the model**.**

Measure	**1 comps**	**2 comps**	**3 comps**	**4 comps**	**5 comps**
Accuracy	0.82698	0.82698	0.85198	0.85198	0.85198
R2	0.62144	0.81563	0.88361	0.90872	0.93983
Q2	0.32104	0.55997	0.58936	0.66351	0.66399

Table S4: Control vs G2, PLS-DA cross-validation values assessed by accuracy, R2, and Q2 with the 4th component best classifying the model.

Measure	**1 comps**	**2 comps**	**3 comps**	**4 comps**	**5 comps**
Accuracy	0.90182	0.94	0.98	1.0	1.0
R2	0.70076	0.86289	0.9125	0.9403	0.95399
Q2	0.6265	0.7579	0.81512	0.82045	0.80772

Table S5: Control vs G3A, PLS-DA cross-validation values assessed by accuracy, R2, and Q2 with the 3rd component best classifying the model.

Measure	**1 comps**	**2 comps**	**3 comps**	**4 comps**	**5 comps**
Accuracy	0.95778	0.97778	0.97778	0.97778	0.97778
R2	0.75691	0.90206	0.9324	0.95158	0.96429
Q2	0.67728	0.81331	0.83296	0.81333	0.79811

Table S6: Control vs G3B, PLS-DA cross-validation values assessed by accuracy, R2, and Q2 with the 4th component best classifying the model.

Measure	**1 comps**	**2 comps**	**3 comps**	**4 comps**	**5 comps**
Accuracy	0.975	1.0	1.0	1.0	1.0
R2	0.84352	0.89891	0.94567	0.96302	0.96919
Q2	0.79515	0.82308	0.86641	0.8747	0.87072

Table S7: G1 and G2, PLS-DA cross-validation values assessed by accuracy, R2, and Q2 with the 3rd component best classifying the model.

Measures	**1 comps**	**2 comps**	**3 comps**	**4 comps**	**5 comps**
Accuracy	0.70556	0.75	0.72778	0.70556	0.73056
R2	0.38008	0.59651	0.69617	0.76632	0.81181
Q2	0.10341	0.17406	0.23483	0.1502	0.00079

Table S8: G2 vs G3A, PLS-DA cross-validation values assessed by accuracy, R2, and Q2 with the 1st component best classifying the model.

Measures	**1 comps**	**2 comps**	**3 comps**	**4 comps**	**5 comps**
Accuracy	0.7002	0.56909	0.57131	0.59313	0.59313
R2	0.41885	0.59078	0.70434	0.78938	0.8282
Q2	0.21117	0.12975	-0.00022	-0.27236	-0.38156

Table S9: G3A and G3B, PLS-DA cross-validation values assessed by accuracy, R2, and Q2 with the 2nd component best classifying the model.

Measures	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	0.7131	0.6123	0.6373	0.56429	0.61429
R2	0.43157	0.65116	0.74697	0.80734	0.83731
Q2	0.22112	0.33691	0.24051	0.15021	0.0056424

Table S10: Key biological pathways that may be connected to the advancement of the disease based on metabolites found to be involved in altered metabolic cycles in individuals with early-stage CKD.

S.No.	Pathway Name	Impact Factor*
1.	Phenylalanine, tyrosine, and tryptophan biosynthesis	1.0
2.	Phenylalanine metabolism	0.357
3.	Citrate cycle (TCA cycle)	0.136
4.	Tyrosine metabolism	0.139
5.	Histidine metabolism	0.221
6.	Pyruvate metabolism	0.191
7.	Inositol phosphate metabolism	0.129

CKD- chronic kidney disease

*A pathway impact score of ≥0.1 was assigned to the metabolic pathways that were found to be relevant in early-stage CKD.