Insights into serum metabolic biomarkers for early detection of incident diabetic kidney disease in Chinese patients with type 2 diabetes by random forest

Jian-Jun Jiang[1,*], Tung-Ting Sham[2,*], Xiu-Fen Gu[1,*], Chi-On Chan[2,3], Nai-Ping Dong[2], Wei-Han Lim[1], Gao-Feng Song[1], Shun-Min Li[1], Daniel Kam-Wah Mok[2,3], Na Ge[1]

[1]Department of Nephrology, Shenzhen Traditional Chinese Medicine Hospital, The Fourth Clinical Medical College of Guangzhou University of Chinese Medicine, Shenzhen, China
[2]The Research Centre for Chinese Medicine Innovation and Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hong Kong, China
[3]State Key Laboratory of Chinese Medicine and Molecular Pharmacology (Incubation), Shenzhen, China
*Equal contribution

Correspondence to: Na Ge, Daniel Kam-Wah Mok; **email:** gena.sz@gzucm.edu.cn, daniel.mok@polyu.edu.hk
Keywords: diabetic kidney disease, biomarkers, untargeted metabolomics, random forest
Received: April 7, 2023   **Accepted:** December 6, 2023   **Published:** February 12, 2024

ABSTRACT

Diabetic kidney disease (DKD) is a leading cause of end-stage renal disease (ESRD) worldwide. Early detection is critical for the risk stratification and early intervention of progressive DKD. Serum creatinine (sCr) and urine output are used to assess kidney function, but these markers are limited by their delayed changes following kidney pathology, and lacking of both sensitivity and accuracy. Hence, it is essential to illustrate potential diagnostic indicators to enhance the precise prediction of early DKD. A total of 194 Chinese individuals include 30 healthy participants (Stage 0) and 164 incidents with type 2 diabetes (T2D) spanning from DKD's Stage 1a to 4 were recruited and their serums were subjected for untargeted metabolomic analysis. Random forest (RF), a machine learning approach, together with univariate linear regression (ULR) and multivariate linear regression (MvLR) analysis were applied to characterize the features of untargeted metabolites of DKD patients and to identify candidate DKD biomarkers. Our results indicate that 2-(α-D-mannopyranosyl)-L-tryptophan (ADT), succinyladenosine (SAdo), pseudouridine and N,N,N-trimethyl-L-alanyl-L-proline betaine (L-L-TMAP) were associated with the development of DKD, in particular, the latter three that were significantly elevated in Stage 2-4 T2D incidents. Each of the four metabolites in combination with sCr achieves better performance than sCr alone with area under the receiver operating characteristic curve (AUC) of 0.81-0.91 in predicting DKD stages. An average of 3.9 years follow-up study of another cohort including 106 Stage 2-3 patients suggested that "urinary albumin-to-creatinine ratio (UACR) + ADT + SAdo" can be utilized for better prognosis evaluation of early DKD (average AUC = 0.9502) than UACR without sexual difference.

INTRODUCTION

Diabetic kidney disease (DKD) is the most common cause of end stage renal disease (ESRD), affecting 20–30% of diabetic patients globally [1]. Standard biomarkers including urinary albumin-to-creatinine ratio (UACR) and estimated glomerular filtration rate (eGFR) are the clinical parameters commonly used to evaluate renal function in clinical practice. However, due to the level of urinary output and serum creatinine

(sCr) can be influenced by many factors, these measures are restricted as they may lack of sensitivity and accuracy [2]. Therefore, there is an urgent need to identify novel biomarkers for the diagnosis and management of DKD.

Metabolomics is a promising tool for detailed characterization of dynamic molecular changes in the intra- and inter-cellular process. It has been applied in multiple fields such as metabolism of drugs or environmental toxicants, screening for new therapeutic targets, discovery and validation of disease biomarkers [3]. Increasing evidence has revealed the association among metabolites, diabetes mellitus (DM) and diabetic complications [4]. Serum metabolic analysis of Korean T2D patients suggested that alanine, arginine, isoleucine, proline, tyrosine, valine, hexose and five phosphatidylcholine diacyls were positively associated with T2D risk [5]. For DKD prediction, Huang et al. utilized targeted metabolomics profiles to evaluate prospective metabolite predictors in the German diabetic individuals of the Region of Augsburg (KORA) cohort, and identified sphingomyelin (SM) C18:1 and phosphatidylcholine diacyl (PC aa) C38:0 as the potential metabolite biomarkers. [6]. In addition to metabolites, elements such as neutrophil gelatinase-associated lipocalin (NGAL), fatty acid-binding protein [7] and cystatin C [8] have been proposed to be correlated with the development of DKD. However, studies exploring the associations between metabolites and the DKD disease development in Chinese are very limited.

In this study, we performed untargeted metabolomics of 194 serum samples collected from 164 Chinese incidents of type 2 diabetes (T2D) and 30 healthy participants (non-T2D and non-diabetes). The metabolites identified were validated by an extra follow up cohort of 106 subjects with a mean follow-up time of 3.9 years. By assessing the predictive power of metabolites via a stringent workflow, we finally identified pseudouridine, L-L-TMAP, ADT and SAdo as the candidate predictors for early DKD.

RESULTS

Baseline characteristics of study participants

An overview of this study design was shown in Figure 1. For the 164 diabetic patients in the discovery and validation cohort, their median diabetic duration was 8 years, median eGFR was 76 (43–104) mL/min/1.73 m^2, and median UACR was 80 (ranges from 10 to 842) mg/g Cr. Approximately 67% of them had a history of $\geq$ one diabetic microvascular or macrovascular complication (Supplementary Table 1). The median eGFR and UACR of 30 healthy participants were 99 (95–112) mL/min/1.73 m^2 and 2.7 (2.3–3.9) mg/g Cr.,
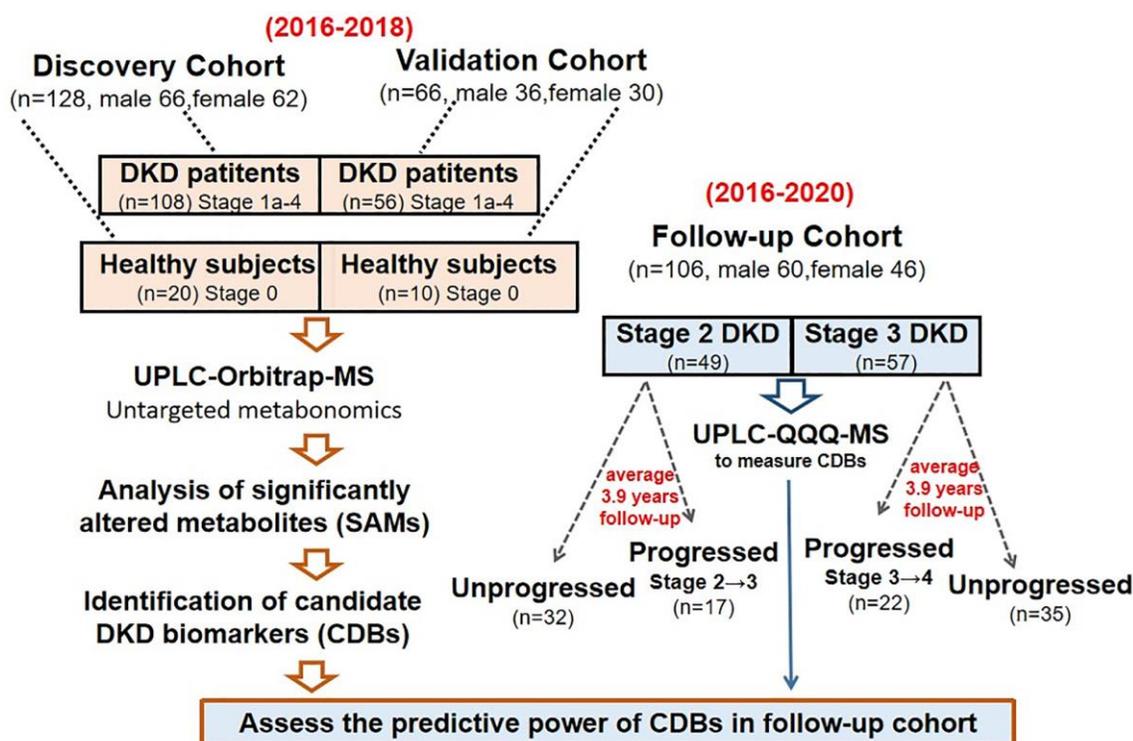
Figure 1. The pipeline of this study. Three independent cohorts were recruited to perform the metabolite biomarker in discovery, validation and follow-up groups, respectively.

respectively. The baseline characteristics of the two cohorts were compared, the patients group showed lower eGFR and an increase of UACR, RRI, systolic blood pressure (SBP), serum creatinine, urea, uric acid, cystatin C and urinary β2-microglobulin (β2-MG) concentrations along with DKD severity (Supplementary Table 2). We noticed that Stage 1a patients showed an enlargement of kidney size in compare with healthy participants by renal ultrasound images and testing body surface-area (BSA) related renal volumes (Supplementary Figure 1); however, for patients at Stage 2–4, their BSA related renal volumes were gradually decreased with DKD progressed (Supplementary Figure 1), which is consistent with previous findings that kidneys tended to be smaller in the most advanced stages of CKD [9]. Our results suggested that patients were likely to have abnormal kidney hypertrophy and enlargement at Stage 1a, followed by gradual renal atrophy and volume depletion at later stages.

Characterization of metabolites in study participants

A total of 7480 compounds were detected and we found that the trend of MS-detected sCr among all the participants showed remarkable consistency with clinically measured sCr (Supplementary Figure 2), suggesting that UPLC-Orbitrap-MS is accurate and efficient for high-throughout metabolites detection. Stepwise filtering was performed based on two criteria: removing metabolites with unstable signals and only retaining metabolites with significant different concentration levels between healthy control and patients, 80 candidates (72 metabolites and 8 ratios) were screened out for a next-step analysis (Supplementary Tables 3 and 4, Supplementary Figure 3). The distribution plot of preprocessed data was shown as Supplementary Figure 4. The fold changes of these metabolites among different stages were calculated and shown in Supplementary Tables 5 and 6.

The 72 metabolites are classified to 7 categories: sulfate metabolites, amino acids, organic acids, acylcarnitine, purine derivatives, steroids and monosaccharides. Metabolomic network based on the 72 metabolites in the discovery set from Stage 0–4 was shown in Supplementary Figure 4. Comparing with healthy group (Stage 0), merely 3 down- and 6 up- regulated metabolites were found in Stage 1a; nevertheless, it increased to 7 down- and 40 up-regulated metabolites in Stage 4. The SAMs were enriched in Tryptophan metabolism (hsa00380) and Phenylalanine metabolism (hsa00360) pathways, indicating that amino acid metabolism disruption is a dominate signature of DKD (Figure 2). Among these significant altered metabolites (SAMs), 1,5-anhydro-D-glucitol (1,5-AG) was remarkably reduced in stages 1–4 compared to healthy group (fold change = −26.5 to −2.60, Supplementary Table 5). As demonstrated by previous studies that 1,5-AG is a potential biomarker for monitoring the progression of diabetes [10, 11], we therefore separately tested the correlation of two clinical glycemic markers – fast blood glucose (FBG) and hemoglobin A1C (HbA1c) with 1,5-AG in our cohort. It showed that 1,5-AG has strong negative correlation with HbA1c and FBG in stage 1a-3 patients (r ranges were −0.95 to −0.64 and −0.87 to −0.42, respectively); however, abnormal correlation was observed between 1,5-AG and FBG in stage 4 discovery sets, with r = 0.25 (Supplementary Figure 5A, 5B). Correlation between 1,5-AG and HbA1c was stronger among stage 1a-3 patients than stage 0–4 (Supplementary Figure 5A, 5B), suggesting that 1,5-AG may serve better as a potential glycemic marker in stage 1a-3 DKD patients than late stage.

Given that DKD is one of the consequences induced by diabetes, we hypothesized that decreased levels of 1,5-AG may be relevant with DKD development. However, our results showed that 1,5-AG exhibited non-significant correlation with neither eGFR nor UACR (Supplementary Figure 5C), indicating that diabetic progression has limited contribution to DKD development (scatter plot of 1,5-AG levels among healthy controls and different stages of patients was shown in Supplementary Figure 5D).

Identification of candidate DKD biomarkers (CDBs)

Receiver Operating Characteristic (ROC) curve analysis and Spearman's coefficient coexpression analysis were used to evaluate the power of each metabolite as well as the combinations of every 2–9 compounds in DKD staging. The metabolites that closely associated ($|r| \geq 0.6$) with eGFR in all participants were shown in Supplementary Figure 6. Strict rank coefficient cut-off values of 0.8 for Stage 0–4 (all participants) and Stage 1–4 (all patients), and 0.7 for Stage 1 and 2 (early-stage patients) were applied to identify biomarkers that closely correlated with eGFR progressive, four metabolites were screened out consist of SAdo ((M-H)⁻ = 382.1005 at 2.89 min), pseudouridne ((M-H)⁻ = 243.0622 at 0.93 min), ADT ((M-H)⁻ = 367.1497 at 2.21 min) and L,L-TMAP ((M+H)⁺ = 229.1546 at 1.06 min) (The regression plots among the four metabolites, eGFR and UACR see Figure 3). Since few studies have investigated the basic features of SAdo, the demonstration of its peak identification was plotted and calibrated (Supplementary Figures 7 and 8). To alleviate the bias induced by sex, age, SBP and UACR, partial correlation analysis (PCA) was performed among the four metabolites, serum cystatin C, MS-detected serum

creatinine (MS-sCr) and log(eGFR). The PCA showed that log(MS-sCr), log(pseudouridine) and log(L,L-TMAP) were strikingly correlated with log(eGFR) ($|r| > 0.9$) in Stage 0–4 (Table 1). In addition, we evaluated the association between the four metabolites and kidney function related factors such as UACR, urinary $\beta2$- microglobulin, renal resistive index and the decrease of total BSA related renal volume, and found they were closely related as well ($|r| > 0.5$, see Table 1). The correlation among interested metabolites, total BSA-related renal volume and renal resistive index was calculated as well (Supplementary Table 7). In conclusion, our results suggested that the four metabolites are possibly involved in DKD progression

and have potential to be utilized as candidate DKD biomarkers (CDBs).

Evaluation of CDB's capacity in staging DKD by random forest (RF)

To evaluate whether CDBs can be applied for staging DKD, RF was employed to assess the classification power of the four CDBs as RF is a powerful supervised classification technique for decision making via building large numbers of decision tree models and merging all predictions from these trees to get an accurate and unprogressed prediction [12]. It exhibited that CDBs can specifically differentiate Stage 1a

Pathway with KEGG database with Homo sapiens	Total	Hits	*p* value	-log(p)	FDR	Impact
Tryptophan metabolism	**79**	**4**	**0.0018**	**6.3081**	**0.0291**	**0.191**
Phenylalanine metabolism	**45**	**4**	**0.0016**	**6.4543**	**0.0291**	**0.151**
Alanine, aspartate and glutamate metabolism	24	2	0.0308	3.4815	0.2479	0.384
D-Glutamine and D-glutamate metabolism	11	2	0.0067	5.0014	0.0769	0.139
Cysteine and methionine metabolism	56	2	0.1369	1.9882	0.9129	0.111
Arginine and proline metabolism	77	6	0.0002	8.5717	0.0076	0.096
Aminoacyl-tRNA biosynthesis	75	5	0.0014	6.5426	0.0291	0.056
Glycine, serine and threonine metabolism	48	3	0.0172	4.0635	0.1719	0.023
Purine metabolism	92	3	0.0889	2.4198	0.6468	0.018
Porphyrin and chlorophyll metabolism	104	2	0.3429	1.0703	1	0.011
Phenylalanine, tyrosine and tryptophan biosynthesis	27	3	0.0034	5.6775	0.0456	0.008
Nitrogen metabolism	39	6	3.53E-06	12.555	0.0003	0
Pyrimidine metabolism	60	3	0.0310	3.474	0.2479	0

Figure 2. KEGG pathway analysis of all SAMs. (**A**) An overview view of pathway analysis; (**B**) Table of the matched pathway with *p*-values from pathway enrichment analysis and pathway impact values from the pathway topology analysis using MetaboAnalyst 4.0 and KEGG database (Hits ≥ 2); (**C**, **D**) Simplified pathways of tryptophan metabolism and phenylalanine metabolism with the change trends of metabolites and their ratios at Stage 4 compared with the normal group.

patients from Stage 1b-4, Stage 1a from Stages 1b-2, Stage 1b from Stage 2, and Stages 1a-1b from Stages 2-4 (average AUC > 0.700, Table 2). Among them, pseudouridine and SAdo achieve better performance than MS-sCr in all the staging process (Supplementary Table 8). Multiple combinations of the 4 metabolites and MS-sCr were generated and used for assessing their ability for DKD stratification. Any one of the four candidate DKD biomarkers combined with MS-sCr can gain higher AUC than MS-sCr alone (Table 2). Among
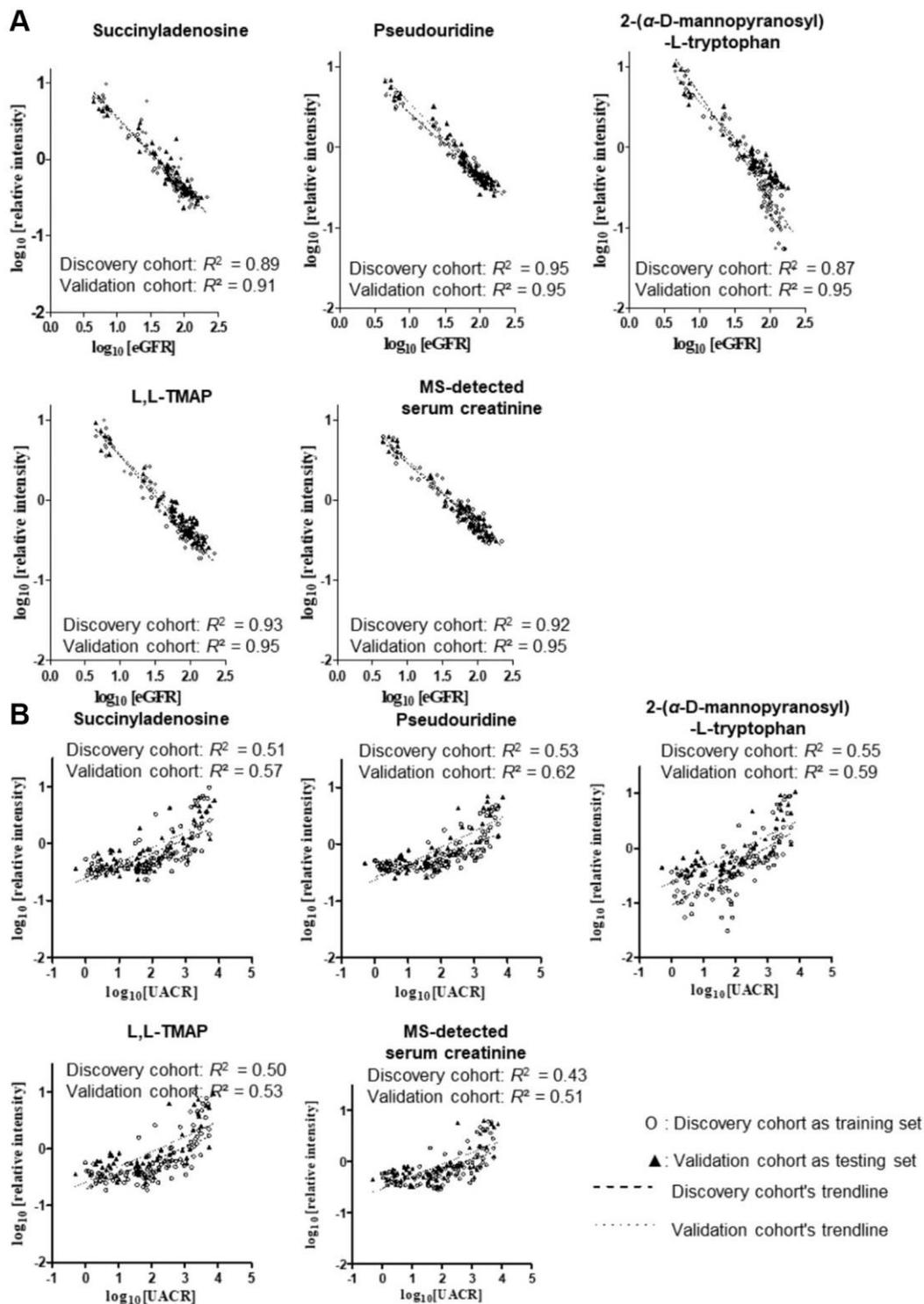
Figure 3. Linear regression analysis among CDBs, sCr, eGFR (**A**) and UACR (**B**) in all stages and T2D patients after log10 transformation, respectively. Four metabolites showed similar strong predictive power with MS-detected sCr as their R2 of the equations were all above 0.85. R2 of UACR prediction model were close to 0.5.

Table 1. Spearman's coefficient correlation *r* analysis among four metabolites, clinically measured sCr, cystatin C, eGFR and interested CKD risk factors in two cohorts.

Stage ranges	Stage 0–4		Stage 1–4		Stage 1 and 2 (eGFR ≥60)	
	Discovery (*n* = 128)	Validation (*n* = 66)	Discovery (*n* = 108)	Validation (*n* = 56)	Discovery (*n* = 69)	Validation (*n* = 30)
Spearman rank correlation with eGFR						
Pseudouridine	−0.896	−0.939	−0.939	−0.953	−0.789	−0.792
ADT	−0.875	−0.928	−0.918	−0.935	−0.733	−0.847
MS-detected creatinine	−0.869	−0.912	−0.900	−0.916	−0.712	−0.739
L,L-TMAP	−0.866	−0.901	−0.921	−0.922	−0.734	−0.739
Succinyladenosine	−0.855	−0.926	−0.899	−0.932	−0.706	−0.818
Serum cystatin C	−	−0.913	−	−0.940	−	−0.802
Partial correlation of metabolites with log (eGFR) after controlling sex, age, SBP and log (UACR)						
log (MS-detected creatinine)	−0.957	−0.975	−0.962	−0.976	−0.850	−0.847
log (Pseudouridine)	−0.941	−0.949	−0.947	−0.952	−0.739	−0.709
log (L,L-TMAP)	−0.924	−0.950	−0.930	−0.953	−0.685	−0.698
log (Succinyladenosine)	−0.867	−0.898	−0.874	−0.898	−0.533	−0.571
log (2-(α-D-Mannopyranosyl)-L-tryptophan)	−0.823	−0.956	−0.844	−0.958	−0.701	−0.755
log (Serum cystatin C)	−	−0.949	−	−0.953	−	−0.831

CKD risk factors (Stage 1–4)	UACR		Urinary *β*2-microglobulin		Total BSV-related renal volume	Renal resistive index
	Discovery (*n* = 108)	Validation (*n* = 56)	Discovery (*n* = 108)	Validation (*n* = 54)	Validation (*n* = 48)	Validation (*n* = 48)
2-(α-D-Mannopyranosyl)-L-tryptophan	0.801	0.746	0.664	0.731	−0.604	0.586
Succinyladenosine	0.795	0.690	0.635	0.780	−0.615	0.556
Pseudouridine	0.794	0.741	0.690	0.793	−0.599	0.588
L,L-TMAP	0.743	0.676	0.670	0.681	−0.596	0.517
MS-detected creatinine	0.681	0.657	0.667	0.685	−0.473	0.459
Serum cystatin C (mg/L)	−	0.676	−	0.675	−0.614	0.544

Table 2. List of mean AUC values for evaluating the predictive power of MS-detected sCr and multiple-metabolite models for differentiating DKD stages in T2D patients using random forest classification in two cohorts.

Classification	Cohort	Serum creatinine	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
eGFR ≥ 119 vs. eGFR < 119	Dis	0.85 ± 0.05	0.93 ± 0.03	0.91 ± 0.03	0.92 ± 0.05	0.94 ± 0.02	0.93 ± 0.02	0.94 ± 0.02	0.95 ± 0.02
	Val	0.88 ± 0.08	0.93 ± 0.03	0.94 ± 0.04	0.94 ± 0.03	0.92 ± 0.03	0.92 ± 0.03	0.94 ± 0.03	0.93 ± 0.03
eGFR ≥ 119 vs. eGFR = 60–118	Dis	0.75 ± 0.08	0.87 ± 0.05	0.84 ± 0.06	0.88 ± 0.07	0.91 ± 0.04	0.89 ± 0.04	0.89 ± 0.05	0.91 ± 0.04
	Val	0.76 ± 0.10	0.85 ± 0.06	0.87 ± 0.07	0.87 ± 0.07	0.83 ± 0.09	0.81 ± 0.09	0.85 ± 0.07	0.89 ± 0.06
eGFR = 90–118 vs. eGFR = 60–89	Dis	0.68 ± 0.09	0.81 ± 0.06	0.79 ± 0.07	0.80 ± 0.07	0.77 ± 0.06	0.83 ± 0.05	0.82 ± 0.06	0.81 ± 0.06
	Val	0.60 ± 0.17	0.97 ± 0.06	0.86 ± 0.08	0.99 ± 0.03	0.97 ± 0.06	0.96 ± 0.05	0.98 ± 0.03	0.95 ± 0.07
eGFR ≥ 90 vs. eGFR < 90	Dis	0.92 ± 0.03	0.96 ± 0.02	0.93 ± 0.02	0.94 ± 0.02	0.94 ± 0.02	0.96 ± 0.02	0.96 ± 0.02	0.95 ± 0.02
	Val	0.93 ± 0.05	0.99 ± <0.01	0.98 ± 0.02	0.97 ± 0.02	1.00 ± <0.01	1.00 ± 0.01	1.00 ± <0.01	0.99 ± 0.01
eGFR ≥ 60 vs. eGFR < 60	Dis	0.95 ± 0.03	0.99 ± 0.01	0.97 ± 0.01	0.98 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	0.98 ± 0.01	0.99 ± 0.01
	Val	0.931± 0.04	0.96 ± 0.03	0.96 ± 0.03	0.94 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.96 ± 0.03	0.97 ± 0.02

eGFR ≥ 30 vs. eGFR < 30	Dis	0.99 ± <0.01	0.99 ± 0.01	1.00 ± <0.01	1.00 ± <0.01	1.00 ± 0.01	1.00 ± <0.01	1.00 ± <0.01	1.00 ± <0.01
	Val	0.97 ± 0.08	0.99 ± 0.04	1.00 ± 0.02	0.99 ± 0.03	0.99 ± 0.04	0.99 ± 0.04	0.99 ± 0.03	0.98 ± 0.06

Data were expressed as mean ± SD. Abbreviations: Dis: discovery cohort; Val: validation cohort. Model 1: MS-detected sCr + pseudouridine; Model 2: MS-detected sCr + Sado; Model 3: pseudouridine + SAdo; Model 4: pseudouridine + L,L-TMAP; Model 5: Model 1 + ADT; Model 6: Model 1 + Sado; Model 7: Model 1 + L,L-TMAP.

these models, the No. 7 model (MS-sCr + pseudouridine + L,L-TMAP) ranks the best in phasing Stage 1a from the rest with average AUC > 0.9 (Table 2).

Comparison of CDBs between male and female patients

Increasing evidence suggested that sexual difference is a significant factor related with DKD progression, which leads to a complex personalized approach for DKD diagnosis and treatment in clinical practice [13, 14]. To investigate the association between the four CDBs and sexual difference, we compared their levels between male and female patients. For SAdo, pseudouridine and ADT at early and later stages, no significant differences was observed (Figure 4). Multiple linear regression analysis which included sex as a covariate showed that CDBs show insignificant sex dependence with eGFR, suggesting that CDBs can be utilized in both male and female patients (Supplementary Table 9).

Predict eGFR using CDBs signatures

We hypothesized that the use of a combinations of multiple biomarkers may be more sensitive and specific than sCr in evaluating the kidney function of diabetic patients. To test the potential of CDBs in predicting DKD, non-parametric methods include univariate linear regression (ULR) and multivariate linear regression (MvLR) were applied to calculate the association among eGFR, UACR, creatinine and four CDBs. The ULR analysis using Stage 0–4 data found a high linear relationship between each CDB and log(eGFR) (training R^2 = 0.87–0.95, root mean square errors (RMSEs) = 0.08–0.13; predictive R^2 = 0.91–0.95) which was very close to MS-detected sCr (training R^2 = 0.95, RMSE = 0.11; predictive R^2 = 0.95) (Figure 5), suggested that CDBs are good covariates to be applied for eGFR prediction. A stepwise MvLR analysis was performed to test the effects of four CDBs and covariates (sex, age, SBP and UACR) in calculating CDB-predicted eGFR (BeGFR) using data from Stage 0–4 and Stage 0–2, respectively. Among all the individuals (Stage 0–4), MS-sCr, pseudouridine, L,L-TMAP and sex are the most significant variables, and were therefore considered as confounding covariants for BeGFR estimation. The predictive outcome R^2 was optimized from 0.971 (log(MS-detected sCr) and sex as covariates) to 0.987 (Table 3).

Renal function of early-stage DKD (Stage 1 and 2) is reversible and manageable [15, 16]; however, most DKD patients are asymptomatic and indolent [17, 18]. Considered that early detection is of great vital for lifetime benefits for DKD patients, we specifically tested the predictive potency of CDBs and eGFR in early-stage participants. Surprisingly, pseudouridine and L,L-TMAP can enhance the predictive power of MS-detected sCr at early stages' patients and healthy participants (Stage 0–2), predictive R^2 in the validation datasets was significantly improved from 0.70 (log(MS-detected creatinine) and sex as covariates) to 0.82 (Table 3), demonstrating that the two CDBs are potential biomarkers for the early detection of DKD. The best model for BeGFR estimation at the early stage (training R^2 = 0.7733, RMSEs = 0.0513) is: log(BeGFR) = −0.675 log(MS-detected sCr) −0.467 log(pseudourdine) + 0.101 (if male) + 1.559. Our results indicated that the combination of multiple biomarkers achieves better performance than standard sCr.

Follow-up study and prognostic assessment

Due to the limitations for purchasing the commercial standards of L-L-TMAP, only three CDBs include ADT, SAdo and pseudouridine and some clinical indexes were measured in the follow-up cohort. The association were assessed between 7 variates (sex, age, eGFR, sCr, ADT, SAdo and pseudouridine) as well as their combinations with DKD progression. The first-time collected serum samples from 106 subjects of Stage 2 and 3 DKD patients were examined to determine the concentration of the metabolites. In Stage 2 unprogressed and Stage 2 progressed groups, for each single variate, ADT ranks the top prognostic power (average AUC = 0.9184) than sCr alone (average AUC = 0.9133). Surprisingly, the combinations of "UACR + ADT + sCr" and "UACR + ADT + age + sex" were extremely associated with the future progression of DKD (AUC ranges from 0.9592 to 1) (Figure 6A, Supplementary Table 10). The same method was applied in Stage 3 unprogressed and Stage 3 progressed patients. Three single variates, including UACR, pseudouridine and ADT, gained the top 3 strongest association with DKD progression (average AUC values = 0.8889, 0.8302 and 0.8117, respectively). For the Stage 3 patients, either of the four metabolites (ADT, sCr, SAdo and pseudouridine) combined with UACR can optimize average AUC value ≥ 0.9012; among the combinations, "UACR +

ADT + pseudouridine + SAdo + Sex" and "UACR + pseudouridine + sCr" achieve the best performance with an average AUC = 0.929 (Figure 6A, Supplementary Table 11). Ignoring the initial DKD phasing status, "UACR + ADT + SAdo" is the best combination for DKD prognostic assessment (average AUC = 0.9502).

In general clinical practice, eGFR was employed as a popular reporter for grading diseases as its levels reflect the status of renal function decline; while UACR was mostly used as the predictive biomarker for disease's progression [7]. Hence, we only compared the changes of UACR and CDBs between progressed and
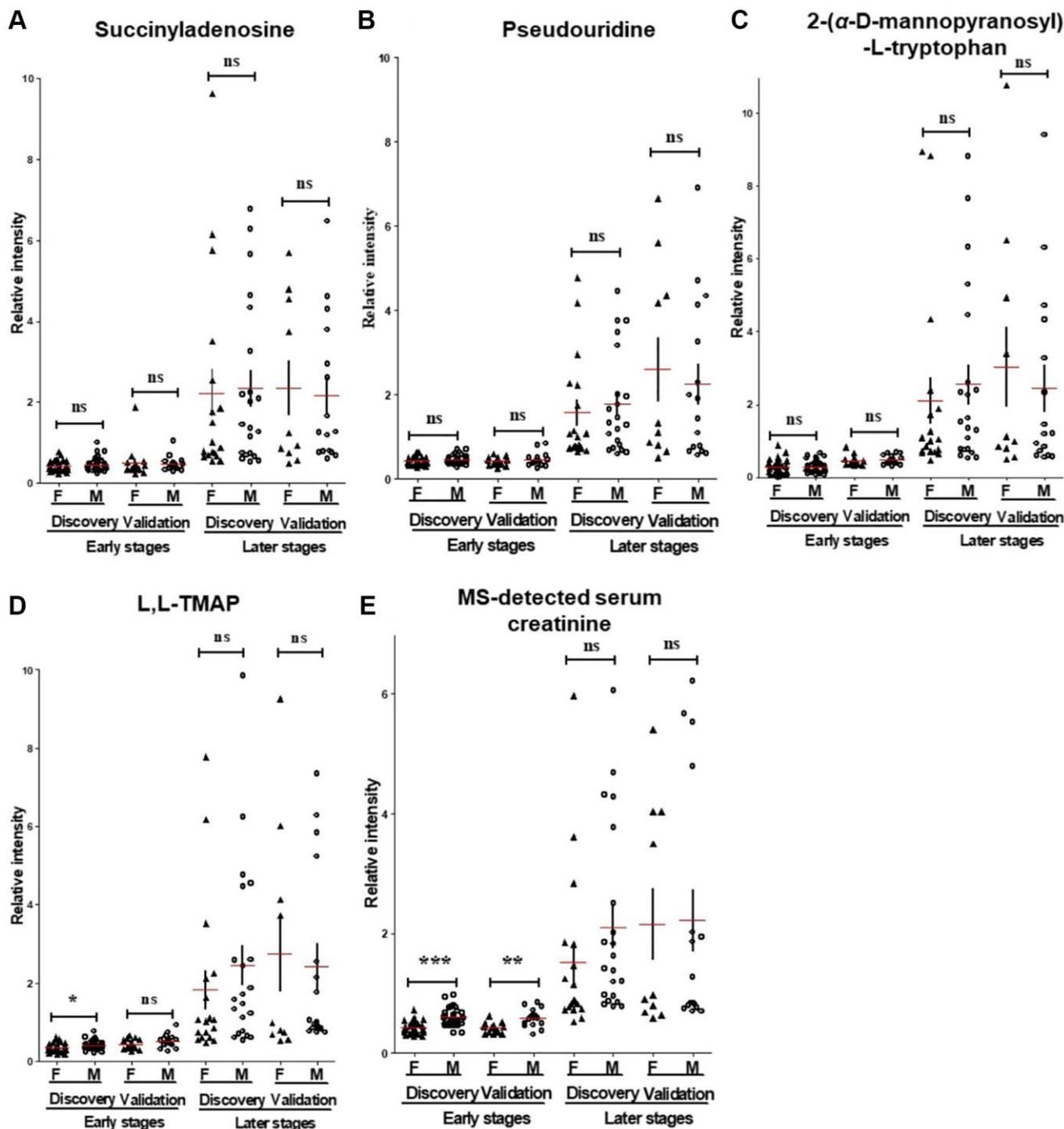
Figure 4. The MS-detected CDBs. (**A–D**) and sCr (**E**) were evaluated and compared between male and female participants at the early and later stages' patients. Unlike with sCr, three of the four CDBs include SAdo, pseudouridine and ADT displayed non-significant differences between male and female patients at early or later stages. F, women; M, men. p value was calculated by Student's t-test and Mann-Whitney U according to the data normality. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, respectively. Horizontal and error bars in the scatter plots represent mean ± SEM.

unprogressed groups in all follow-up individuals, it showed that the levels of ADT, pseudouridine and AUCR were significant different between the two categories while SAdo showed slightly but non-significant differences (Figure 6B). Among stage 2 incidents, concentrations of ADT, pseudouridine, Sado and AUCR were remarkably different in progressed patients comparing with the unprogressed (Figure 6C). Further logistic regression analysis indicated that ADT and pseudouridine are risk factors for DKD

development (Figure 6D). With every single increase of standard deviation (SD) of ADT, the risk of DKD progression is enhanced for 2.151 folds; for pseudouridine, the risk scores are 1.741 folds (Figure 6D). To better evaluate the associations between the levels of risk factors and the DKD progressing, we used the duration of time for progressing to later stage in follow-up individuals for survival curve analysis. It showed that patients with higher levels of either pseudouridine or ADT had significant less survival
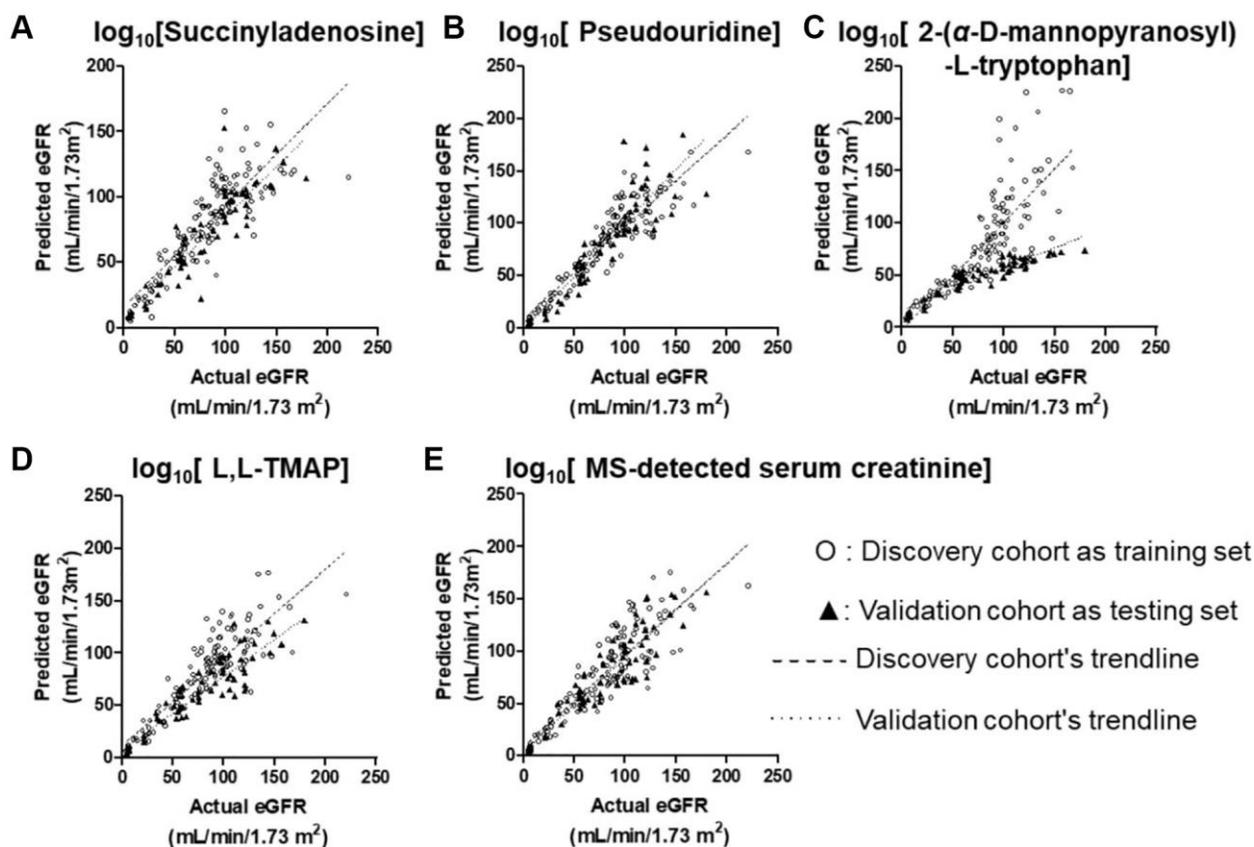
log [eGFR]	Discovery cohort (training)				Validation cohort (testing)
	β (95% CI)	p value	R^{2*}	RMSE	R^2 †
log [MS-detected creatinine]	-1.16 (-1.23 to -1.10)	***	0.9110	0.1052	0.9512
log [pseudouridine]	-1.19 (-1.25 to -1.14)	***	0.9448	0.0829	0.9515
log [L,L-TMAP]	-0.93 (-0.98 to -0.88)	***	0.9200	0.0997	0.9467
log [succinyladenosine]	-0.92 (-0.98 to -0.86)	***	0.8843	0.1200	0.9148
log [2-(α-D-mannopyranosyl)-L-tryptophan]	-0.64 (-0.68 to -0.59)	***	0.8479	0.1376	0.9516

Figure 5. Univariate linear regression plots of BeGFR against MDRD eGFR using the four CDBs. (**A–D**) and MS-detected sCr (**E**) for all participants at Stages 0–4 after $\log_{10}$ transformation. Univariate linear regression analysis of each selected metabolites with all participants' log(MDRD eGFR) resulted in a high linear relationship (training R^2 = 0.85–0.94, root mean square errors (RMSEs) = 0.08–0.13; predictive R^2 = 0.91–0.95), which was similar with that of MS-detected sCr (training R^2 = 0.95, RMSE = 0.11; predictive R^2 = 0.95). ***$p <$ 0.001 β, unstandardized coefficient of linear regression. *R^2 was calculated based on the log(BeGFR) against log (eGFR) using the equation of the model and data of the discovery cohort. †R^2 was measured based on that using the equation of the model of the discovery cohort and data of the validation cohort.

Table 3. Multivariate linear regression analyses of biomarkers with log (eGFR) trained with discovery set and tested with validation set among all participants.

	Stages 0–4			Stages 0-2		
	Discovery (training)		Validation (testing)	Discovery (training)		Validation (testing)
Prediction of log (BeGFR)	R^{2*}	RMSE	$R^{2\dagger}$	R^{2*}	RMSE	$R^{2\dagger}$
	0.9562	0.0740	0.9714	0.6802	0.0606	0.7050
	β (95% CI)	p value		β (95% CI)	p value	
log (MS-detected creatinine)	**–1.23 (–1.28 to –1.19)**	***		**–0.93 (–1.06 to –0.79)**	***	
sex	**0.15 (0.13 to 0.18)**	***		**0.12 (0.09 to 0.16)**	***	
All biomarkers and common covariates	R^{2*}	RMSE	$R^{2\dagger}$	R^{2*}	RMSE	$R^{2\dagger}$
	0.9403	0.0514	0.9855	0.7949	0.0503	0.8348
	β (95% CI)	p value		β (95% CI)	p value	
log (MS-detected creatinine)	–0.71 (–0.86 to –0.56)	***		–0.71 (–0.87 to –0.54)	***	
log (pseudouridine)	–0.35 (–0.58 to –0.13)	**		–0.23 (–0.50 to 0.07)	ns	
log (L,L-TMAP)	–0.14 (–0.28 to –0.01)	*		–0.06 (–0.22 to 0.10)	ns	
log (succinyladenosine)	–0.01 (–0.11 to 0.09)	ns		0.01 (–0.13 to 0.14)	ns	
log (2-(α-D-mannopyranosyl)-L-tryptophan)	–0.02 (–0.09 to 0.04)	ns		–0.07 (–0.15 to 0.003)	ns	
sex	0.11 (0.08 to 0.13)	***		0.10 (0.07 to 0.13)	***	
age	–0.0009 (–0.003 to 0.001)	Ns		–0.0005 (–0.003 to 0.002)	ns	
SBP	0.0003 (–0.0003 to 0.0009)	ns		0.0003 (–0.0003 to 0.0009)	ns	
BMI	0.002 (–0.001 to 0.006)	ns		0.0026 (–0.001 to 0.01)	ns	
log (UACR)	0.006 (–0.009 to 0.02)	ns		0.0049 (–0.01 to 0.02)	ns	
The best model by stepwise method using variables with $p < 0.05$	R^{2*}	RMSE	$R^{2\dagger}$	R^{2*}	RMSE	$R^{2\dagger}$
	0.9754	0.0514	0.9870	0.7733	0.0513	0.8200
	β (95% CI)	p value		β (95% CI)	p value	
log (MS-detected creatinine)	–0.64 (–0.77 to –0.51)	***		–0.68 (–0.82 to –0.53)	***	
log (L,L-TMAP)	–0.13 (–0.25 to –0.01)	*		N/A	N/A	
log (pseudouridine)	–0.44 (–0.59 to –0.30)	***		–0.46 (–0.62 to −0.31)	***	
sex	0.10 (0.08 to 0.12)	***		0.10 (0.07 to 0.13)	***	

β, unstandardized coefficient of linear regression. Sex, female = 1 and male = 2. RMSE, root mean square error. $^{*}R^2$ was based on the predicted log (eGFR) against actual log (eGFR) using the equation of the model and data of discovery set. $^{\dagger}R^2$ was based on that using the equation of the model of discovery set and data of validation set. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, Abbreviation: ns: indicates no significance.

probabilities ($p < 0.05$), which is similar to UACR (Supplementary Figure 9). Taken together, our cross-sectional study indicated that abnormal metabolism is involved in DKD progression and our follow-up study validated the predictive power of CDBs in DKD development.

DISCUSSION

A precise assessment of renal function in the clinical settings, would be instructive for management of DKD, such as for the prediction and intervention of the disease progression, CKD staging, for assessing the need for dialysis therapy, and adjustment of nephrotoxic agents dosage for patients [19]. In the past decades, eGFR has been applied as the best overall measurement of kidney function in medical practice; however, it also has some limits on accuracy and reliability [20, 21]. To overcome the limitations, over 70 equations have been developed for estimating eGFR. We applied Modification of Diet in Renal Disease (MDRD) formula to calculate eGFR for DKD classification; nevertheless, other methods include (CKD-EPI)creatinine [22], CKD-EPIcystatin C [23] and

CKD-EPI[creatinine–cystatin C] [23] equations were calculated as well. CKD-EPI[creatinine] eGFR was strongly correlated with MDRD eGFR (Pearson's r = 0.9523 and 0.9729 in discovery and validation sets, respectively). CKD-EPI[cystatin C] and CKD-EPI[creatinine–cystatin C] eGFR also show high correlation with MDRD eGFR with
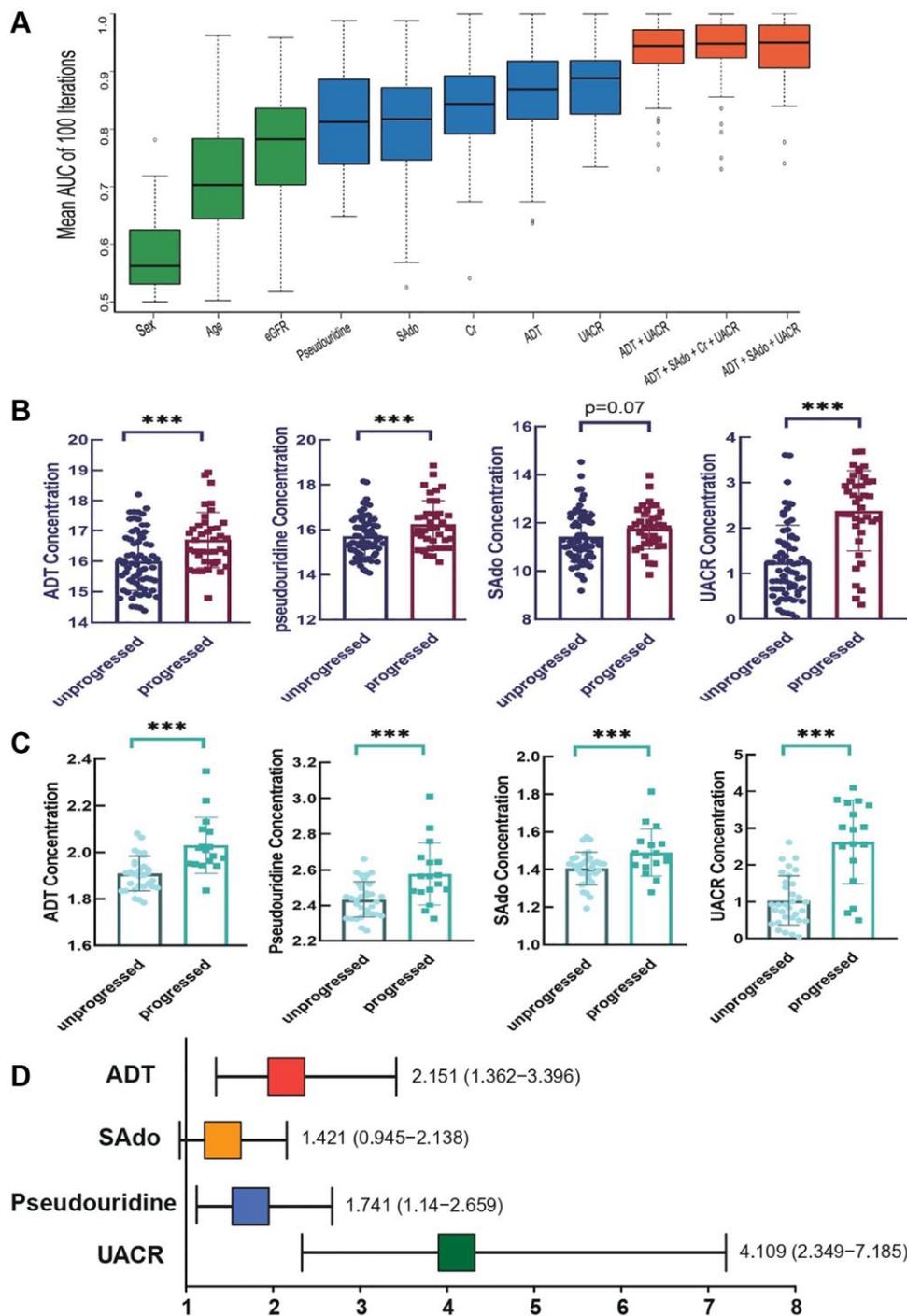
Figure 6. Evaluation of the prognostic performance of CDBs in follow-up cohort. (**A**) The distribution of AUC values using single and combinations of variate(s) in follow-up patients. With stratified random sampling and random forest, AUC of distinction between and progressed patients were calculated 100 times with single and multiple variables. Results of AUC average and standard deviation indicated that ADT_SAdo_UACR (AUC average: 0.9502; CI: 0.9062–0.9805) manifested the best prediction, followed with UACR_ADT_SAdo_sCr (AUC average: 0.9482; CI:0.9248-0.9805) and ADT_UACR (AUC average: 0.9443; CI: 0.9141–0.9727). (**B**) Comparisons of the three CDB levels and UACR between "progressed" and "unprogressed" groups in all the follow-up individuals. (**C**) In stage 2 patients, levels of three CDBs and UCAR were remarkably different in "progressed" vs. "unprogressed", $^{**}p < 0.01$ and $^{***}p < 0.001$, respectively. (**D**) Risk sores of three CDBs and UCAR in DKD progression by logistic regression analysis.

Pearson's r = 0.9468 in discovery sets and 0.9681 in validation sets (Supplementary Table 12).

Machine learning approaches such as RF, decision tree, logistic regression and XG Boost have greatly advanced the development of biomedical science especially for the prognostic prediction of human diseases. RF was applied to assess the covariates associated with DKD development as it is one of the most efficient and widely used algorithms that leverages a collection of decision trees for making decisions; on the other hand, we used logistic regression to the estimates the risk probability of DKD progression using CDB levels considering this algorithm is useful to obtain odds ratio in the presence of more than one explanatory variable [12]. Metabolomics analysis can be classified into two categories, namely the non-targeted and the targeted approach. Considering that the first one is an unbiased metabolomic analysis that can discover new biomarkers [24], the non-targeted approach has been adopted to gain a more comprehensively and systematically knowledge of the progressive DKD. All the DKD patients suffer from dysregulated metabolic milieu including hyperglycemia and insulin resistance that lead to renal functions being impaired. In the design of this study, subjects with various degree of renal function impairments are recruited and these would be one of the major variations among the subjects. However, it is unavoidable that these subjects would also have different status of hyperglycemia and insulin resistance although most of them have a longer history of diabetes. Thus, although this cohort may not be a good one for metabolomics study of diabetes, but metabolties related to the progression of diabetes may also revealed in this study. The endogenous metabolite, 1,5-anhydro-D-glucitol (1,5-AG), is an example of these which correlate with eGFR, but show a stronger correlation with serum hemoglobin A1c (HbA1c) and fasting blood-glucose (FBG) which are important clinical markers for hyperglycemia. Apart from that, Liu et al. found the catabolism of amino acids in plasma of individuals of DKD with T2D was accelerated [25]. The targeted metabolic nuclear magnetic resonance (NMR) spectroscopy of European T2D patients revealed that the amino acids glycine, phenylalanine, the energy metabolites citrate and glycerol were negatively associated with eGFR, while alanine, valine and pyruvate depicted opposite association in diabetics (positive) and non-diabetics (negative) [26]. Accumulating evidence suggested that aromatic amino acids (phenylalanine) and branched-chain amino acids (BCAAs) such as leucine and valine were associated with an increased risk of developing T2D [27, 28]. Our results showed consistent findings that amino acids were significantly changed among different stages, especially leucine, valine and phenylalanine (Supplementary Tables 5 and 6). Although the markers correlated with the progression of diabetes is not the focus of this study, but our data clearly supported they are being affected during the development of diabetes. Metabolites related to both hyperglycemia and renal functions are being identified in the analysis also demonstrated the non-targeted metabolomics analysis performed in this study is of very good quality and the data is capable of revealing various differences in the host metabolism.

Since DKD is asymptomatic until later stages, its early detection is of great significance to provide an opportunity for preventing or delaying its progression and decreasing morbidity and mortality. Small molecules are extensively metabolized by kidney and the impaired renal function can lead to the changes of serum metabolites, hence, they may be used to estimate filtration (e.g., the established marker creatinine) or precede and potentially contribute to the development of kidney diseases [29]. In this study, pseudouridine, L-L-TMAP, ADT and SAdo were identified as the candidate biomarkers for optimizing DKD stratification and eGFR prediction. Pseudouridine has been identified as a non-traditional kidney function marker in previous study as it shows significant correlation with eGFR in general population [30] while TMAP has shown better performance than creatinine in accurately identifying patients with a single kidney [31]. Yonemura et al. revealed that the concentration of serum ADT is a more reliable diagnostic marker than that of serum creatinine as a measure of normal renal function [32]. Our study reconfirmed the potential of pseudouridine, L-TMAP, ADT in measuring renal function; in addition, to the best of our knowledge, we reported that SAdo is a new candidate biomarker and can be utilized to predict the progression of early stages' DKD for the first time. Interestingly, the concentration of the four serum biomarkers were not only strongly correlated with eGFR but also associated with non-GFR renal injury indicators (nGRI) including UACR, urinary β2-microglobulin, RRI and kidney sizes. These four nGRI were usually served as indicators of albuminuria [19, 33], renal proximal tubular reabsorption dysfunction [34], renal arterial damage and resistance [35], and kidney hyperfiltration and degeneration [36], respectively. Our results suggested the four metabolites are indicators of glomerular filtration dysfunction and kidney pathophysiology damages as well. In contrast with serum creatinine that can be easily affected by sex and muscle metabolism, we found SAdo, ADT and pseudouridine are sexual independent.

Prognostic markers play important role in DKD patients stratification, treatment choice and future outcome assessment. Our follow-up results of Stage-2 and Stage 3 patients offered further evidence to the hypothesis that

the four biomarkers are prognostic markers for disease progression (both renal function decline and UACR increment) in patients with early DKD. Similar results about pseudouridine and ADT in disease progression were acquired in our study, comparing with other DKD follow up cohort [37]. For the first time, our follow-up study gave a clinical evidence-based proof for succinyladenosine as a DKD prognostic marker and turned out to have good prognostic value, especially at early stage. These markers would facilitate both doctors and patients on their treatment selection and aid in clinical practice.

CONCLUSION

For the first time, we demonstrated that SAdo is a new potential biomarker for eGFR estimation and DKD prognostic assessment. Consistent with previous studies, the predictive potential of pseudouridine, L-TMAP, ADT in measuring renal function was further confirmed in our cohorts. These four serum biomarkers were not only strongly correlated with eGFR but also closely associated with non-GFR renal injury indicators. Unlike serum creatinine with noticeable sexual difference, SAdo, ADT and pseudouridine are sexual independent. Our follow-up study validated the prognostic power of the above biomarkers for both renal function decline and UACR increment in early DKD patients. This study provided comprehensive insights into the signatures of metabolites in Chinese DKD patients and identified four candidate biomarkers for better monitoring of DKD.

MATERIALS AND METHODS

Study design and participants

The 194 serum samples were collected spanning five DKD stages. As these two groups of samples were subjected to untargeted metabolites analysis at different times, we analyzed them separately to remove batch effect and separately referred them as the discovery and validation sets (Figure 1). Participants were required to cease taking unnecessary medications and fasted for 8 hours before serum collection. We examined all participants' clinical parameters, reviewed and recorded their medical history, medical complications and dietary habits (see Supplementary Table 1). Based on the ADA and KDIGO criteria [33, 38] for diagnosis of diabetes, participants were firstly classified into healthy group (stage 0) and diabetic group. The healthy and diabetic groups are age and sex matched (Supplementary Table 2). We followed the MDRD formula: "eGFR $(mL/min/1.73 \ m^2) = 186 \times (serum \ creatinine)^{-1.154}$

$\times$ (age in years)$^{-0.203} \times 0.742$ (if female) $\times 1.210$ (if African American)" to calculate the eGFR values [39]. According to the eGFR levels, diabetic individuals were further stratified to five types, which are stage 1a-b and stage 2 to 4 (criteria see Supplementary Materials and Methods); the healthy group participants were regarded as stage 0. The renal resistive index (RRI) was calculated as (peak systolic velocity - end diastolic velocity)/peak systolic velocity derived from the kidney doppler ultrasonography. Additional methods for measuring eGFR values were shown in Supplementary Table 12 [22]. All the DKD-related clinical parameters were measured in The Fourth Clinical Medical College of Guangzhou University of Chinese Medicine (Shenzhen Traditional Chinese Medicine Hospital) followed by the standard procedures.

Measurements of serum untargeted metabolites

Serum samples and an equal volume of quality control (QC) samples (Supplementary Table 3) were deproteinated with cold methanol that contains internal standards. Ultra-Performance Liquid Chromatography-Orbitrap-Mass Spectrometry (UPLC-Orbitrap-MS) analysis was conducted on a Waters ACQUITY UPLC system coupled to a Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer for mass spectrometry (MS) analysis. For detailed steps of UPLC-Orbitrap-MS, please see the Supplementary Materials and Methods.

Untargeted metabolites analysis

UPLC-Orbitrap-MS data from the discovery and validation cohorts were analyzed separately. Data were firstly processed by Progenesis QI 2.3 software (Nonlinear Dynamics, Waters, Milford, MA, USA) for peak detection and alignment, and then subjected to Matlab (MathWorks, Natick, MA, USA) for exclusion of unreliable features with missing rates >40% and missing value imputation [40, 41]. We performed baseline correction via cubic spline interpolation to align the baseline levels of data obtained at different times [42] (Supplementary Figures 10–84). Unstable signals with a coefficient of variation (CV%) >30% across the QC samples were filtered out. Compounds were identified upon matching their mass to charge ratio (m/z) and mass fragmentation patterns against available reference standards and Human Metabolome Database (hmdb.ca) (Supplementary Table 13) [43]. To gain a unique view of DKD, only metabolites that repeat the same mass fragmentation pattern, retention time and show the same trend of significant statistical differences in both discovery and validation sets were kept for further investigation.

Evaluation of CDBs' performance on DKD disease stage classification using random forest

Random forest (RF) algorithm is applicable for evaluating the performance of metabolites on differentiating disease stages [44]. We used the RF package scikit-learn [45] that was implemented by Python (version 3.8) to evaluate the classification power of candidate DKD biomarkers (n_estimators = 10–100). Area under curve (AUC) using 1–9 metabolite models was determined by RF, respectively. To avoid overfitting, samples in each cohort were randomly and evenly divided into training and testing sets for the establishment and performance evaluation of the models. This step was repeated 100 times to obtain the mean AUC values using the testing set of the two cohorts.

Significant altered metabolites analysis

One-way ANOVA followed by Fisher's LSD post-hoc test was used to identify significant altered metabolites (SAMs). Metabolites with p-value < 0.05 and false discovery rate (FDR) < 0.1 among any two of the stages were regarded as SAMs.

Metabonomic networks analysis

Networks of metabolites (Pubmed ID listed in Supplementary Table 13) were generated through MetaMapp [46] with default parameters. CytoScape [47] was used to visualize the networks. Their metabolic pathway output was generated on the basis of their KEGG reaction pairs while their chemical and structural relationships were constructed by their Tanimoto similarity.

Linear regression of CDBs for log(eGRF) estimation

Using CDBs and covariates such as sex, age, SBP and UACR, univariate linear regression (ULR) and multivariate linear regression (MvLR) were utilized to determine log(eGFR). Discovery group was used as the training set. The unstandardized regression coefficients (β) of the training set's model were applied to generate equations for log(eGFR) prediction in testing set -- the validation cohort. Variables that contributed to the model with $p < 0.05$ were selected by stepwise linear regression analysis for the best model in favor of a simpler model.

Follow-up study and targeted metabolites prognostic assessment

To identify the progression of DKD, we recruited an extra cohort of 106 patients at stage 2 and 3 for an average 3.9 years of follow-up study. Patients' serums were collected since the starting point and the sCr, Bun and eGFR were measured every 3 months. The serum samples collected at the first time were subjected to targeted metabolomic analysis for the quantification of ADT, SAdo and pseudouridine by an ultra-high performance liquid chromatography (Shimadzu, Kyoto, Japan) coupled with the AB SCIEX Q-Trap 5500 triple quadrupole mass spectrometer (AB SCIEX, Toronto, Canada). During the follow-up period, patients who remained at their original stage were regarded as the "unprogressed" group; those who progressed to later stages accompanying with a 25% drop in eGFR [48] were "progressed" group. Random forest algorithm [20] were used to evaluate covariates that associated with DKD progression. The prognostic power of biomarkers during DKD stage progression was assessed via performing 100 iterations for each variate. Logistic regression analysis was used to evaluate the risk scores of interested metabolites by SPSS. The average AUC values were calculated and compared among all the variates. Detailed methods of targeted metabolite analysis with UPLC-QQQ-MS/MS and random forest analysis can be found in the Supplementary Materials and Methods (Supplementary Tables 14–16).

Data availability

The raw data were submitted to National Genomics Data Center database under the bioproject PRJCA013833 (https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA0138 33). All the data analyzed during the current study are available from the corresponding author on reasonable request.

Abbreviations

β2-MG: β2-microglobulin; BSA: body surface-area; CKD: chronic kidney disease; DKD: diabetic kidney disease; ESRD: end-stage renal disease; FBG: fasting blood glucose; QC: quality control; RRI: renal resistive index; SBP: systolic blood pressure; L,L-TMAP: N,N,N-trimethyl-L-alanyl-L-proline betaine; RMSEs: root mean square errors; UPLC-Orbitrap-MS: Ultra-Performance Liquid Chromatography-Orbitrap-Mass Spectrometry; UACR: urinary albumin-to-creatinine ratio; SAM: significantly altered metabolites; ADT: 2-(α-D-mannopyranosyl)-L-tryptophan; SAdo: succinyladenosine; sCr: serum creatinine; MS-sCr: Mass Spectrometry detected serum creatinine; T2D: type 2 diabetes; T1D: type 1 diabetes; CDBs: candidate DKD biomarkers; PCA: partial correlation analysis; BeGFR: CDB-predicted eGFR; SD: standard deviation.

AUTHOR CONTRIBUTIONS

N. G., C.C. and D.M. conceived and designed the study. N.G., W.L., and X.G. collected the samples and

acquired clinical data. T.S., N.G. and N.D. conducted statistical analysis. J.J., X.G. and T.S. performed metabolomics experiments, interpreted the data, made the figures and wrote the manuscript. N.D. conducted the metabolomics data preprocessing. J.J., T.S., N.G., X.G., D.M., S.L., G.S. and C.C. revised the manuscript. All authors approved the final version of the manuscript.

CONFLICTS OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ETHICAL STATEMENT AND CONSENT

The use of protocol for this study and informed consent forms were approved by the Research Ethics Board at the Shenzhen Hospital of Guangzhou University of Chinese Medicine (Shenzhen Traditional Chinese Medicine Ethics (Research) (2016) No. 8). The authors declared that all the procedures were in accordance with the approved guidelines, and written informed consent was obtained from all subjects prior to this study.

REFERENCES

1. Gheith O, Farouk N, Nampoory N, Halim MA, Al-Otaibi T. Diabetic kidney disease: world wide difference of prevalence and risk factors. J Nephropharmacol. 2015; 5:49–56.
PMID:28197499

2. Ostermann M, Zarbock A, Goldstein S, Kashani K, Macedo E, Murugan R, Bell M, Forni L, Guzzi L, Joannidis M, Kane-Gill SL, Legrand M, Mehta R, et al. Recommendations on Acute Kidney Injury Biomarkers From the Acute Disease Quality Initiative Consensus Conference: A Consensus Statement. JAMA Netw Open. 2020; 3:e2019209.
https://doi.org/10.1001/jamanetworkopen.2020.19209
PMID:33021646

3. Clish CB. Metabolomics: an emerging but powerful tool for precision medicine. Cold Spring Harb Mol Case Stud. 2015; 1:a000588.
https://doi.org/10.1101/mcs.a000588
PMID:27148576

4. Wang Y, Xia F, Wan H, Chen C, Chen Y, Zhang W, Wang N, Lu Y. Metabolites in the association between early-life famine exposure and type 2 diabetes in adulthood over a 5-year follow-up period. BMJ Open Diabetes Res Care. 2021; 9:e001935.
https://doi.org/10.1136/bmjdrc-2020-001935
PMID:33888542

5. Yang SJ, Kwak SY, Jo G, Song TJ, Shin MJ. Serum metabolite profile associated with incident type 2 diabetes in Koreans: findings from the Korean Genome and Epidemiology Study. Sci Rep. 2018; 8:8207.
https://doi.org/10.1038/s41598-018-26320-9
PMID:29844477

6. Huang J, Huth C, Covic M, Troll M, Adam J, Zukunft S, Prehn C, Wang L, Nano J, Scheerer MF, Neschen S, Kastenmüller G, Suhre K, et al. Machine Learning Approaches Reveal Metabolic Signatures of Incident Chronic Kidney Disease in Individuals With Prediabetes and Type 2 Diabetes. Diabetes. 2020; 69:2756–65.
https://doi.org/10.2337/db20-0586
PMID:33024004

7. Tsai IT, Wu CC, Hung WC, Lee TL, Hsuan CF, Wei CT, Lu YC, Yu TH, Chung FM, Lee YJ, Wang CP. FABP1 and FABP2 as markers of diabetic nephropathy. Int J Med Sci. 2020; 17:2338–45.
https://doi.org/10.7150/ijms.49078
PMID:32922199

8. Bagshaw SM, Bellomo R. Early diagnosis of acute kidney injury. Curr Opin Crit Care. 2007; 13:638–44.

https://doi.org/10.1097/MCC.0b013e3282f07570
PMID:17975383

9. Rigalleau V, Garcia M, Lasseur C, Laurent F, Montaudon M, Raffaitin C, Barthe N, Beauvieux MC, Vendrely B, Chauveau P, Combe C, Gin H. Large kidneys predict poor renal outcome in subjects with diabetes and chronic kidney disease. BMC Nephrol. 2010; 11:3.
https://doi.org/10.1186/1471-2369-11-3
PMID:20199663

10. Selvin E, Rawlings AM, Grams M, Klein R, Steffes M, Coresh J. Association of 1,5-anhydroglucitol with diabetes and microvascular conditions. Clin Chem. 2014; 60:1409–18.
https://doi.org/10.1373/clinchem.2014.229427
PMID:25200356

11. Kim WJ, Park CY. 1,5-Anhydroglucitol in diabetes mellitus. Endocrine. 2013; 43:33–40.
https://doi.org/10.1007/s12020-012-9760-6
PMID:22847316

12. Boulesteix AL, Janitza S, Kruppa J, König IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdiscip Rev Data Min Knowl Discov. 2012; 2:493–507.
https://doi.org/10.1002/widm.1072

13. Shepard BD. Sex differences in diabetes and kidney disease: mechanisms and consequences. Am J Physiol Renal Physiol. 2019; 317:F456–62.
https://doi.org/10.1152/ajprenal.00249.2019
PMID:31241989

14. Piani F, Melena I, Tommerdahl KL, Nokoff N, Nelson RG, Pavkov ME, van Raalte DH, Cherney DZ, Johnson RJ, Nadeau KJ, Bjornstad P. Sex-related differences in diabetic kidney disease: A review on the mechanisms and potential therapeutic implications. J Diabetes Complications. 2021; 35:107841.
https://doi.org/10.1016/j.jdiacomp.2020.107841
PMID:33423908

15. Mogensen CE. Kidney function and glomerular permeability to macromolecules in early juvenile diabetes. Scand J Clin Lab Invest. 1971; 28:79–90.
https://doi.org/10.3109/00365517109090666
PMID:5093522

16. Tang E, Bansal A, Novak M, Mucsi I. Patient-Reported Outcomes in Patients with Chronic Kidney Disease and Kidney Transplant-Part 1. Front Med (Lausanne). 2018; 4:254.
https://doi.org/10.3389/fmed.2017.00254
PMID:29379784

17. Mogensen CE, Christensen CK, Vittinghus E. The stages in diabetic renal disease. With emphasis on the stage of incipient diabetic nephropathy. Diabetes. 1983 (Suppl 2); 32:64–78.
https://doi.org/10.2337/diab.32.2.s64
PMID:6400670

18. Mogensen CE. Glomerular filtration rate and renal plasma flow in long-term juvenile diabetics without proteinuria. Br Med J. 1972; 4:257–9.
https://doi.org/10.1136/bmj.4.5835.257
PMID:5083884

19. Andrassy KM. Comments on 'KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease'. Kidney Int. 2013; 84:622–3.
https://doi.org/10.1038/ki.2013.243
PMID:23989362

20. Levey AS, Coresh J, Tighiouart H, Greene T, Inker LA. Strengths and limitations of estimated and measured GFR. Nat Rev Nephrol. 2019; 15:784.
https://doi.org/10.1038/s41581-019-0213-9
PMID:31578495

21. Porrini E, Ruggenenti P, Luis-Lima S, Carrara F, Jiménez A, de Vries APJ, Torres A, Gaspari F, Remuzzi G. Estimated GFR: time for a critical appraisal. Nat Rev Nephrol. 2019; 15:177–90.
https://doi.org/10.1038/s41581-018-0080-9
PMID:30518813

22. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J, and CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009; 150:604–12.
https://doi.org/10.7326/0003-4819-150-9-200905050-00006
PMID:19414839

23. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T, Kusek JW, Manzi J, Van Lente F, Zhang YL, Coresh J, Levey AS, and CKD-EPI Investigators. Estimating glomerular filtration rate from serum creatinine and cystatin C. N Engl J Med. 2012; 367:20–9.
https://doi.org/10.1056/NEJMoa1114248
PMID:22762315

24. Lelli V, Belardo A, Timperio AM. From Targeted Quantification to Untargeted Metabolomics. Metabolomics—Methodology and Applications in Medical Sciences and Life Sciences. 2021.
https://doi.org/10.5772/intechopen.96852

25. Liu JJ, Ghosh S, Kovalik JP, Ching J, Choi HW, Tavintharan S, Ong CN, Sum CF, Summers SA, Tai ES, Lim SC. Profiling of Plasma Metabolites Suggests Altered Mitochondrial Fuel Usage and Remodeling of Sphingolipid Metabolism

in Individuals With Type 2 Diabetes and Kidney Disease. Kidney Int Rep. 2016; 2:470–80.
https://doi.org/10.1016/j.ekir.2016.12.003
PMID:29142974

26. Barrios C, Zierer J, Würtz P, Haller T, Metspalu A, Gieger C, Thorand B, Meisinger C, Waldenberger M, Raitakari O, Lehtimäki T, Otero S, Rodríguez E, et al. Circulating metabolic biomarkers of renal function in diabetic and non-diabetic populations. Sci Rep. 2018; 8:15249.
https://doi.org/10.1038/s41598-018-33507-7
PMID:30323304

27. Chen S, Akter S, Kuwahara K, Matsushita Y, Nakagawa T, Konishi M, Honda T, Yamamoto S, Hayashi T, Noda M, Mizoue T. Serum amino acid profiles and risk of type 2 diabetes among Japanese adults in the Hitachi Health Study. Sci Rep. 2019; 9:7010.
https://doi.org/10.1038/s41598-019-43431-z
PMID:31065046

28. Liao X, Liu B, Qu H, Zhang L, Lu Y, Xu Y, Lyu Z, Zheng H. A High Level of Circulating Valine Is a Biomarker for Type 2 Diabetes and Associated with the Hypoglycemic Effect of Sitagliptin. Mediators Inflamm. 2019; 2019:8247019.
https://doi.org/10.1155/2019/8247019
PMID:31827381

29. Sekula P, Goek ON, Quaye L, Barrios C, Levey AS, Römisch-Margl W, Menni C, Yet I, Gieger C, Inker LA, Adamski J, Gronwald W, Illig T, et al. A Metabolome-Wide Association Study of Kidney Function and Disease in the General Population. J Am Soc Nephrol. 2016; 27:1175–88.
https://doi.org/10.1681/ASN.2014111099
PMID:26449609

30. Sekula P, Dettmer K, Vogl FC, Gronwald W, Ellmann L, Mohney RP, Eckardt KU, Suhre K, Kastenmüller G, Oefner PJ, Köttgen A. From Discovery to Translation: Characterization of C-Mannosyltryptophan and Pseudouridine as Markers of Kidney Function. Sci Rep. 2017; 7:17400.
https://doi.org/10.1038/s41598-017-17107-5
PMID:29234020

31. Velenosi TJ, Thomson BKA, Tonial NC, RaoPeters AAE, Mio MA, Lajoie GA, Garg AX, House AA, Urquhart BL. Untargeted metabolomics reveals N, N, N-trimethyl-L-alanyl-L-proline betaine (TMAP) as a novel biomarker of kidney function. Sci Rep. 2019; 9:6831.
https://doi.org/10.1038/s41598-019-42992-3
PMID:31048706

32. Yonemura K, Takahira R, Yonekawa O, Wada N, Hishida A. The diagnostic value of serum concentrations of 2-(alpha-mannopyranosyl)-L-tryptophan for normal renal function. Kidney Int.
2004; 65:1395–9.
https://doi.org/10.1111/j.1523-1755.2004.00521.x
PMID:15086480

33. Levin A, Stevens PE, Bilous RW, Coresh J, De Francisco AL, De Jong PE, Griffith KE, Hemmelgarn BR, Iseki K, Lamb EJ, Levey AS, Riella MC, Shlipak MG, et al. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl. 2013; 3:1–150.
https://doi.org/10.1038/kisup.2012.73

34. Schardijn GH, Statius van Eps LW. Beta 2-microglobulin: its significance in the evaluation of renal function. Kidney Int. 1987; 32:635–41.
https://doi.org/10.1038/ki.1987.255
PMID:3323598

35. Afsar B, Elsurer R. Increased renal resistive index in type 2 diabetes: Clinical relevance, mechanisms and future directions. Diabetes Metab Syndr. 2017; 11:291–6.
https://doi.org/10.1016/j.dsx.2016.08.019
PMID:27594114

36. McArdle Z, Schreuder MF, Moritz KM, Denton KM, Singh RR. Physiology and Pathophysiology of Compensatory Adaptations of a Solitary Functioning Kidney. Front Physiol. 2020; 11:725.
https://doi.org/10.3389/fphys.2020.00725
PMID:32670095

37. Solini A, Manca ML, Penno G, Pugliese G, Cobb JE, Ferrannini E. Prediction of Declining Renal Function and Albuminuria in Patients With Type 2 Diabetes by Metabolomics. J Clin Endocrinol Metab. 2016; 101:696–704.
https://doi.org/10.1210/jc.2015-3345
PMID:26684276

38. American Diabetes Association. Classification and diagnosis of diabetes. Diabetes Care. 2015; 38:S8–16.
https://doi.org/10.2337/dc15-S005
PMID:25537714

39. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, Hogg RJ, Perrone RD, Lau J, Eknoyan G, and National Kidney Foundation. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Ann Intern Med. 2003; 139:137–47.
https://doi.org/10.7326/0003-4819-139-2-200307150-00013
PMID:12859163

40. Armitage EG, Godzien J, Alonso-Herranz V, López-Gonzálvez Á, Barbas C. Missing value imputation strategies for metabolomics data. Electrophoresis. 2015; 36:3050–60.

https://doi.org/10.1002/elps.201500352
PMID:26376450

41. Wei R, Wang J, Su M, Jia E, Chen S, Chen T, Ni Y. Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. Sci Rep. 2018; 8:663.
https://doi.org/10.1038/s41598-017-19120-0
PMID:29330539

42. van der Kloet FM, Bobeldijk I, Verheij ER, Jellema RH. Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. J Proteome Res. 2009; 8:5132–41.
https://doi.org/10.1021/pr900499r
PMID:19754161

43. Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, Dizon R, Sayeeda Z, Tian S, Lee BL, Berjanskii M, Mah R, Yamamoto M, et al. HMDB 5.0: the Human Metabolome Database for 2022. Nucleic Acids Res. 2022; 50:D622–31.
https://doi.org/10.1093/nar/gkab1062
PMID:34986597

44. Kokla M, Virtanen J, Kolehmainen M, Paananen J, Hanhineva K. Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: a comparative study. BMC Bioinformatics. 2019; 20:492.
https://doi.org/10.1186/s12859-019-3110-0
PMID:31601178

45. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011; 12:2825–30.

46. Barupal DK, Haldiya PK, Wohlgemuth G, Kind T, Kothari SL, Pinkerton KE, Fiehn O. MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. BMC Bioinformatics. 2012; 13:99.
https://doi.org/10.1186/1471-2105-13-99
PMID:22591066

47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–504.
https://doi.org/10.1101/gr.1239303
PMID:14597658

48. KDIGO KDIGOCWG: Clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl. 2012; 2013:1–150.

SUPPLEMENTARY MATERIALS

Sample preparation and UPLC-Orbitrap-MS conditions

Recruitment of participants

At study enrollment stage, a designed baseline clinical examination of each participant and structured interview were performed by trained recruiters. The exclusion criteria were acute renal failure, rapidly increasing proteinuria or nephrotic syndrome, refractory hypertension, serious infections, signs or symptoms of other systemic disease, known renal tubular acidosis, pregnancy, type 1 diabetes, gestational diabetes, chronic liver disease, serious cardiovascular diseases, alcoholics or malignancy. For healthy control, they had not received any treatment like antibiotics, probiotics and hormone therapy in the past two months, did not have proteinuria or history of kidney disease and their oral glucose tolerance test and other related clinical test are in normal levels. All baseline clinical information is shown in Supplementary Tables 1 and 2.

Criteria of stages' classification for participants

Based on the MDRD eGFR values, participants were classified into five stages by following criteria: Stage 1 (eGFR $\geq$ 90 ml/min/1.73 m^2) which consists of Stage 1a (eGFR $\geq$ 120 mL/min/1.73 m^2) and Stage 1b (eGFR within 90–120 mL/min/1.73 m^2); Stage 2, 60–89 mL/min/1.73 m^2; Stage 3, 30–59 mL/min/1.73 m^2; Stage 4, < 30 mL/min/1.73 m^2 [1, 2].

Serum preparation

60 µL serum was deproteinated with 240 µL cold methanol containing 0.5 ppm L-tryptophan (indole-D$_5$, 98%, Cambridge Isotope Laboratories, Tewksbury, MA, USA) and 0.5 ppm cholic acid-2,2,4,4-D$_4$ and 50 ppm C19:1n9c. They were vortexed for 1 min and stood at −20°C overnight for complete deproteination. Then, they were centrifuged at 18700 × g for 20 min. 250 µL supernatant was collected and dried under nitrogen gas and stored at −80°C. The dried supernatant was reconstituted with initial UPLC gradient (5% acetonitrile in water), vortexed for 30 s and was centrifuged at 18700 × g for 20 min. The supernatant was transferred to a glass insert in an amber HPLC vial prior to UPLC-Orbitrap-MS analysis or UPLC-QQQ-MS/MS analysis.

Standard solution and quality control sample preparation

For targeted metabolites quantitation, standards of selected metabolites were purchased from Toronto Research Chemicals (North York, Toronto, Canada) and

Sigma-Aldrich (St. Louis, MO, USA), which were used for preparation of standard solutions. Standard solutions were gradient diluted into ten levels, respectively (Supplementary Table 14). The gradient diluted standard solutions were mixed and then dried by TurboVap® blowdown evaporator (Biotage Sweden AB, Ystrad Mynach, United Kingdom) for later use. 20 µL aliquots from each sample of all groups were mixed and aliquoted as QC samples. QC samples were injected between every six-sample injections to monitor the stability of the instruments throughout the UPLC-Orbitrap-MS signal acquisition. The order of injection for all samples was randomized. Recovery rate of detected metabolites was calculated through parallel serum samples spiked with a known amount of each metabolite standard at three concentration levels. Recovery rate equation: ((Detected concentration − endogenous blank sample concentration) × 100%)∕spiked concentration. (Supplementary Table 15).

UPLC condition

UPLC-Orbitrap-MS analysis
3 µL aliquot was injected into a Waters ACQUITY UPLC system. UPLC separation was performed on a Waters ACQUITY UPLC HSS T3 column (2.1 mm × 100 mm, 1.8 µm) with HSS T3 guard column (2.1 mm × 5 mm, 1.8 µm, Waters Corporation, Milford, MA, USA). The mobile phase consisted of combinations of A (0.1% formic acid in water, v/v) and B (0.1% formic acid in acetonitrile, v/v) at a flow rate of 0.3 mL/min with elution gradient as follows: 0–1.5 min, 5% B; 2 min, 35% B; 4 min, 50% B; 8 min, 55% B; 11–14 min, 95% B. A 3-min post-run time was set to fully equilibrate the column. Column and sample chamber temperature were 40°C and 4°C respectively.

UPLC-QQQ-MS/MS analysis
2 µL aliquot was injected into a SHIMADZU A30 UPLC system. Chromatographic separation was performed on the Luna Omega 1.6 µm Polar C18 reversed-phase column (Phenomenex, Torrance, CA, USA) with Polar C18 security guard column (2.1 mm, Phenomenex, Torrance, CA, USA). The mobile phase A (0.1% formic acid in ultrapure water, v/v) and mobile phase B (100% acetonitrile) were delivered at 0.3 ml/min. Gradient elution was as follows: 2–60% B at 0–3.2 min, 60% maintained at 3.21–3.5 min, 2% B at 3.51–5 min to equilibrate the column before a new injection. Column and sample chamber temperature were 40°C and 4°C respectively.

Mass spectrometry condition

UPLC-Orbitrap-MS analysis
Mass spectrometry analysis was conducted by a Thermo Scientific Orbitrap Fusion Lumos Tribrid

mass spectrometer equipped with a heated electrospray ionization (H-ESI) interface (Thermo Fisher Scientific, Waltham, MA, USA). The mass-spectrometric parameters were set as follows: spray voltage, 2300 V and 3500 V in ESI negative and positive ionization modes respectively; ion transfer tube and vaporizer temperature, 300°C. Nitrogen gas was used as the sheath gas and the aux gas with a flow rate of 25 and 10 L/min, respectively. The analyzer was operated in a data-dependent acquisition mode, with full MS scans of mass range at 90–1000 m/z with detection in the Orbitrap (120000 resolution) and with auto gain control targeted at 20000 count and a maximum injection time at 100 ms.

UPLC-QQQ-MS/MS analysis

Selected metabolites were detected under positive ion multiple reaction monitoring (MRM) mode. Turbo ion spray source was set at a source temperature of 500°C. Ion spray voltage was 5500 V, Ion Source Gas1 (GS1) and Ion Source Gas2 (GS2) had a flow of 50 psi, the curtain gas had a flow of 25 psi, the CAD gas setting was 'medium', and the declustering potential was optimized one by one according to the metabolite. Q1/Q3 mass and MRM conditions for each metabolite were listed in Supplementary Table 16.

Baseline correction

Batch correction was then performed by smoothing through QC samples in sequential injections using cubic splines, a very flexible smoother that can catch the variations of ion abundances caused by the systematic bias in instrumental responses, with a very wide range of curve shapes (e.g., linear, nonlinear) (van der Kloet et al. 2009) to ([3–5]). The penalty for smoothing spline was set to 0.01, which was found to be fitted well to the variations (Supplementary Figures 10–84). The ion abundance for metabolite i at k^{th} injection after batch correction ($x'_{k,i}$) then becomes

$$x'_{k,i} = C_{QC,i}, r_{k,i}, \text{ where } r_{k,i} = \frac{x_{k,i}}{f_{k,i}}$$

where $C_{QC,i}$ was the true concentration of metabolite i in QC samples, which served as a scaling factor to map the corrected ion abundance $r_{k,i}$ to the corrected raw ion abundance ($x'_{k,i}$). However, it is impossible to obtain the true concentration of any metabolite, thus median ion abundance of the metabolite in QC samples could be used as an estimation of $C_{QC,i}$. $x_{k,i}$ and $f_{k,i}$ are observed and fitted raw ion abundance in sample at k^{th} injection. Since results obtained from statistical analysis (e.g., Pearson correlation, Student's t-test) using $x'_{k,I}$ will be the same with those using $r_{k,i}$ as the two kinds of ion abundance only differ in a constant multiplier $C_{QC,i}$, $r_{k,i}$ was used in subsequent statistical analysis instead of the raw ion abundance.

Random forest of metabolite prediction on DKD stage progression in follow up cohort

AUC of variate(s) on prediction DKD stage progression were calculated by random forest (RF). Stratified random sampling was used between progressed and unprogressed group in follow up cohort. Samples were split into a training set (70% of sample size) for modelling and a testing set (the rest 30% samples) for prediction. To avoid overfitting, this stratified random sampling procedure was repeated 100 times and the AUC of testing set was calculated 100 times. Finally, the AUC average and standard deviation were executed to exhibit the performance prediction of metabolites on DKD stage progression.

Supplementary References

1. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, Hogg RJ, Perrone RD, Lau J, Eknoyan G, and National Kidney Foundation. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Ann Intern Med. 2003; 139:137–47.
https://doi.org/10.7326/0003-4819-139-2-200307150-00013
PMID:12859163

2. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T, Kusek JW, Manzi J, Van Lente F, Zhang YL, Coresh J, Levey AS, and CKD-EPI Investigators. Estimating glomerular filtration rate from serum creatinine and cystatin C. N Engl J Med. 2012; 367:20–9.
https://doi.org/10.1056/NEJMoa1114248
PMID:22762315

3. van der Kloet FM, Bobeldijk I, Verheij ER, Jellema RH. Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. J Proteome Res. 2009; 8:5132–41.
https://doi.org/10.1021/pr900499r
PMID:19754161

4. Armitage EG, Godzien J, Alonso-Herranz V, López-González Á, Barbas C. Missing value imputation strategies for metabolomics data. Electrophoresis. 2015; 36:3050–60.
https://doi.org/10.1002/elps.201500352
PMID:26376450

5. Wei R, Wang J, Su M, Jia E, Chen S, Chen T, Ni Y. Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. Sci Rep. 2018; 8:663.
https://doi.org/10.1038/s41598-017-19120-0
PMID:29330539

Supplementary Figures
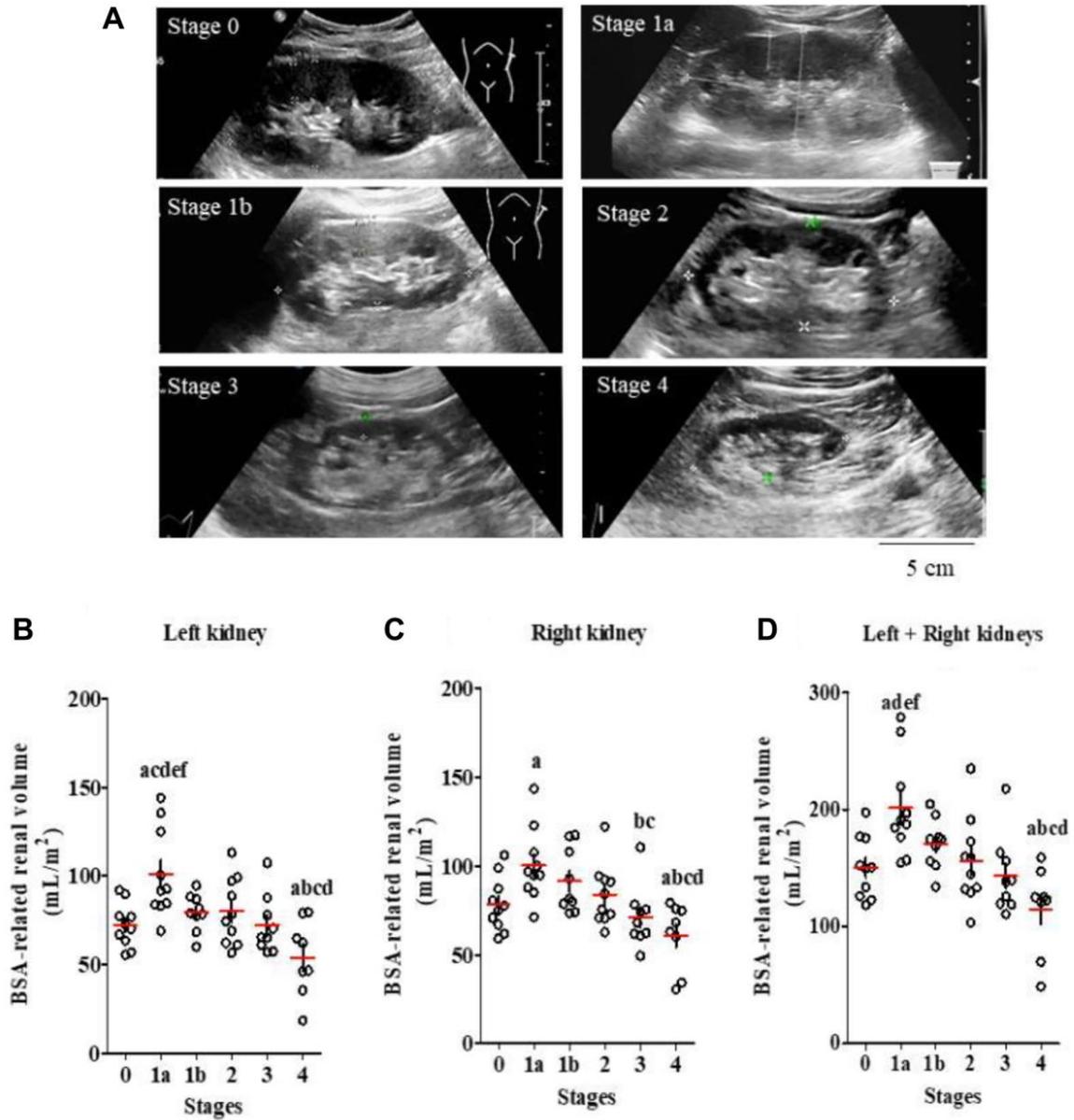
Supplementary Figure 1. (**A**) Images of ultrasonic scanning of the left kidneys in healthy subjects and diabetic patients in different stages and (**B**–**D**) the comparison of BSA-related renal volumes of both kidneys between stages in the validation cohort (significance level for superscript case letters: [a,b,c,d,e,f] represent comparison with Stages 0, 1a, 1b, 2, 3 and 4, respectively): Student's *t*-test, $p \leq 0.05$. Horizontal bars in the scatter plots represent mean ± SEM. *n* = 10 per stage.
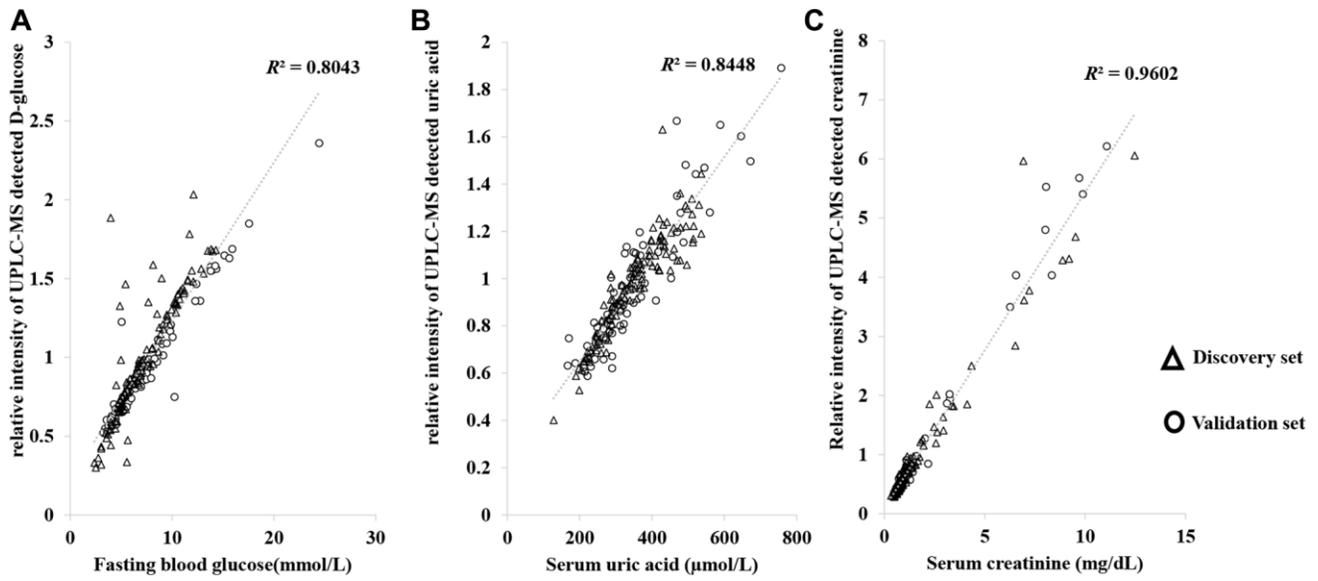
Supplementary Figure 2. Linear relationship of UPLC-Orbitrap-MS detected (**A**) serum D-glucose, (**B**) serum uric acid and (**C**) MS-detected serum creatinine against clinically measured FBG, serum uric acid, serum creatinine, respectively. Trendlines were formed with both cohorts.
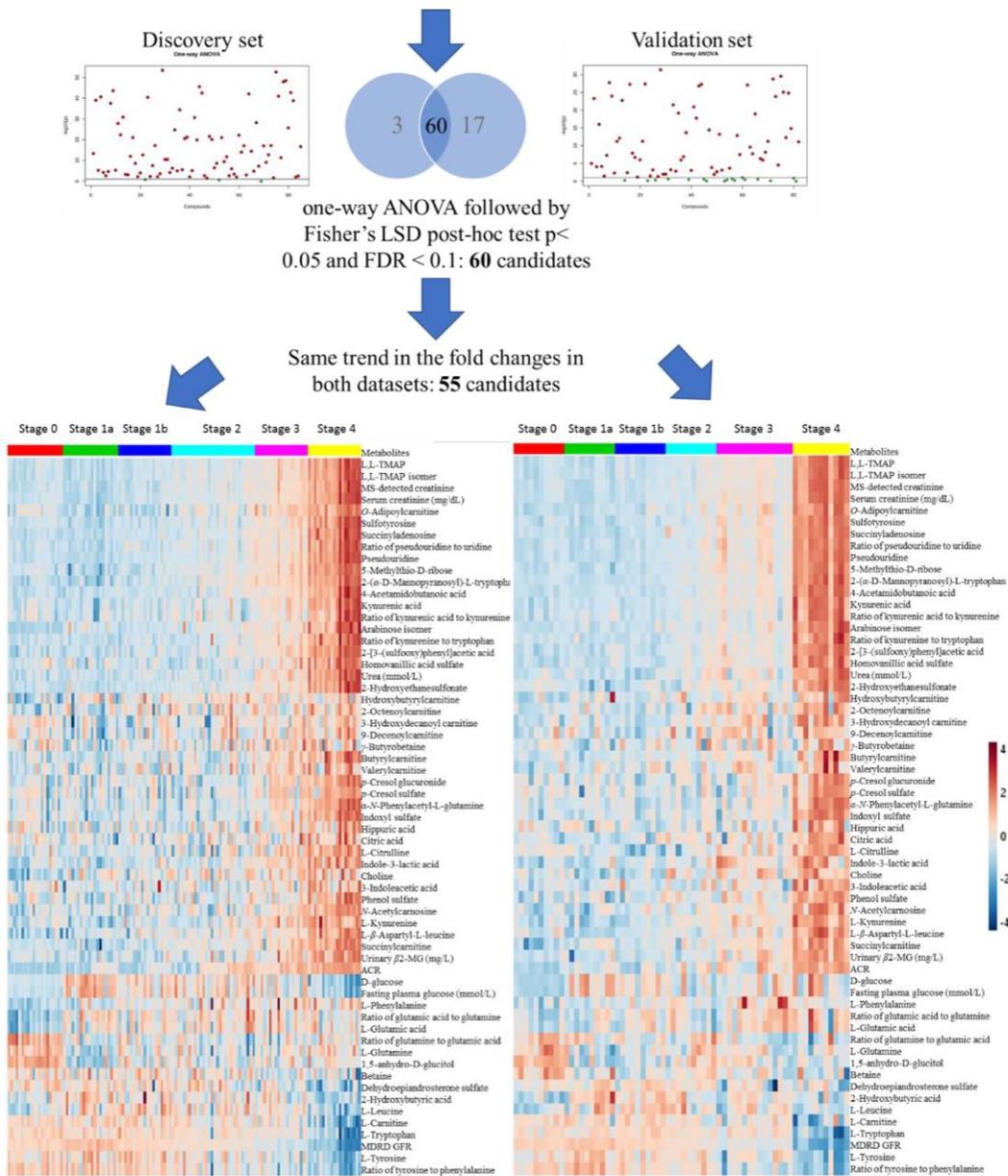
Supplementary Figure 3. Pipeline of screening of potential candidates.
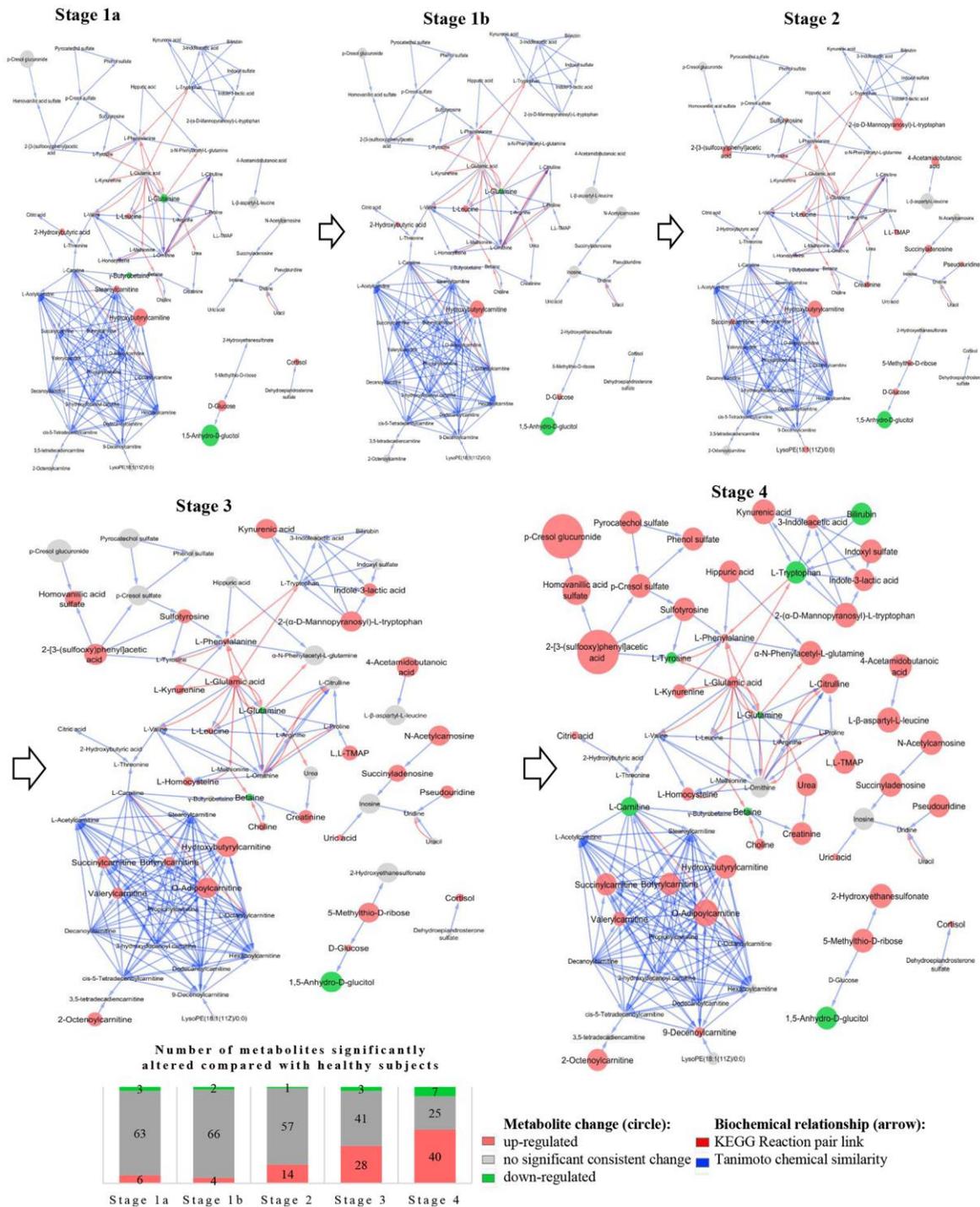
Supplementary Figure 4. Metabolomic network diagrams of SAMs. Metabolites of nucleoside, amino acid, glucose and short-medium-long chain carnitine were accumulated gradually with the progressed DKD. These metabolites were categorized into six classes: sulfate metabolites, amino acids and organic acids, acylcarnitine, purine derivatives, steroids, monosaccharides and their derivatives. Succinyladenosine and pseudouridine, originated from posttranslational modifications of nucleosides, increased significantly in DKD stage 4 patients. Uremic retention solutes like indoxyl sulfate and p-cresyl sulfate, which are products of dietary tryptophan and tyrosine respectively, are significantly elevated at DKD Stage 4. Tryptophan, tyrosine, phenylalanine, glutamine and citrulline are the significantly altered amino acid. Several metabolites decreased along with the declined eGFR, which were tyrptophan, tyrosine, glutamine, 1,5-anhydro-D-glucitol, carnitine, bilirubin and betaine. Red circle: upregulated metabolites; green circle: downregulated metabolites; grey circle denotes metabolites without significant or consistent changes in both cohorts. Circle sizes were proportional to the absolute value of fold change with respect to the normal group ($p \leq 0.05$).
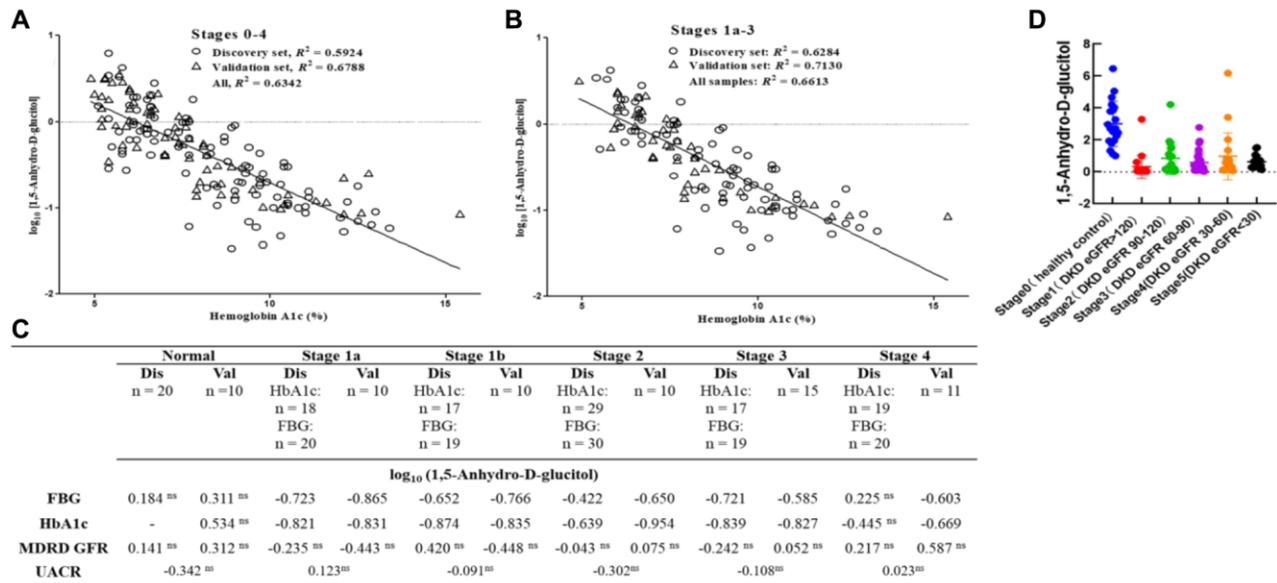
Supplementary Figure 5. Linear regression of UPLC-Orbitrap-MS detected log(1,5-anhydro-D-glucitol) against hemoglobin A1c in (**A**) Stages 0–4 and (**B**) Stages 1a-3 of DKD. (**C**) Correlation between 1,5-anhydro-D-glucitol, hemoglobin A1c, FBG, MS-detected D-glucose, MDRD GFR and UACR in each stage. (**D**) The scatter plots of 1,5-anhydro-D-glucitol among different stages. Abbreviations: Dis: discovery set; Val: validation set; ns: no statistical significance ($p > 0.05$). HbA1c: hemoglobin A1c. -, not available.

A Stages 1-4

MS-detected serum creatinine*
L,L-TMAP*
Pseudouridine*
Succinyladenosine*
2-(α-D-Mannopyranosyl)-L-tryptophan*
5-Methylthio-D-ribose
Sulfotyrosine
4-Acetamidobutanoic acid*
Succinylcarnitine*
2-[3-(sulfooxy)phenyl]acetic acid
2-Hydroxyethanesulfonate*
O-Adipoylcarnitine*
Indole-3-lactic acid*
Kynurenic acid*
N-Acetylcarnosine*
Arabinose isomer
L-Kynurenine*
Butyrylcarnitine*
L-β-Aspartyl-L-leucine
Indoxyl sulfate*
α-N-Phenylacetyl-L-glutamine*
Homovanillic acid sulfate*
p-Cresol glucuronide*
3-Hydroxydecanoyl carnitine
p-Cresol sulfate*
L-Citrulline*
Phenol sulfate
9-Decenoylcarnitine
Valerylcarnitine*
γ-Butyrobetaine*
3-Indoleacetic acid*
L-Carnitine*
L-Tryptophan*
L-Tyrosine*

■ Discovery
■ Validation

-0.6 -0.2 0.2 0.6 1 **Spearman's r**

B Early stages (GFR ≥ 60 mL/min/1.73m²)

MS-detected serum creatinine*
L,L-TMAP*
Pseudouridine*
Succinyladenosine*
2-(α-D-Mannopyranosyl)-L-...
3-Hydroxydecanoyl carnitine
Succinylcarnitine*
N-Acetylcarnosine*
γ-Butyrobetaine*
5-Methylthio-D-ribose
Sulfotyrosine
Indole-3-lactic acid*
9-Decenoylcarnitine
2-Hydroxyethanesulfonate*
O-Adipoylcarnitine*
2-[3-(sulfooxy)phenyl]acetic acid
4-Acetamidobutanoic acid*
Arabinose isomer
L-β-Aspartyl-L-leucine
L-Kynurenine*
Valerylcarnitine*
Kynurenic acid*
Butyrylcarnitine*
Phenol sulfate
Indoxyl sulfate*
L-Tryptophan*
Homovanillic acid sulfate*
p-Cresol sulfate*
α-N-Phenylacetyl-L-glutamine*
L-Citrulline*
3-Indoleacetic acid*
p-Cresol glucuronide*
L-Carnitine*
L-Tyrosine*

■ Discovery
■ Validation

-0.3 0.1 0.5 0.9 **Spearman's r**

*Identification of metabolites with reference standards.

Supplementary Figure 6. List of metabolites that correlated (|r| ≥ 0.6) with eGFR at (**A**) Stages 1–4 and (**B**) early stages (GFR ≥ 60 mL/min/1.73 m²) patients. Metabolites showed strong correlation with eGFR, especially in early diabetic patients.
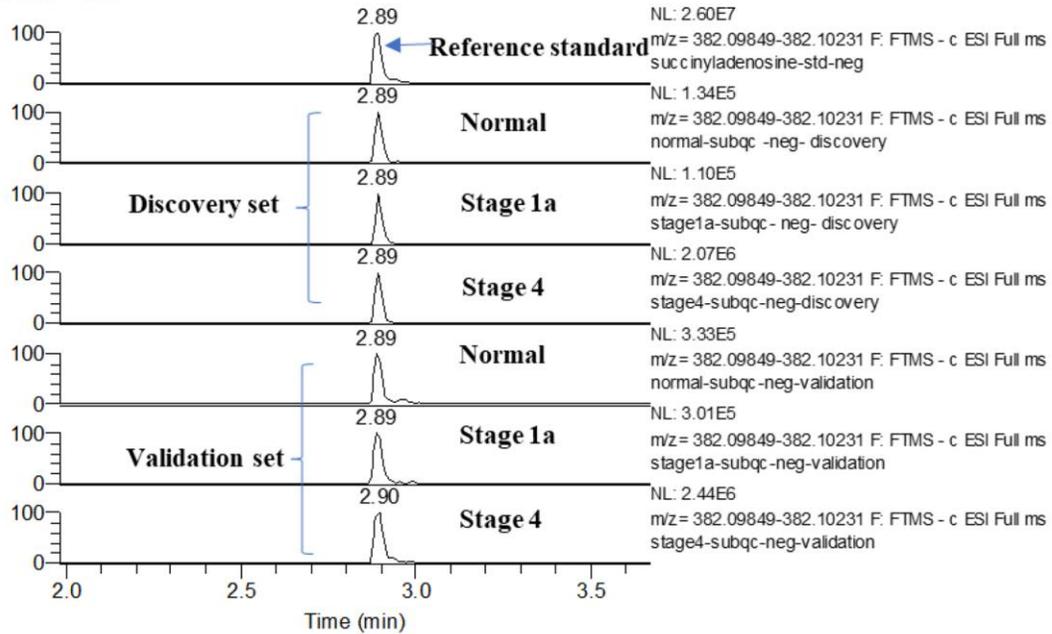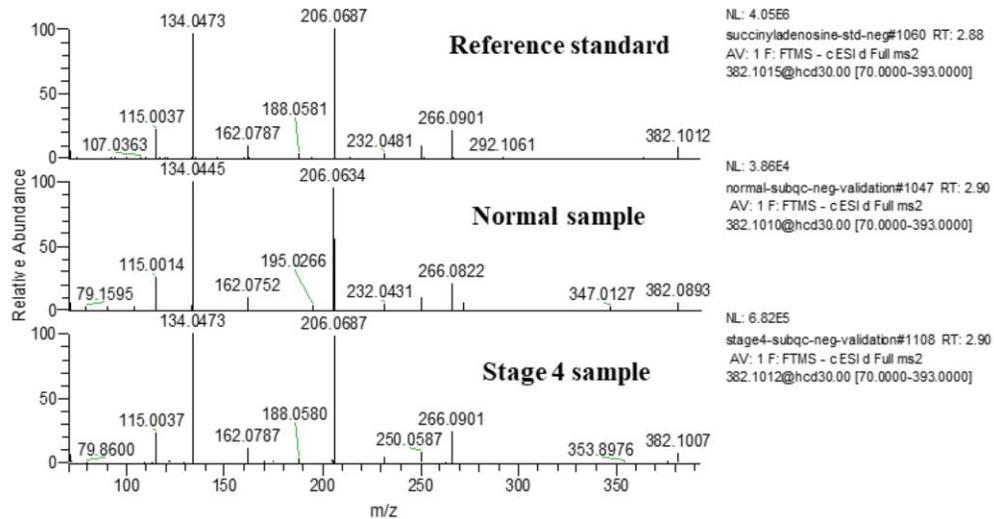
Supplementary Figure 7. Demonstration of peak identification of the adduct (M-H)- of succinyladenosine compared with its reference standard.

Calibration for 2-A-D-L-Try247.1: y = 10369.89374 x + -97.51104 (r = 0.99863, r² = 0.99726) (weighting: 1 / x)

2-(a-D-Manopyranosly)-L-tryptophan   R²=0.99726

Calibration for succinyladenosine252: y = 1.52691e5 x + 8232.42355 (r = 0.99866, r² = 0.99733) (weighting: 1 / x)

succinyladenosin R²=0.99733

Calibration for pseudouridine191: y = 3282.86684 x + 5649.52326 (r = 0.99903, r² = 0.99806) (weighting: 1 / x)

ß-pseudouridine R²=0.99506
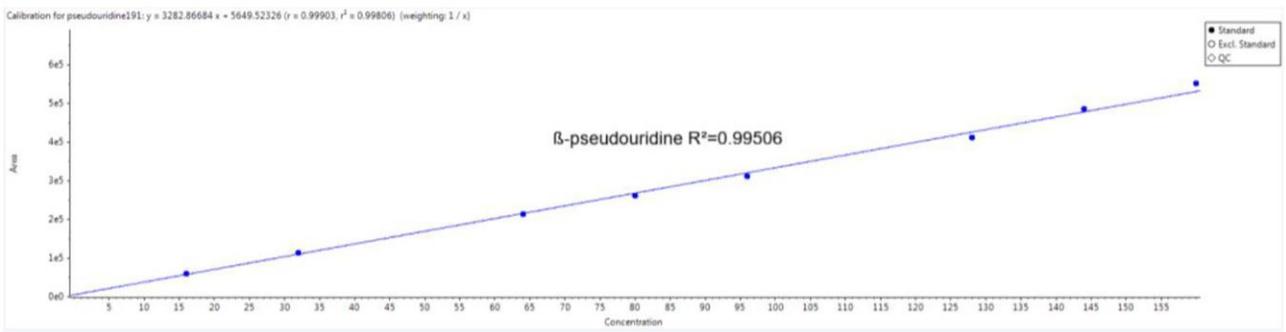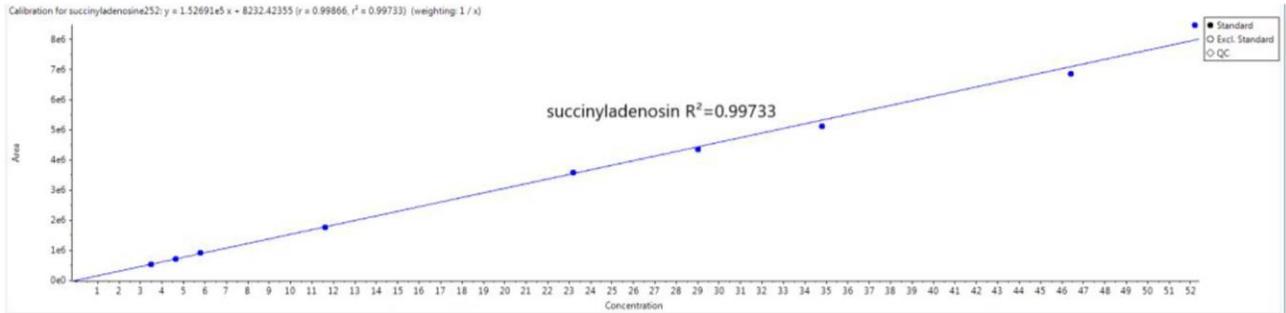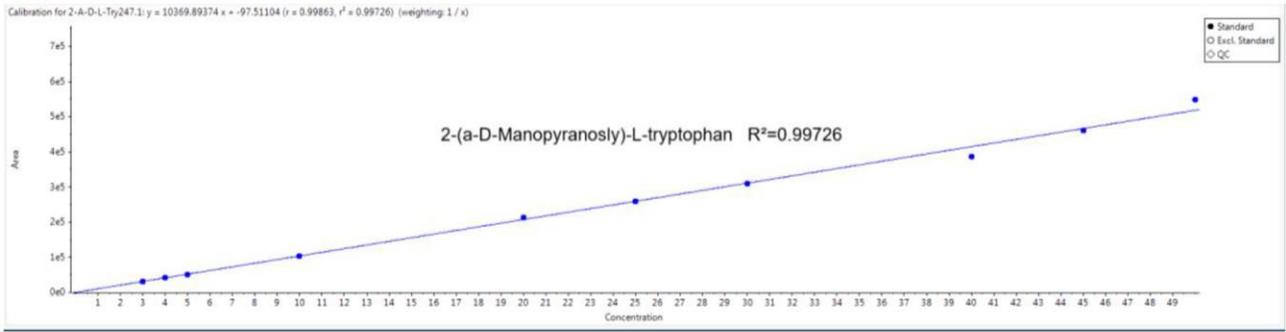
Supplementary Figure 8. Calibration curves of selected metabolites.
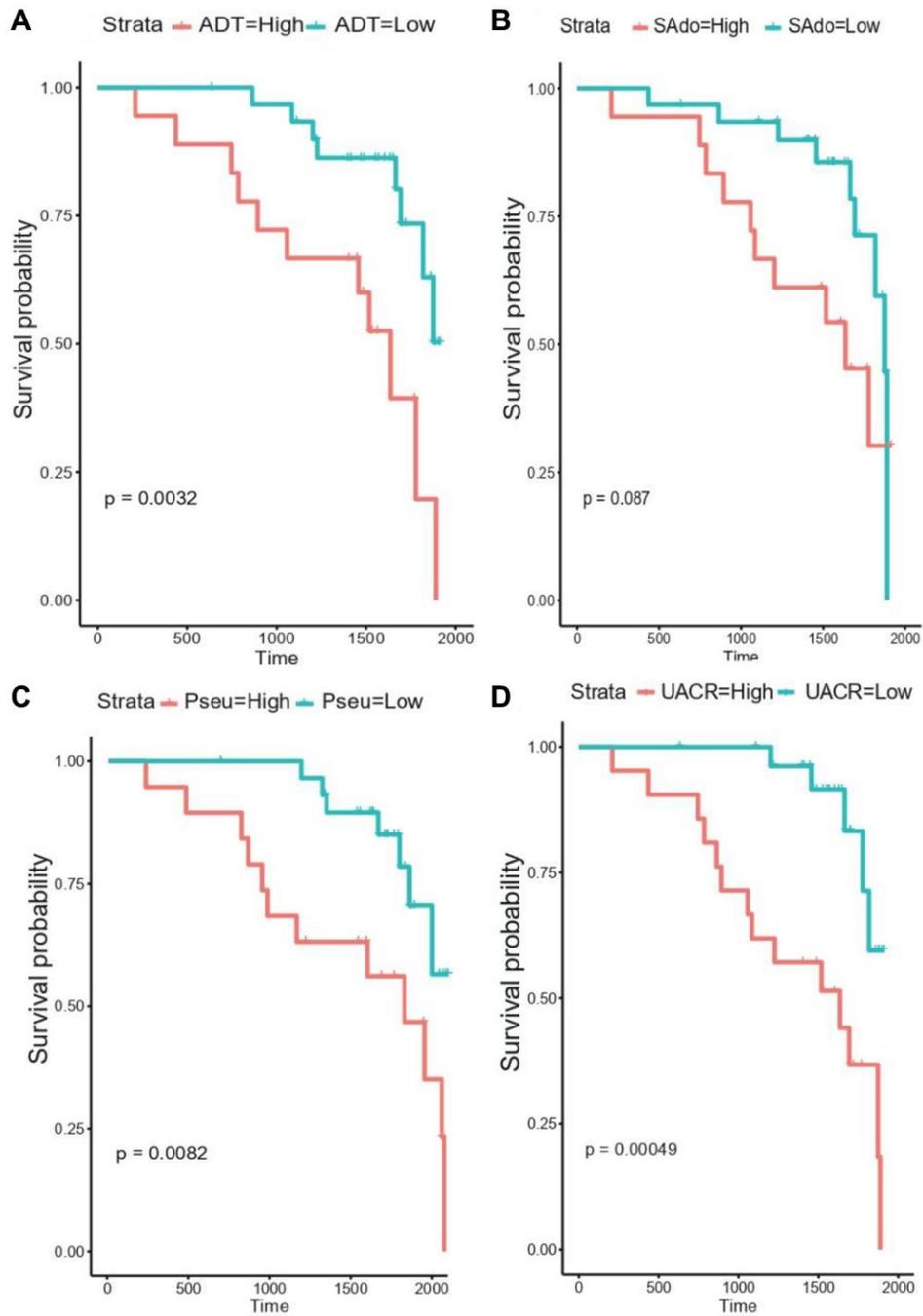
Supplementary Figure 9. Survival curve of follow-up cohort using the baseline levels of three CDBs and UACR. (**A**) ADT; (**B**) SAdo; (**C**) pseudouridine; (**D**) UACR.
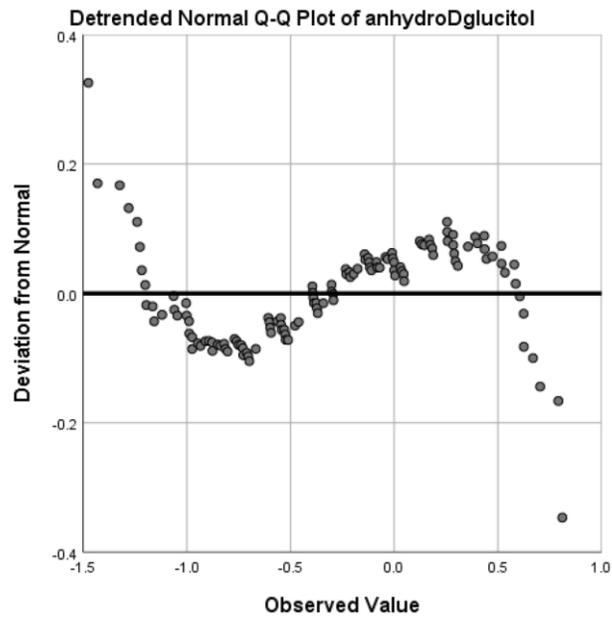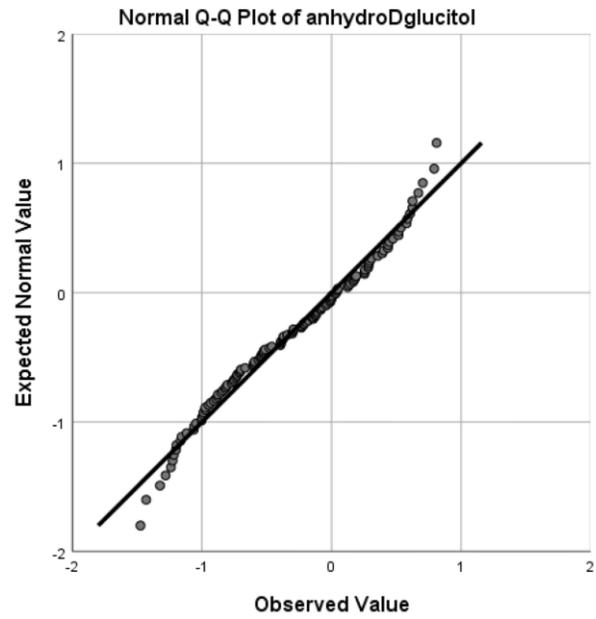
Supplementary Figure 10. The quantile-quantile (Q-Q) plot of normalized anhydroDglucitol which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
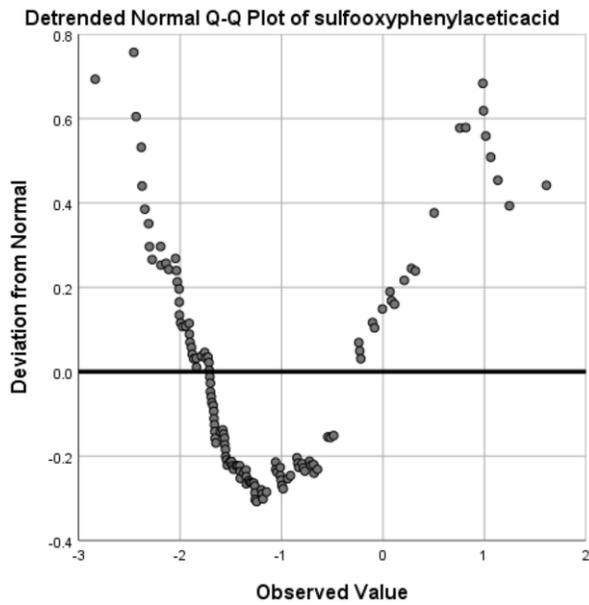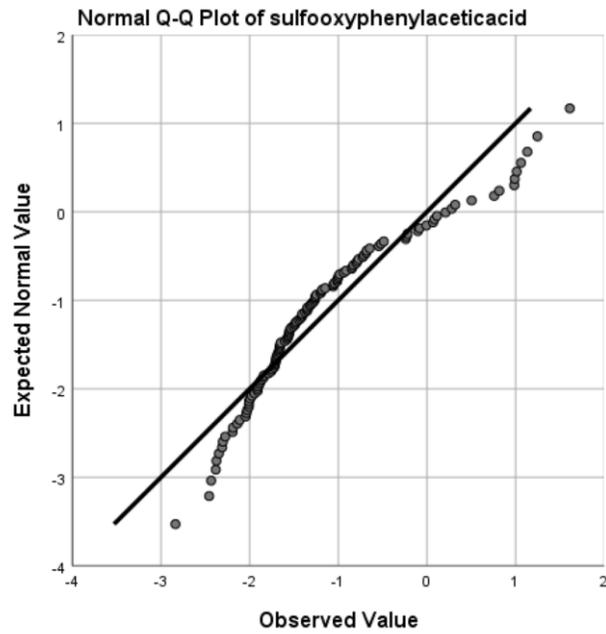
Supplementary Figure 11. The quantile-quantile (Q-Q) plot of normalized sulfooxyphenylaceticacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
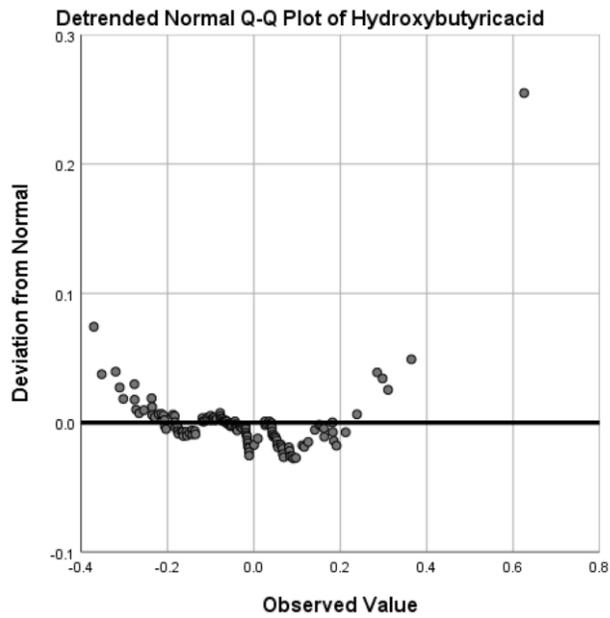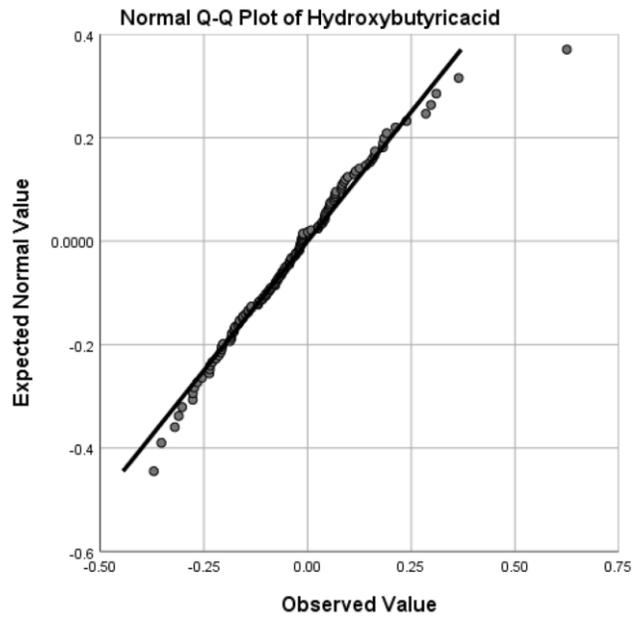
Supplementary Figure 12. The quantile-quantile (Q-Q) plot of normalized Hydroxybutyricacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Hydroxyethanesulfonate

Detrended Normal Q-Q Plot of Hydroxyethanesulfonate
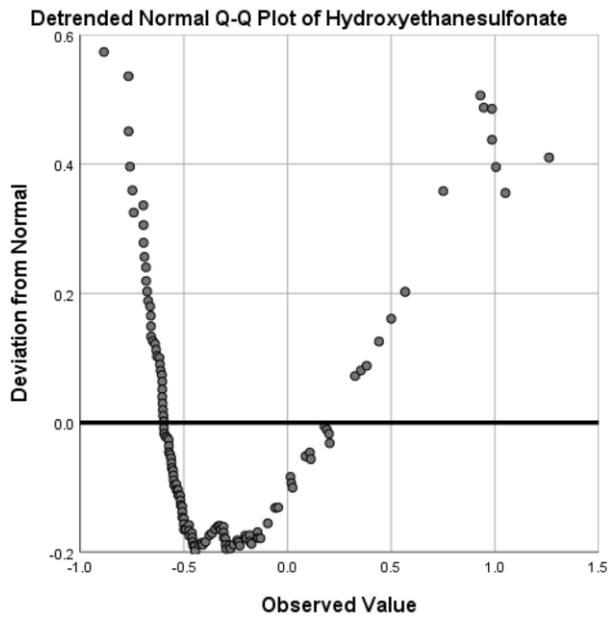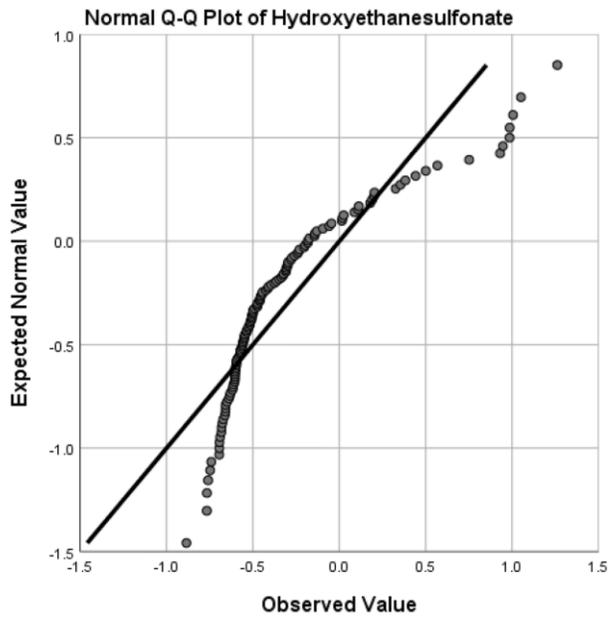
Supplementary Figure 13. The quantile-quantile (Q-Q) plot of normalized Hydroxyethanesulfonate which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Octenoylcarnitine

Detrended Normal Q-Q Plot of Octenoylcarnitine
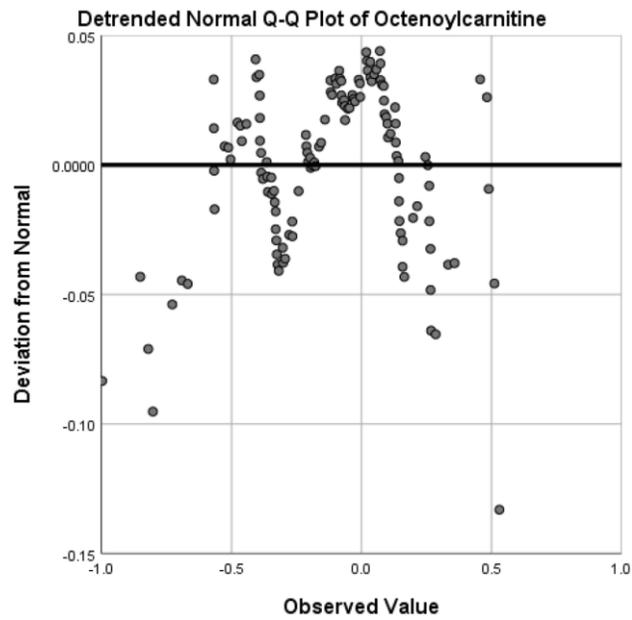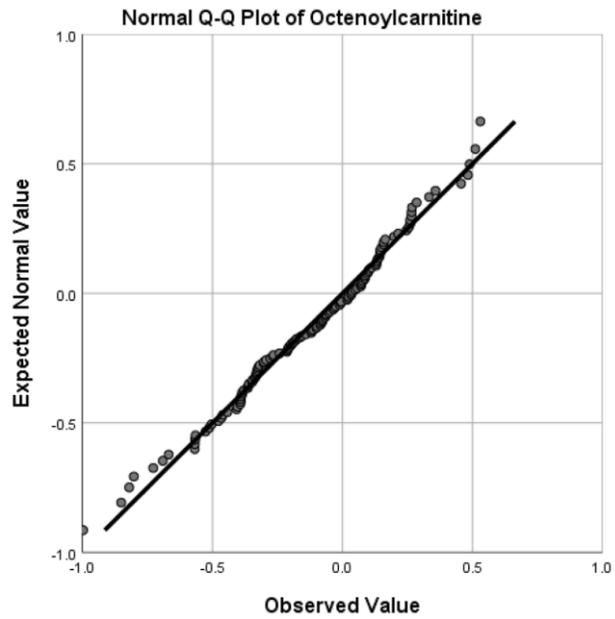
Supplementary Figure 14. The quantile-quantile (Q-Q) plot of normalized Octenoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
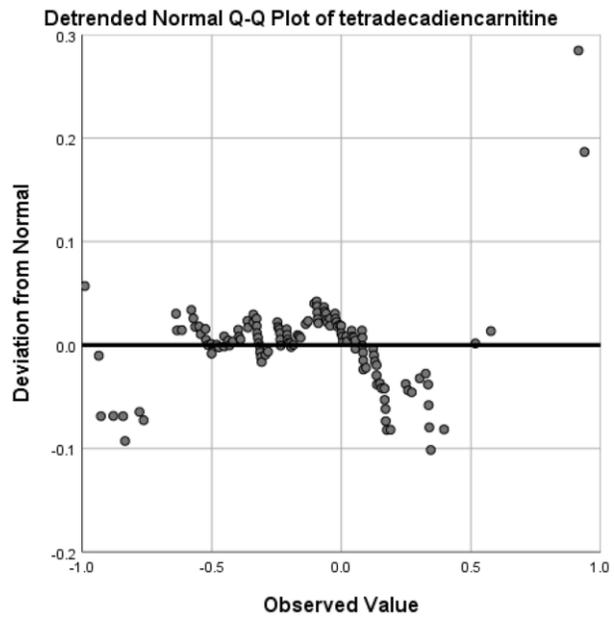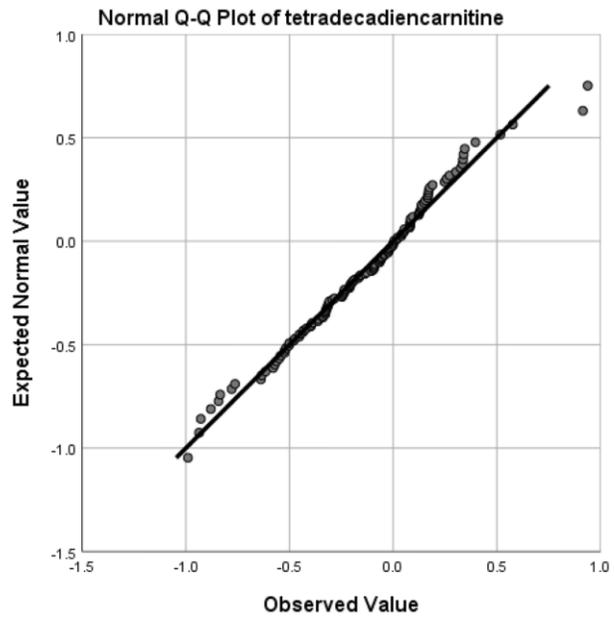
Supplementary Figure 15. The quantile-quantile (Q-Q) plot of normalized tetradecadiencarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of hydroxydecanoylcarnitine

Detrended Normal Q-Q Plot of hydroxydecanoylcarnitine
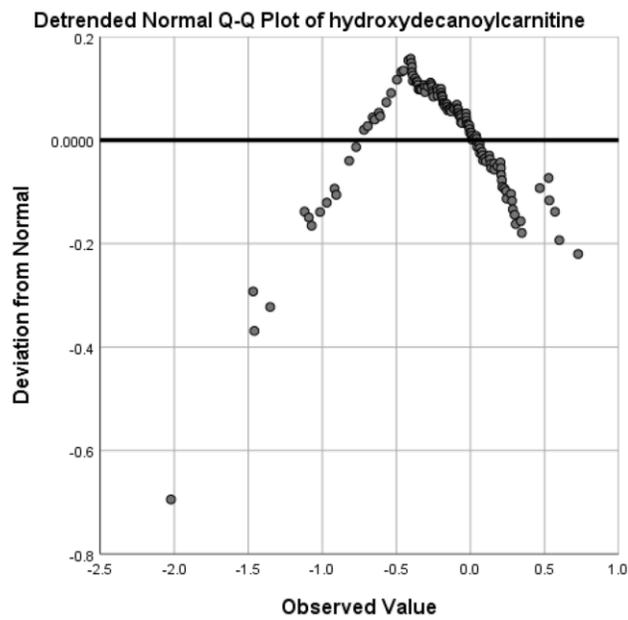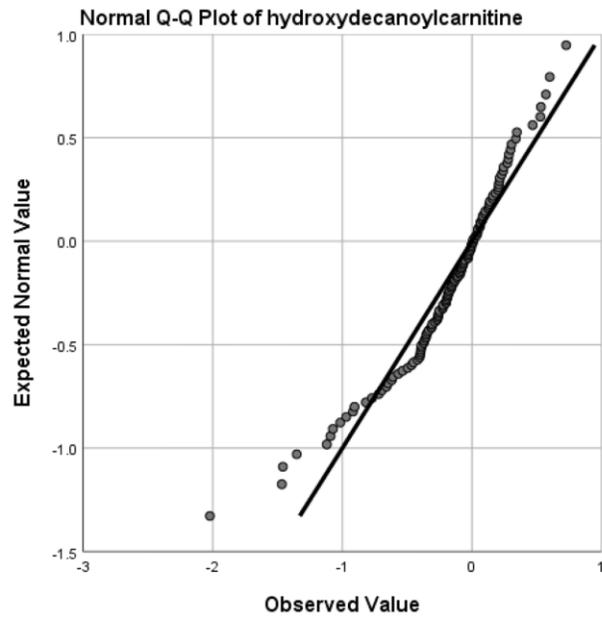
Supplementary Figure 16. The quantile-quantile (Q-Q) plot of normalized hydroxydecanoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
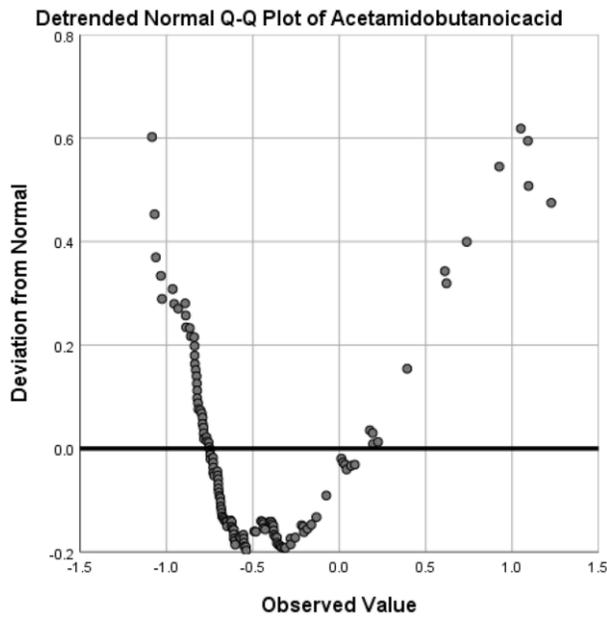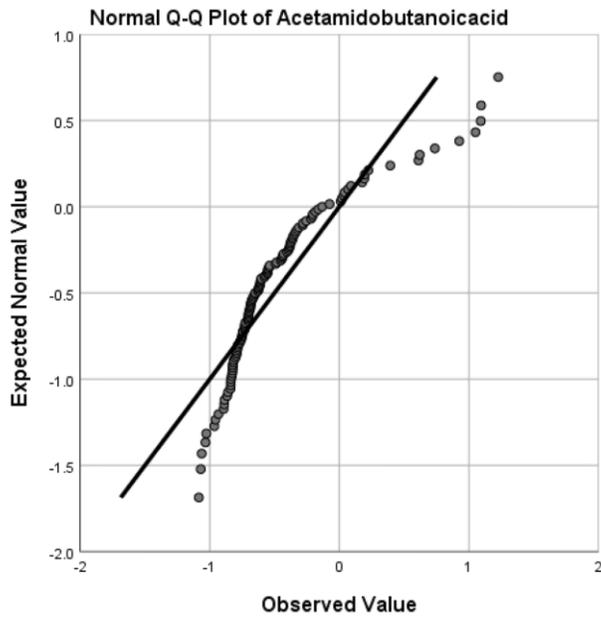
Supplementary Figure 17. The quantile-quantile (Q-Q) plot of normalized Acetamidobutanoicacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
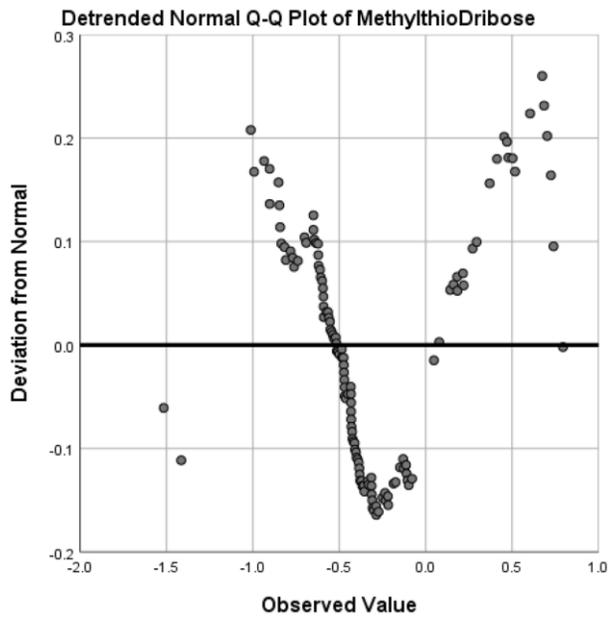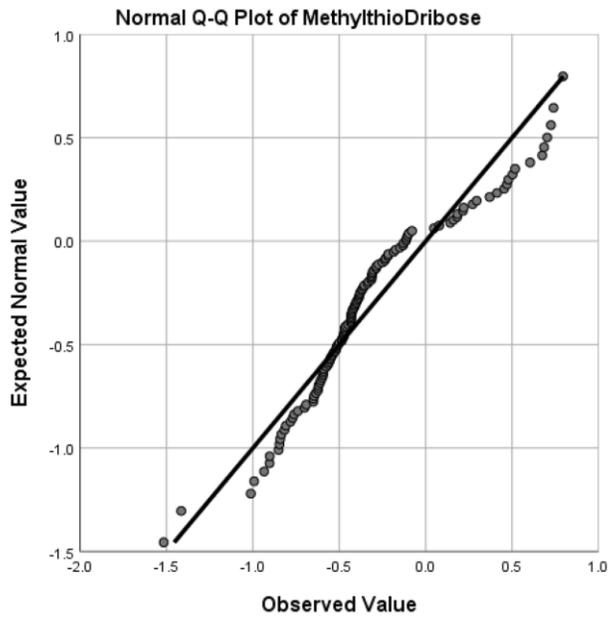
Supplementary Figure 18. The quantile-quantile (Q-Q) plot of normalized MethylthioDribose which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
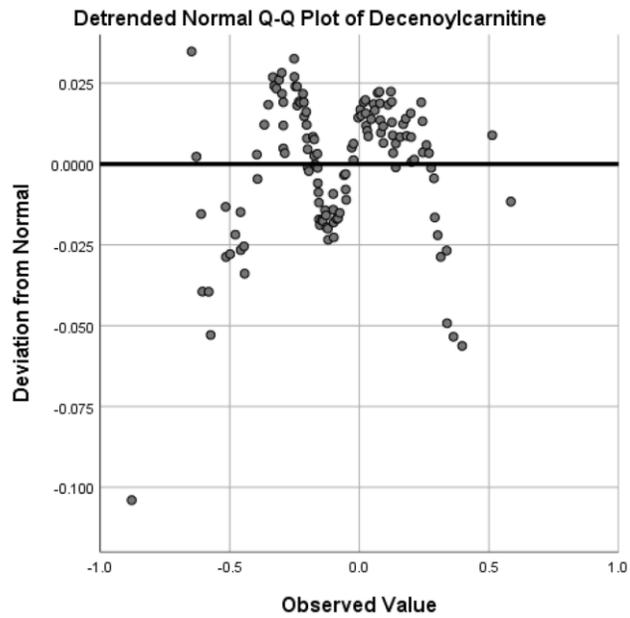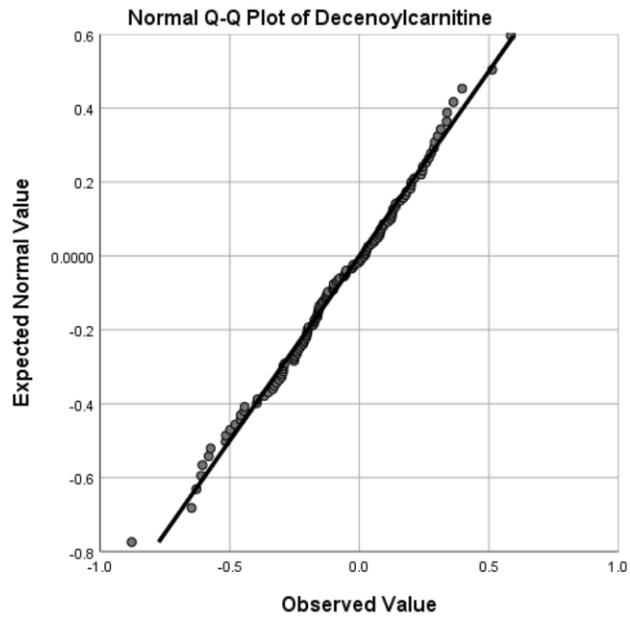
Supplementary Figure 19. The quantile-quantile (Q-Q) plot of normalized Decenoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of alphaNPhenylacetylLglutamine

Detrended Normal Q-Q Plot of alphaNPhenylacetylLglutamine
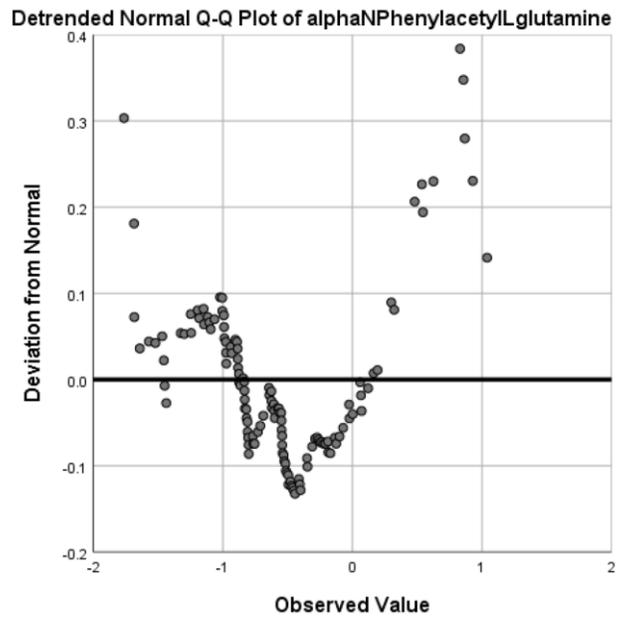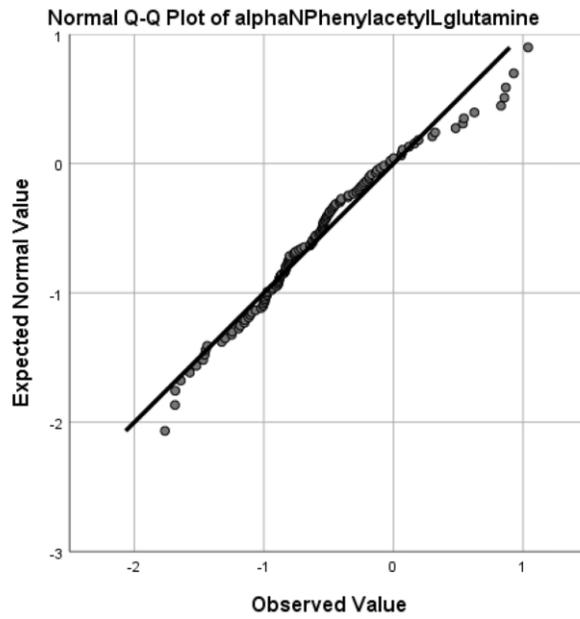
Supplementary Figure 20. The quantile-quantile (Q-Q) plot of normalized alphaNPhenylacetylLglutamine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Arabinoseisomer

Detrended Normal Q-Q Plot of Arabinoseisomer
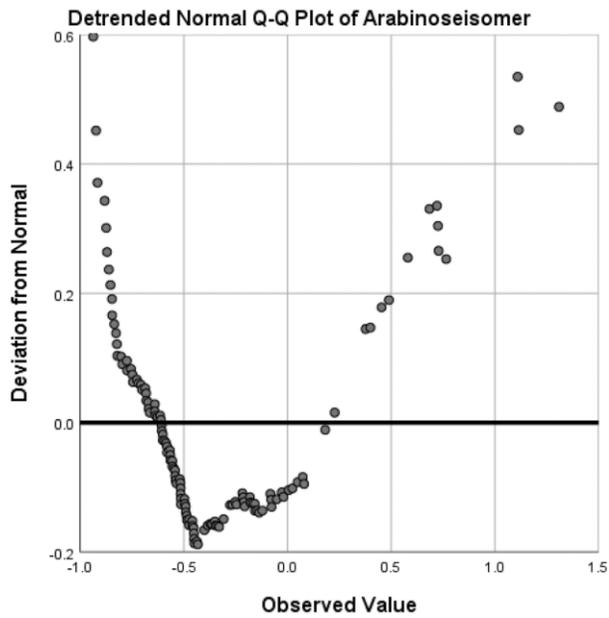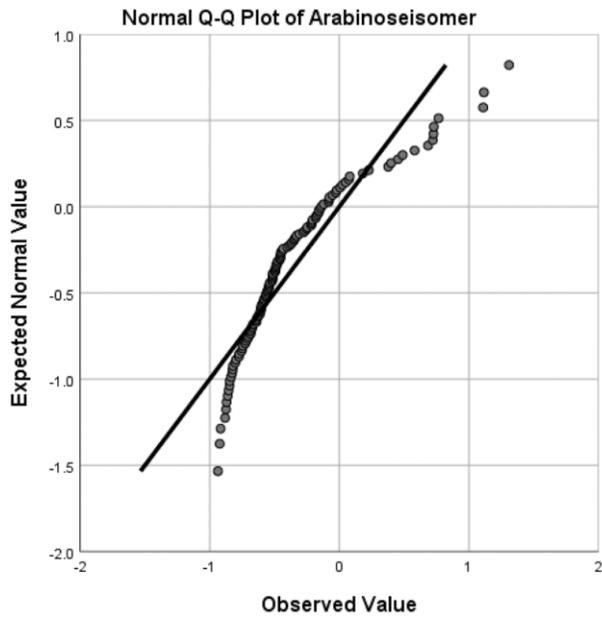
Supplementary Figure 21. The quantile-quantile (Q-Q) plot of normalized Arabinoseisomer which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
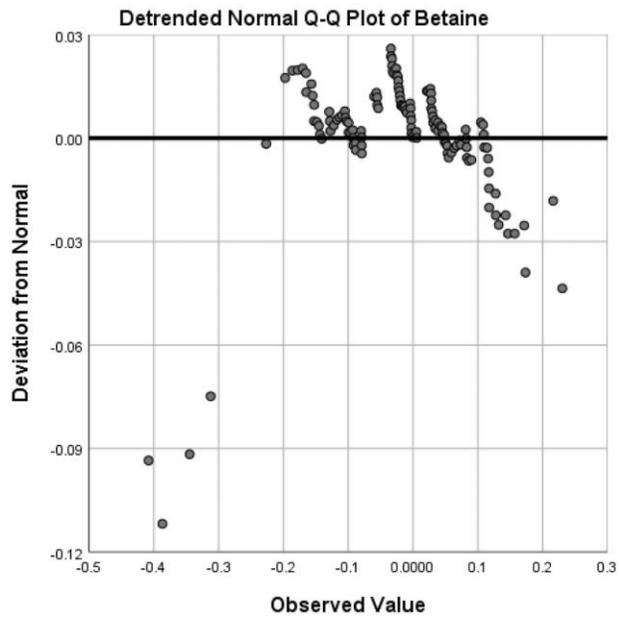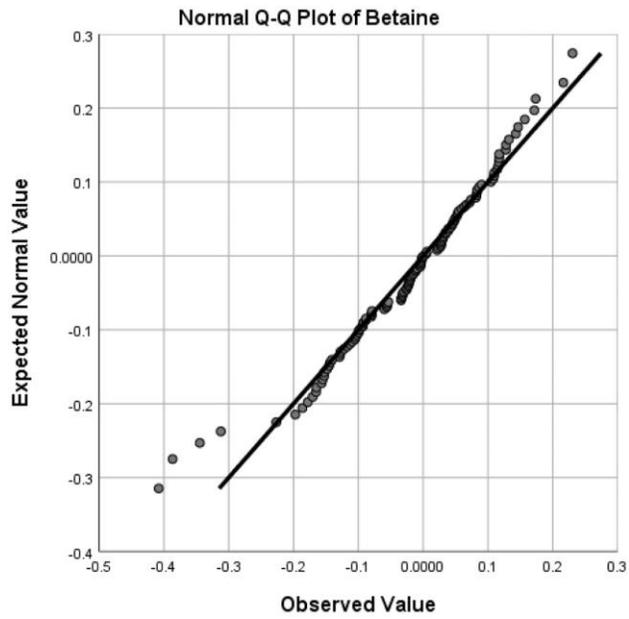
Supplementary Figure 22. The quantile-quantile (Q-Q) plot of normalized Betaine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
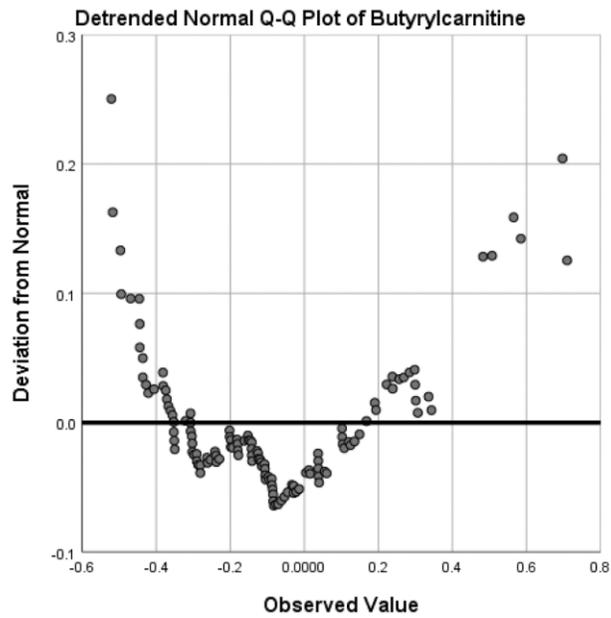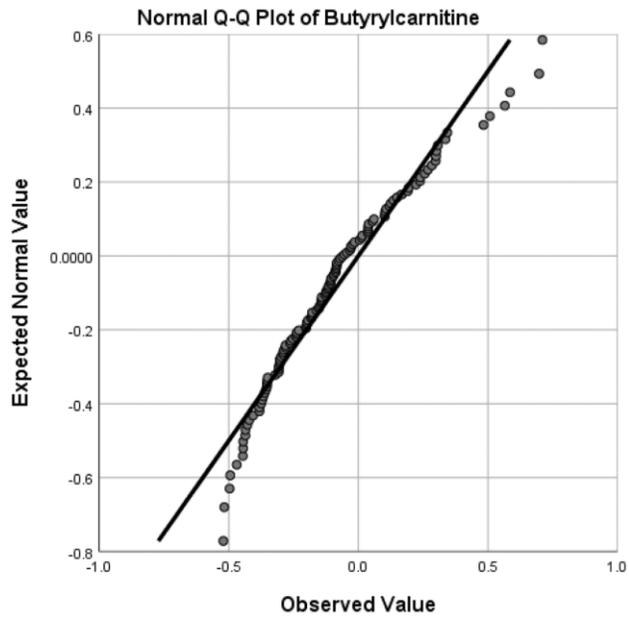
Supplementary Figure 23. The quantile-quantile (Q-Q) plot of normalized Butyrylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
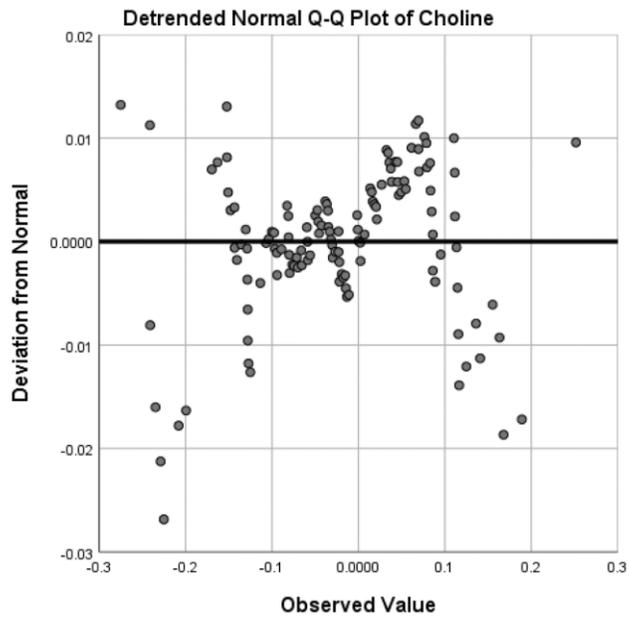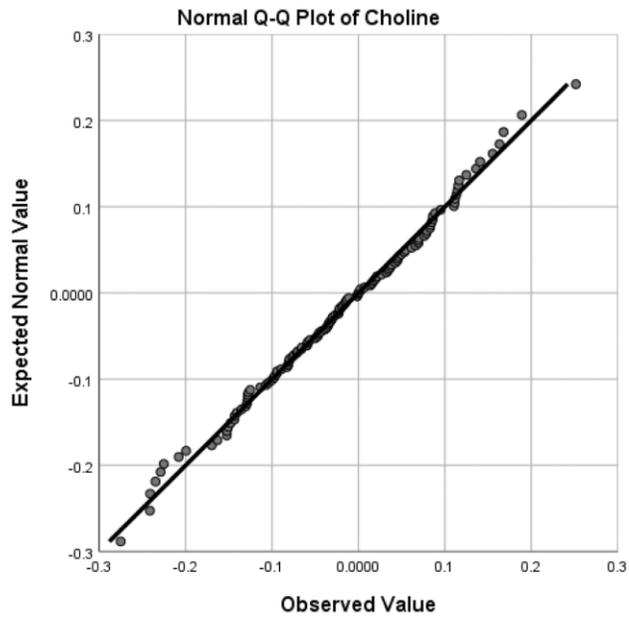
Supplementary Figure 24. The quantile-quantile (Q-Q) plot of normalized Choline which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of cis5Tetradecenoylcarnitine

Detrended Normal Q-Q Plot of cis5Tetradecenoylcarnitine
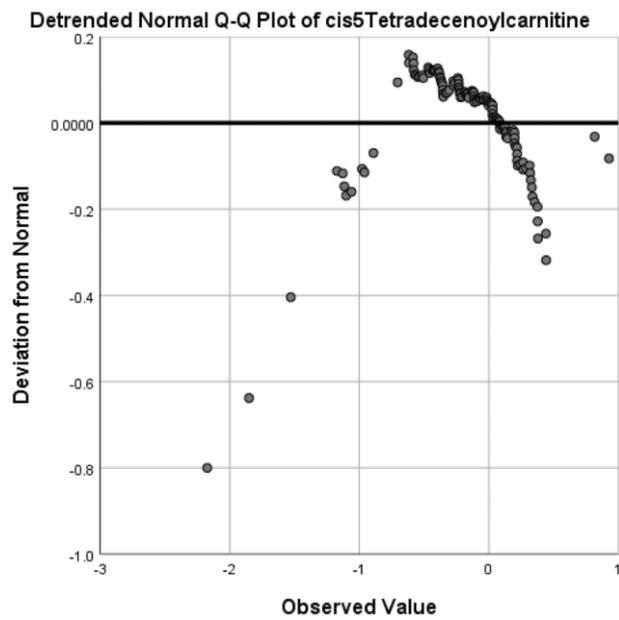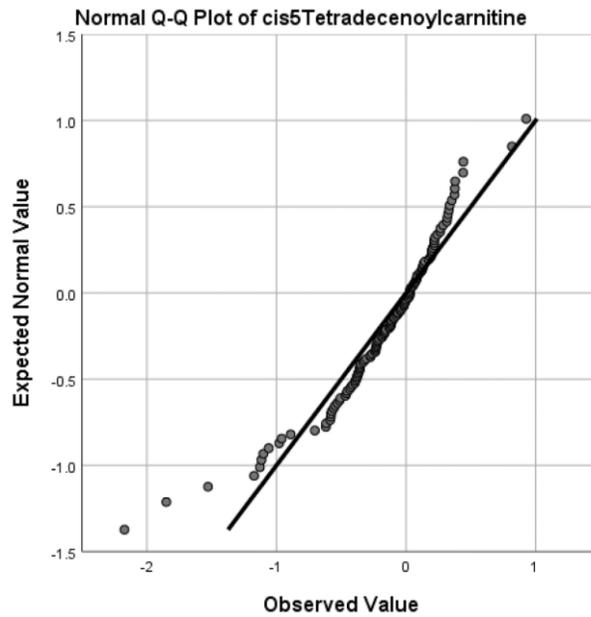
Supplementary Figure 25. The quantile-quantile (Q-Q) plot of normalized cis5Tetradecenoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
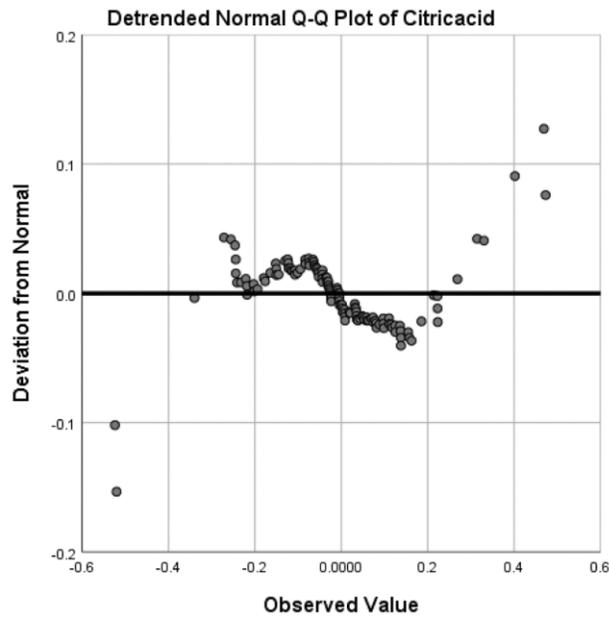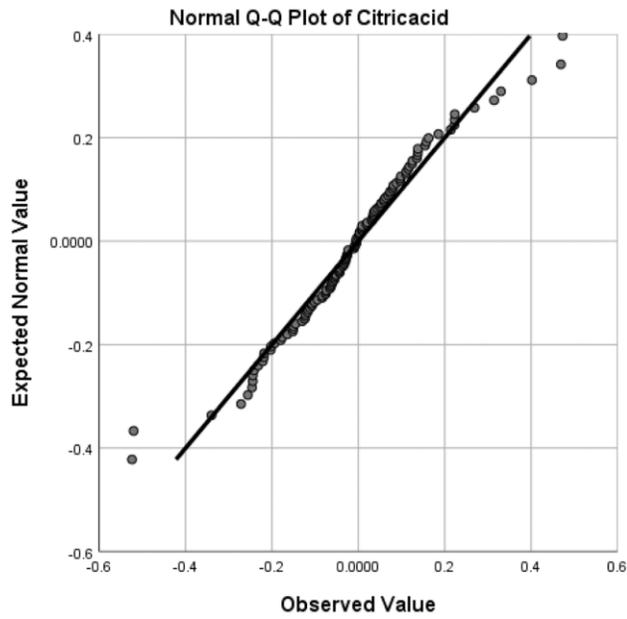
Supplementary Figure 26. The quantile-quantile (Q-Q) plot of normalized Citricacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Citrulline

Detrended Normal Q-Q Plot of Citrulline
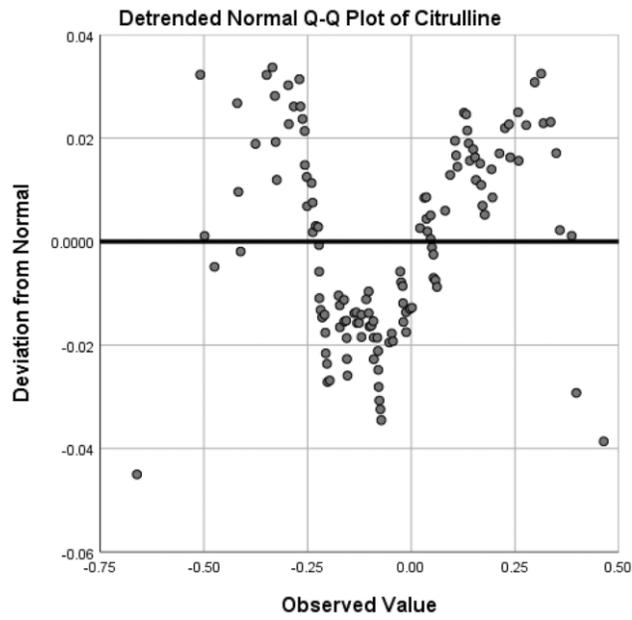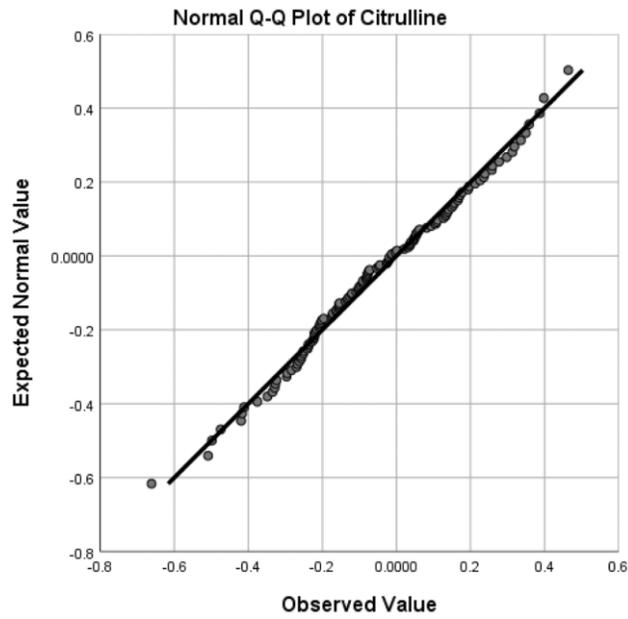
Supplementary Figure 27. The quantile-quantile (Q-Q) plot of normalized Citrulline which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
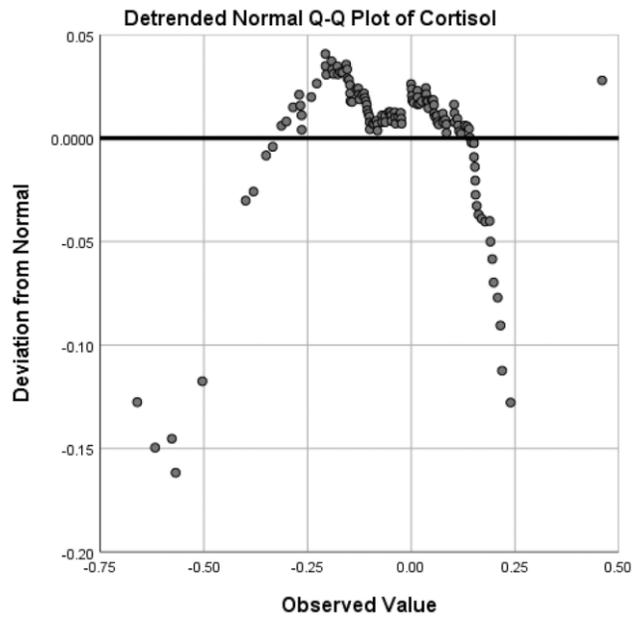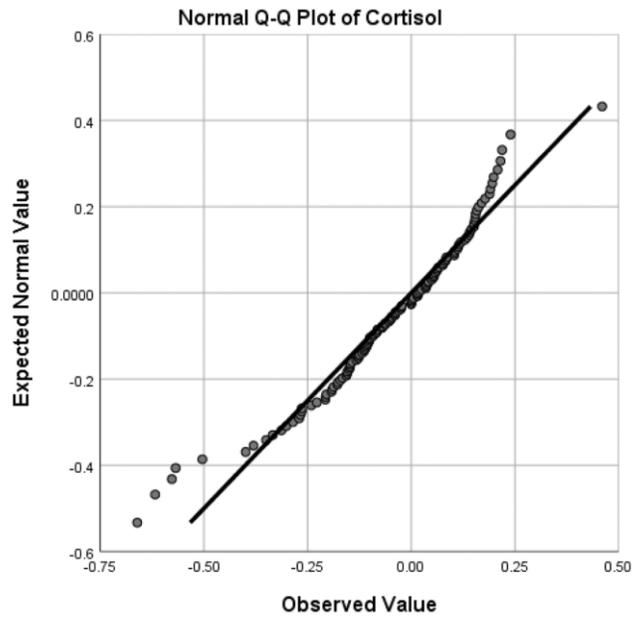
Supplementary Figure 28. The quantile-quantile (Q-Q) plot of normalized Cortisol which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
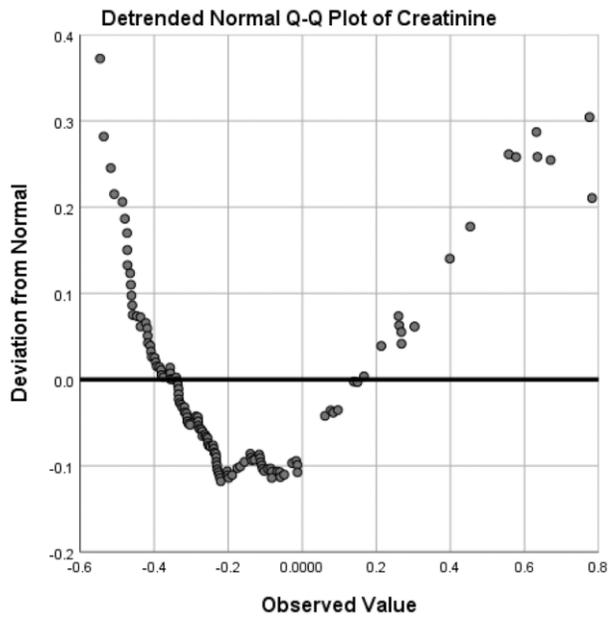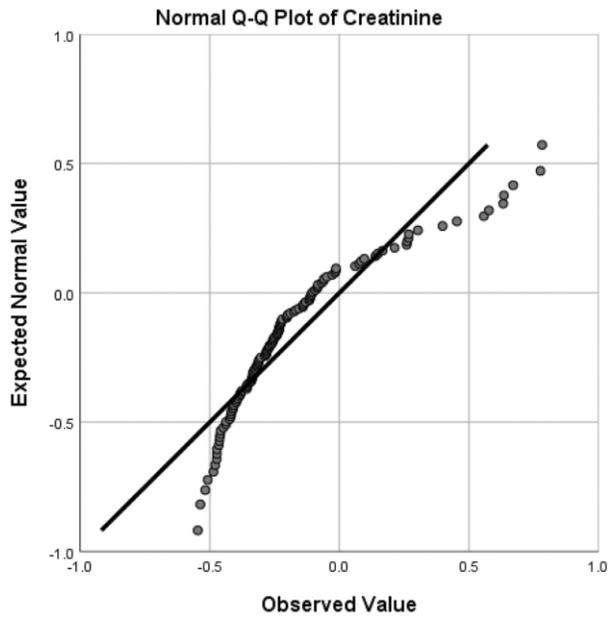
Supplementary Figure 29. The quantile-quantile (Q-Q) plot of normalized Creatinine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
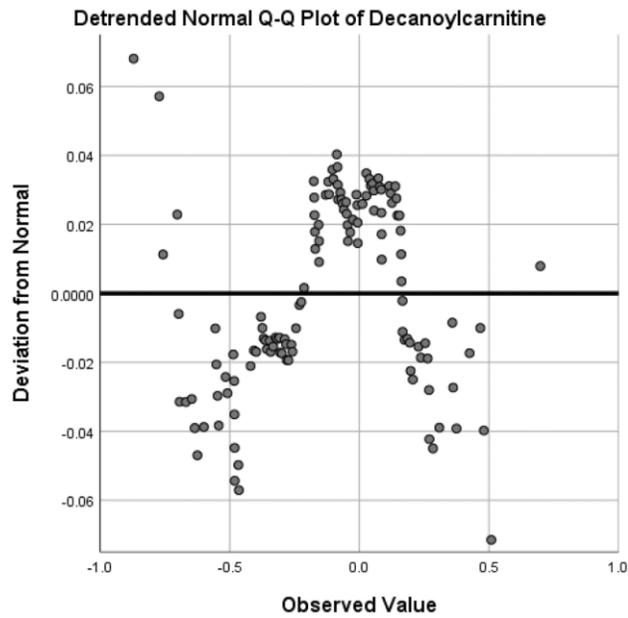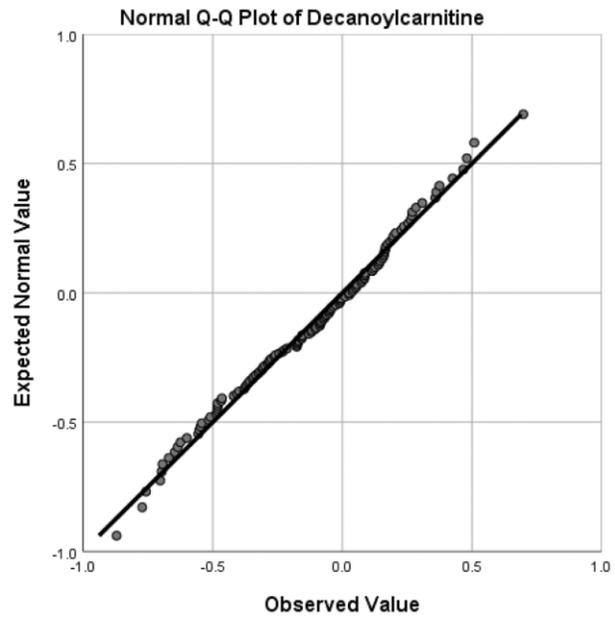
Supplementary Figure 30. The quantile-quantile (Q-Q) plot of normalized Decanoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
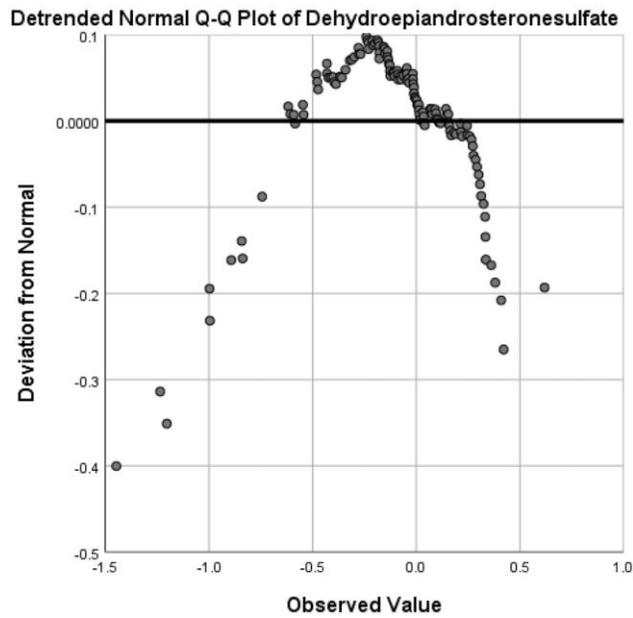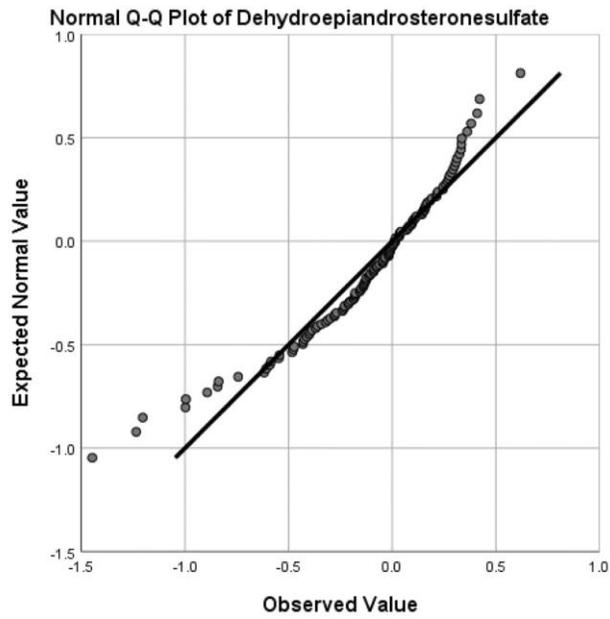
Supplementary Figure 31. The quantile-quantile (Q-Q) plot of normalized Dehydroepiandrosteronesulfate which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
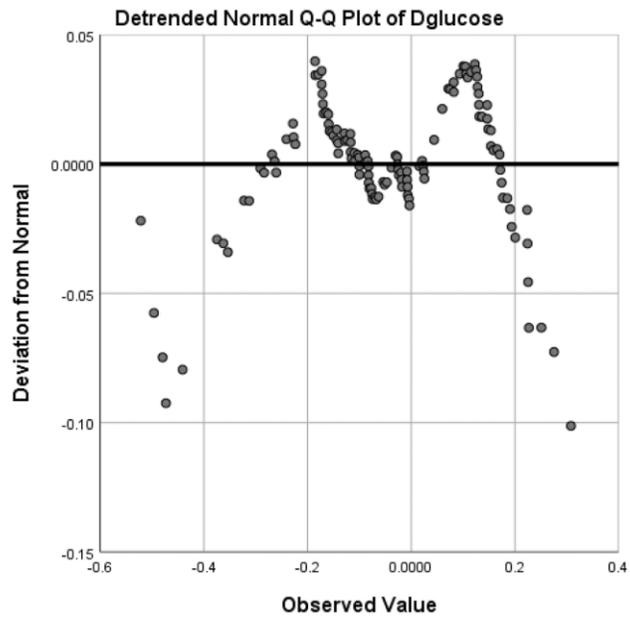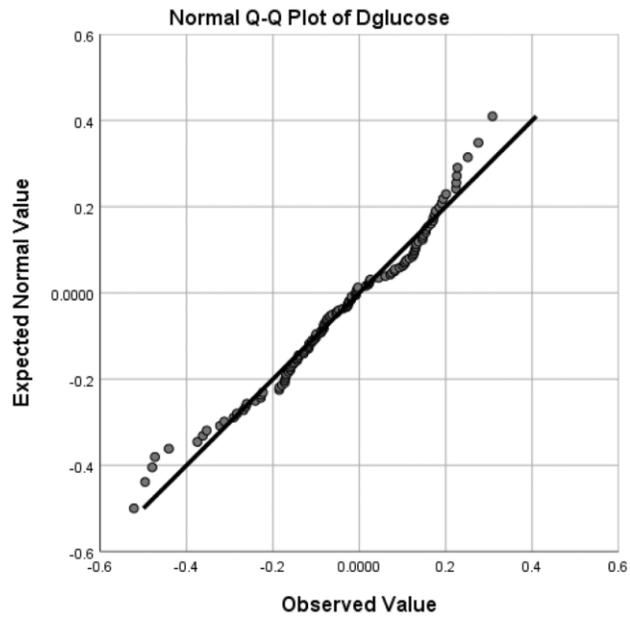
Supplementary Figure 32. The quantile-quantile (Q-Q) plot of normalized Dglucose which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Dodecanoylcarnitine

Detrended Normal Q-Q Plot of Dodecanoylcarnitine
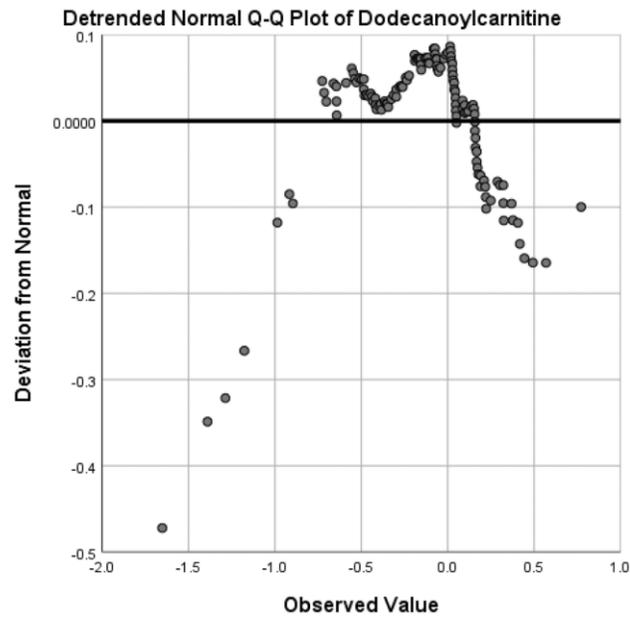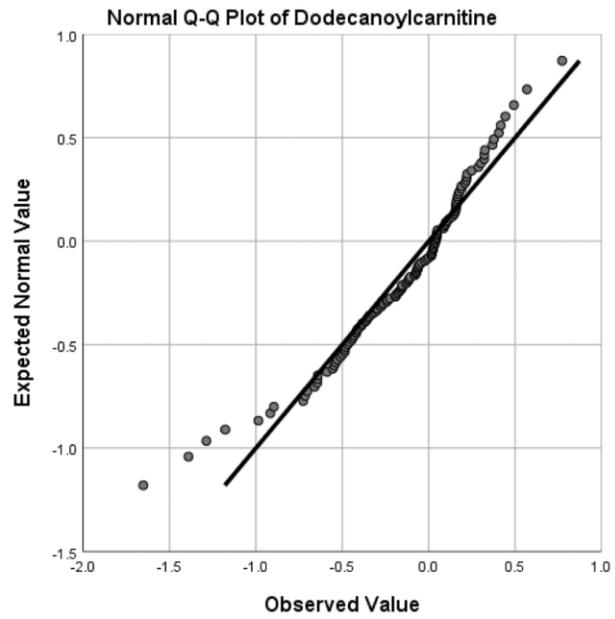
Supplementary Figure 33. The quantile-quantile (Q-Q) plot of normalized Dodecanoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of GammaButyrobetaine

Detrended Normal Q-Q Plot of GammaButyrobetaine
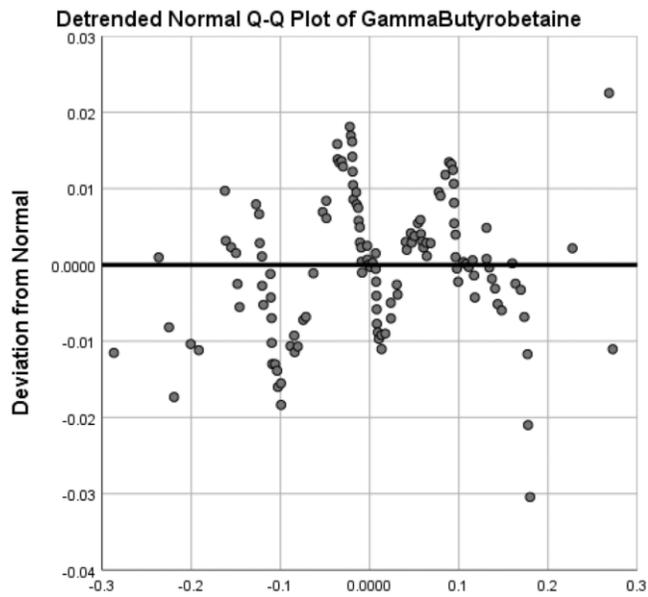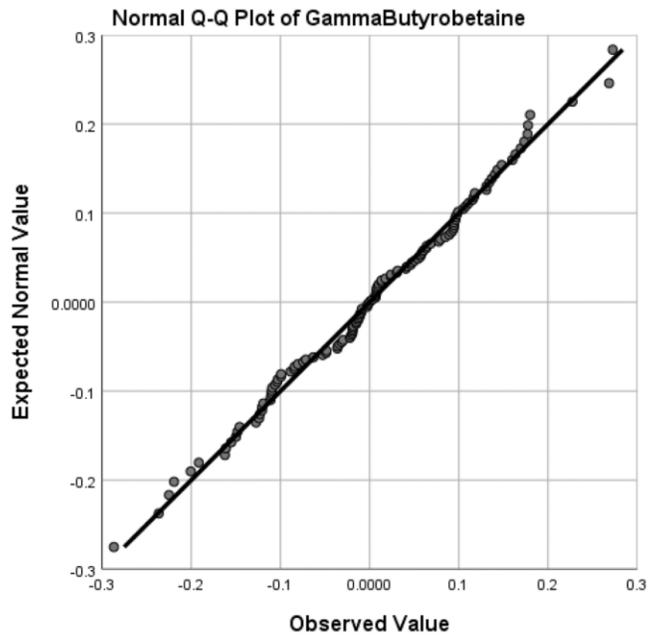
Supplementary Figure 34. The quantile-quantile (Q-Q) plot of normalized GammerButyrobetaine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
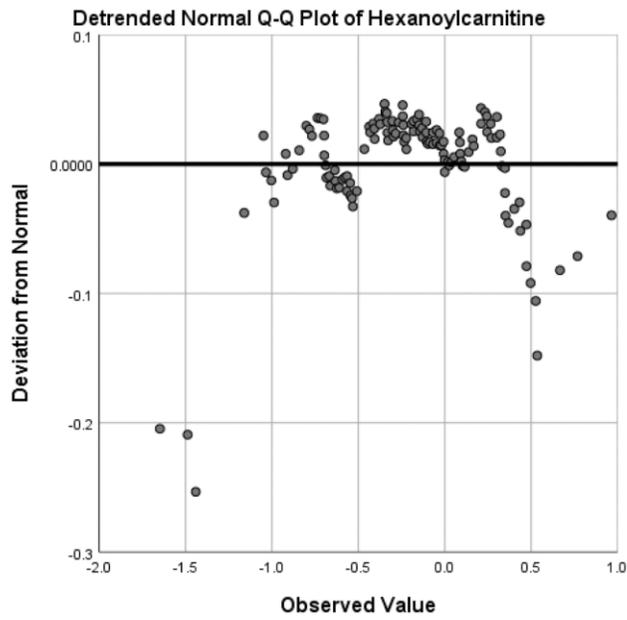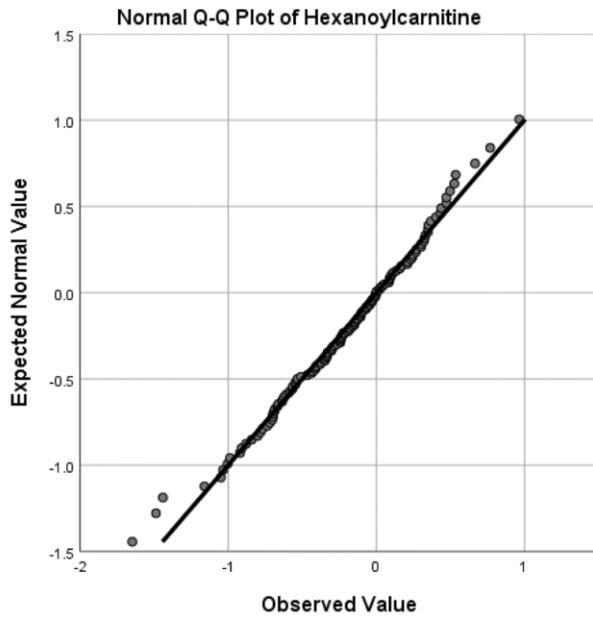
Supplementary Figure 35. The quantile-quantile (Q-Q) plot of normalized Hexanoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
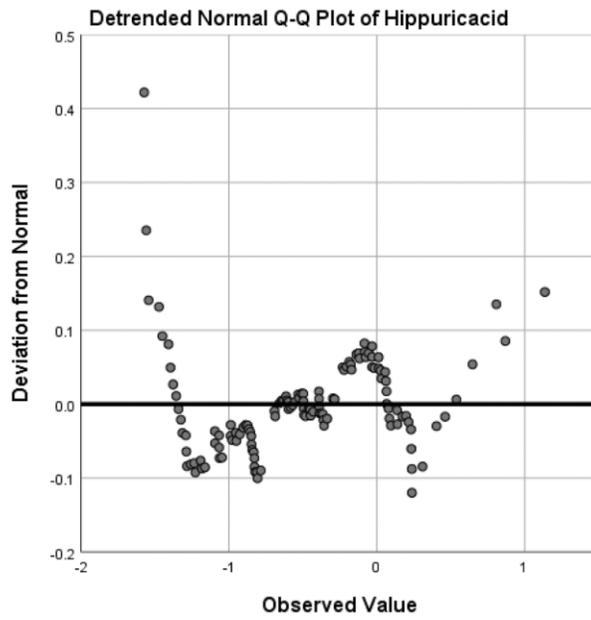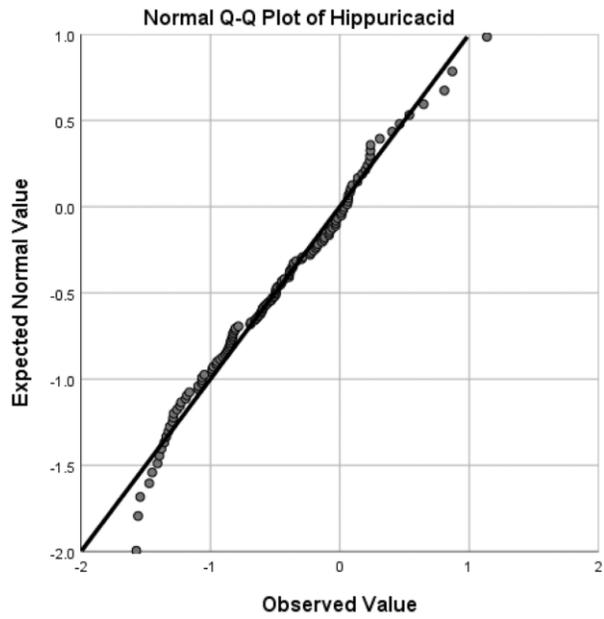
Supplementary Figure 36. The quantile-quantile (Q-Q) plot of normalized Hippuricacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Homovanillicacidsulfate

Detrended Normal Q-Q Plot of Homovanillicacidsulfate
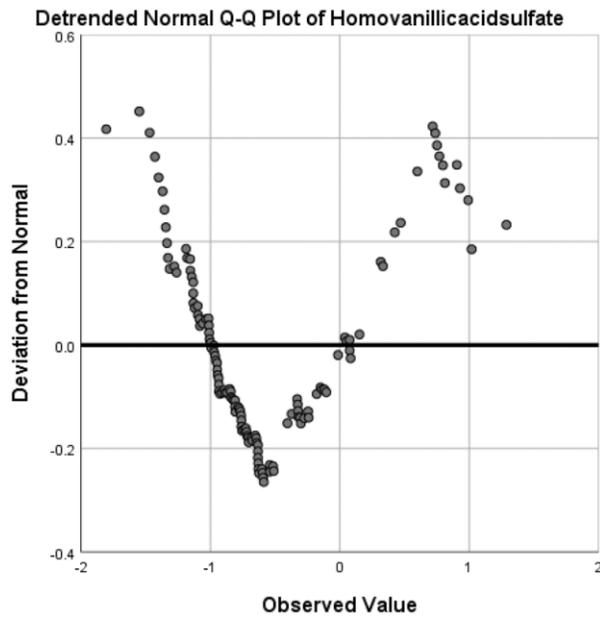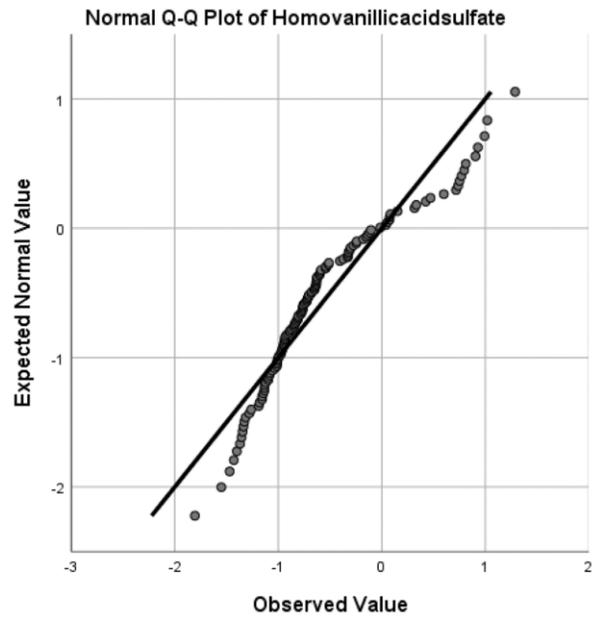
Supplementary Figure 37. The quantile-quantile (Q-Q) plot of normalized Homovanillicacidsulfate which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Hydroxybutyrylcarnitine

Detrended Normal Q-Q Plot of Hydroxybutyrylcarnitine
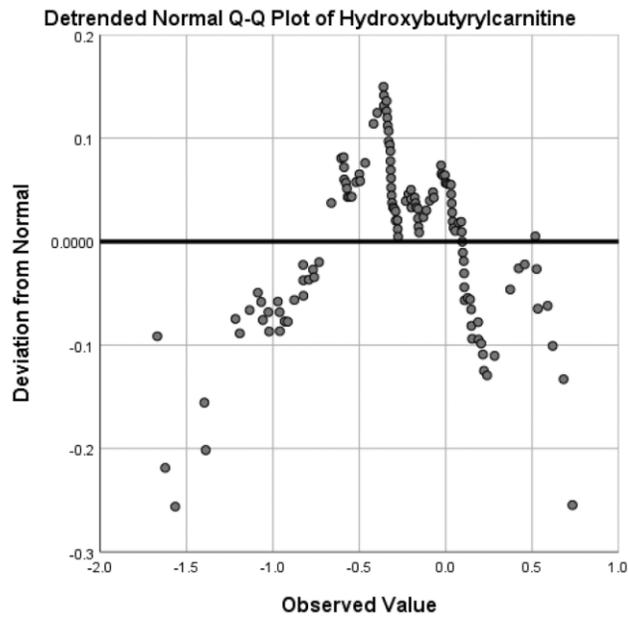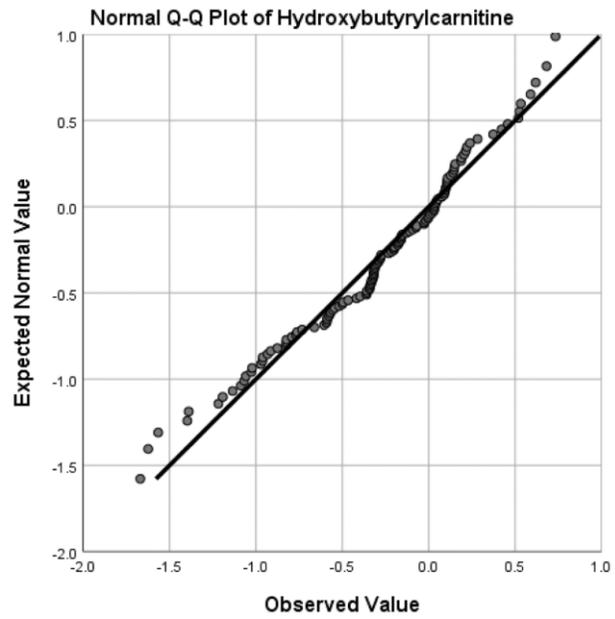
Supplementary Figure 38. The quantile-quantile (Q-Q) plot of normalized Hydroxybutyrylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
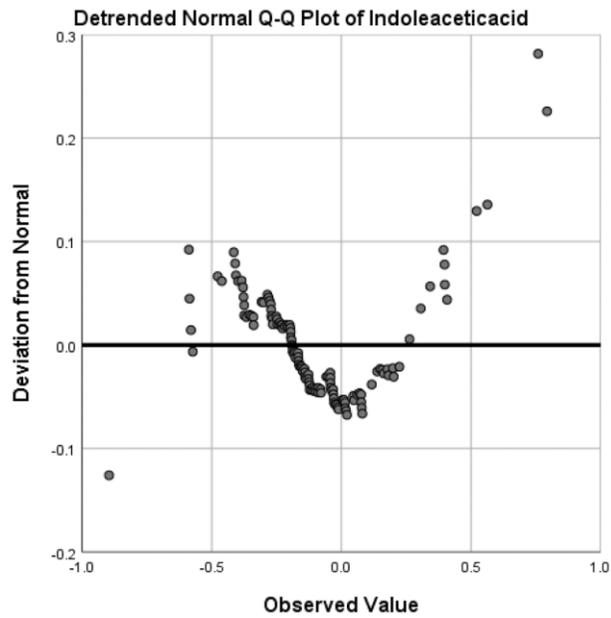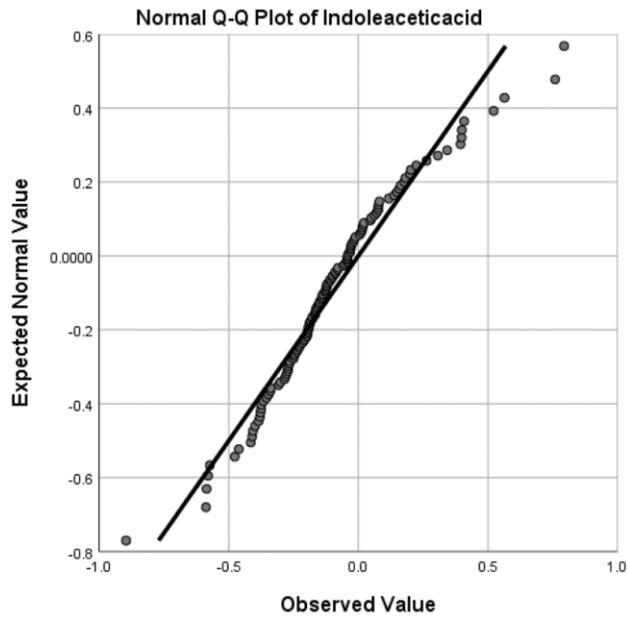
Supplementary Figure 39. The quantile-quantile (Q-Q) plot of normalized Indoleaceticacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
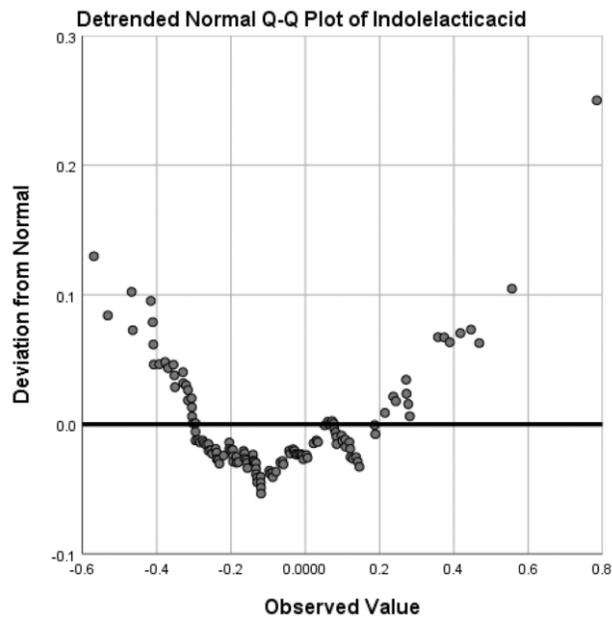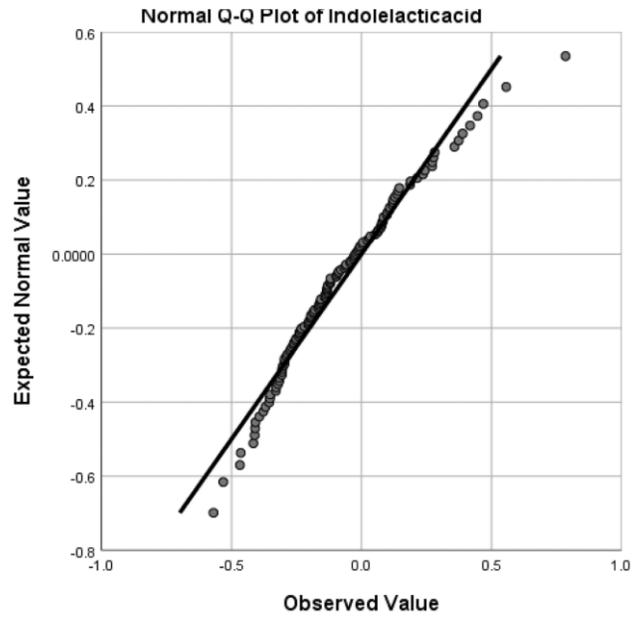
Supplementary Figure 40. The quantile-quantile (Q-Q) plot of normalized Indolelacticacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
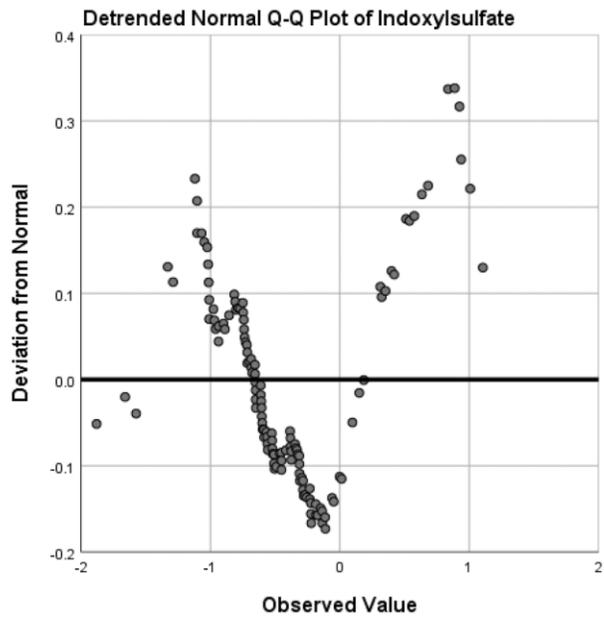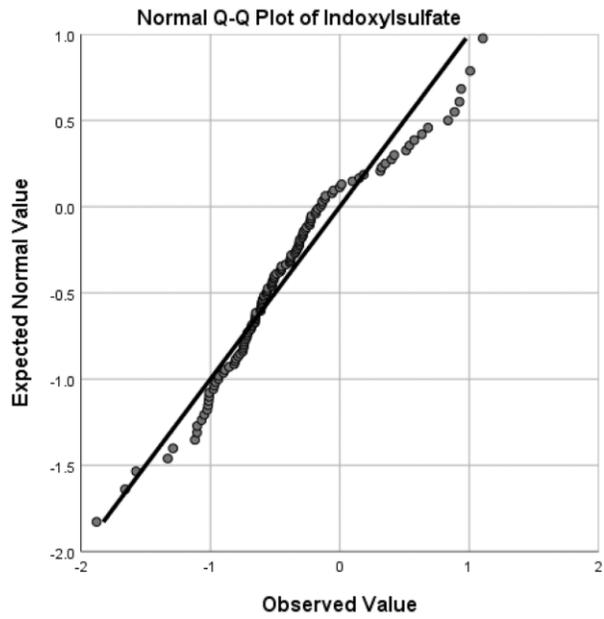
Supplementary Figure 41. The quantile-quantile (Q-Q) plot of normalized Indoxylsulfate which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Inosine

Detrended Normal Q-Q Plot of Inosine
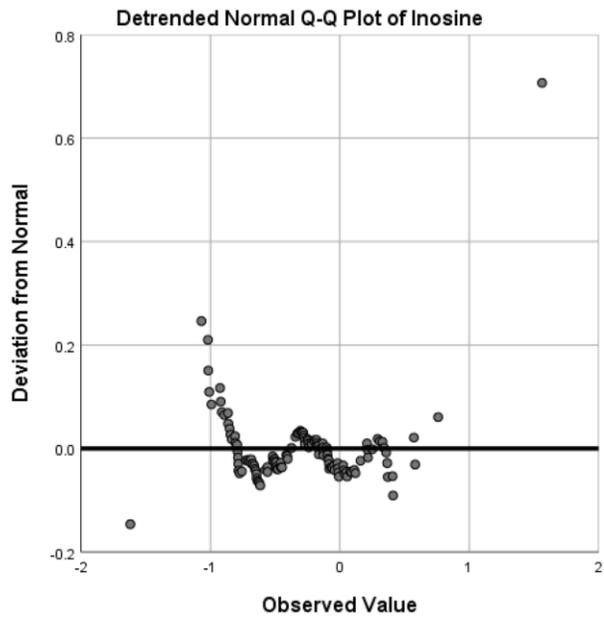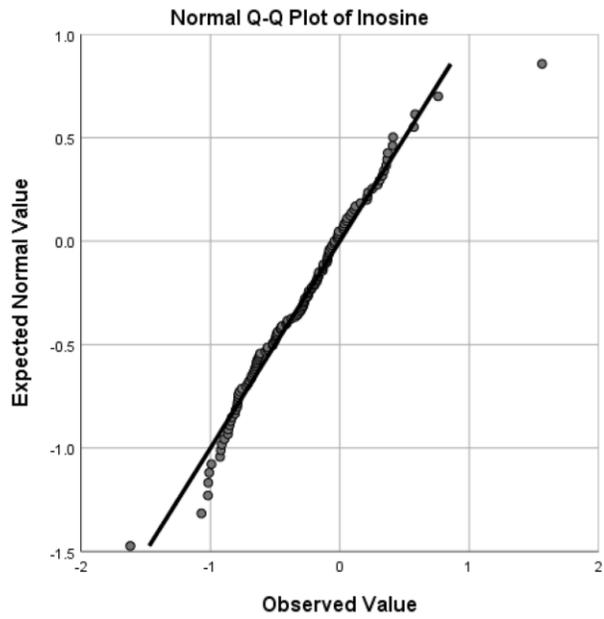
Supplementary Figure 42. The quantile-quantile (Q-Q) plot of normalized Inosine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Kynurenicacid

Detrended Normal Q-Q Plot of Kynurenicacid
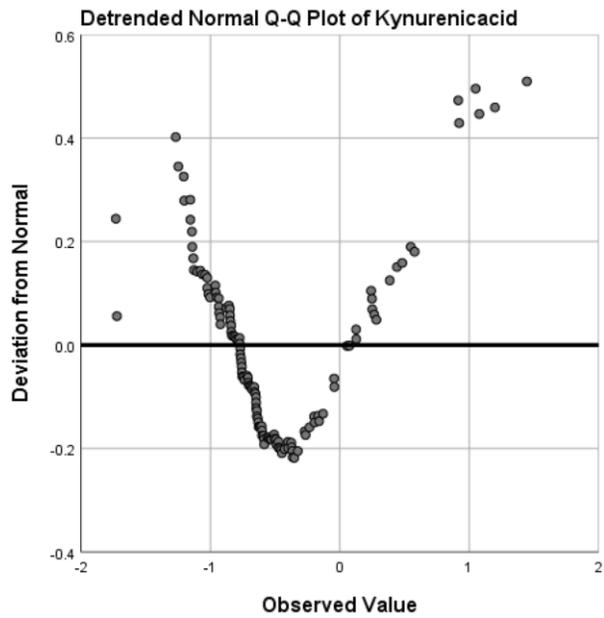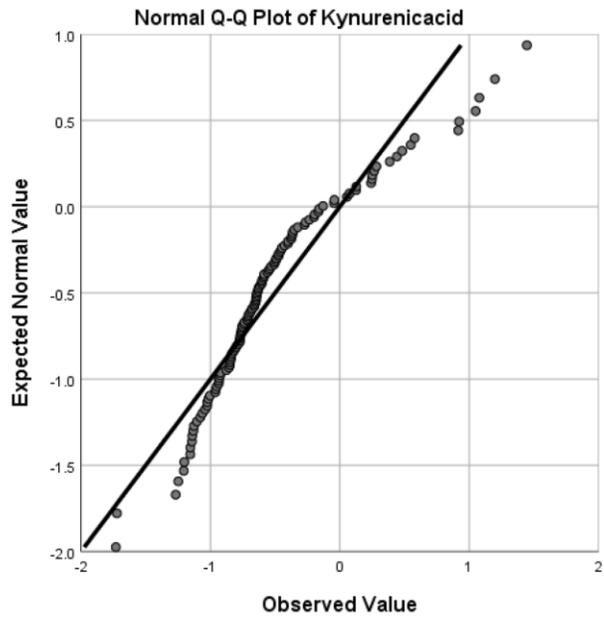
Supplementary Figure 43. The quantile-quantile (Q-Q) plot of normalized Kynurenicacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
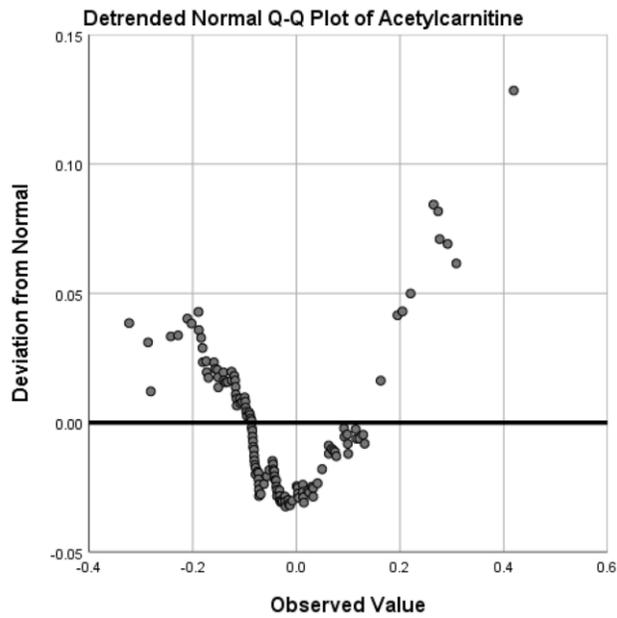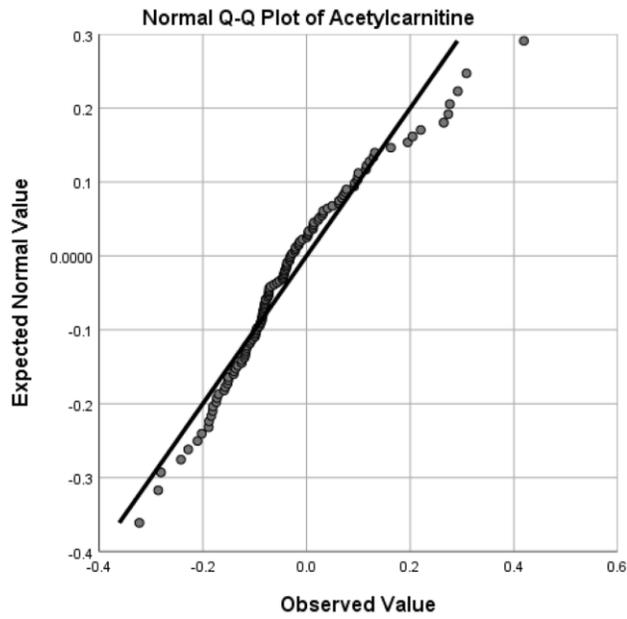
Supplementary Figure 44. The quantile-quantile (Q-Q) plot of normalized Acetylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
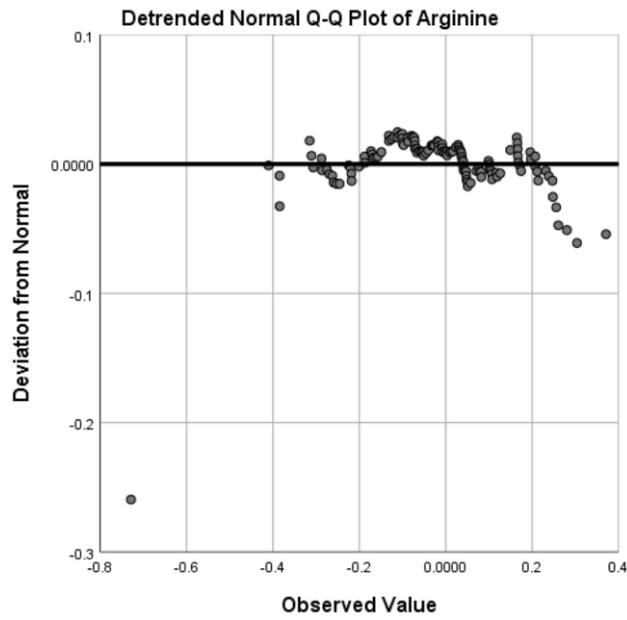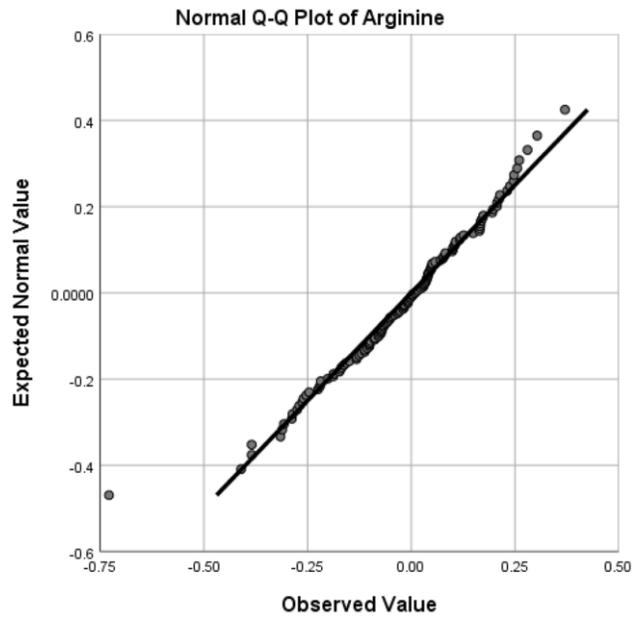
Supplementary Figure 45. The quantile-quantile (Q-Q) plot of normalized Arginine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
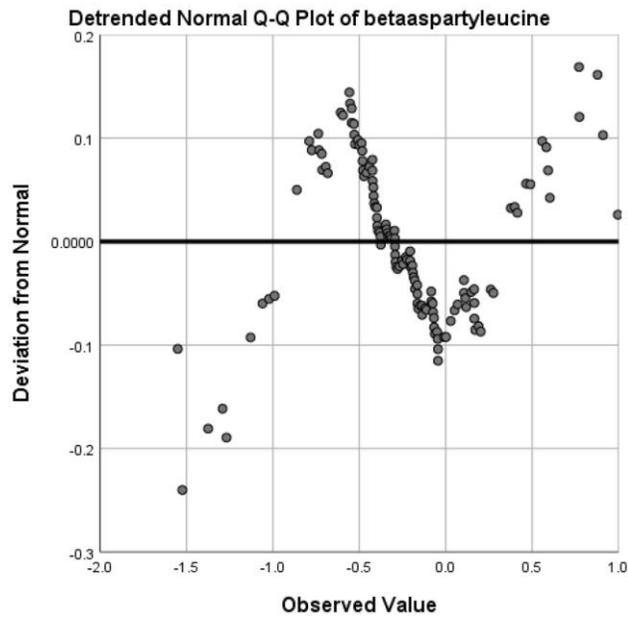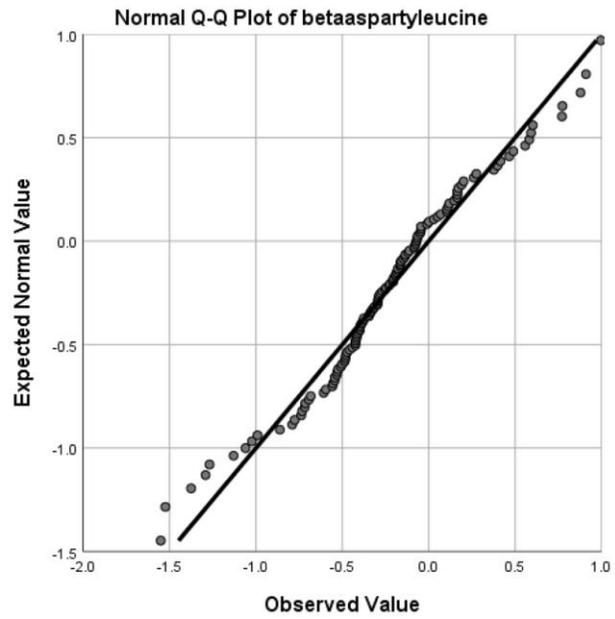
Supplementary Figure 46. The quantile-quantile (Q-Q) plot of normalized betaaspartyleucine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
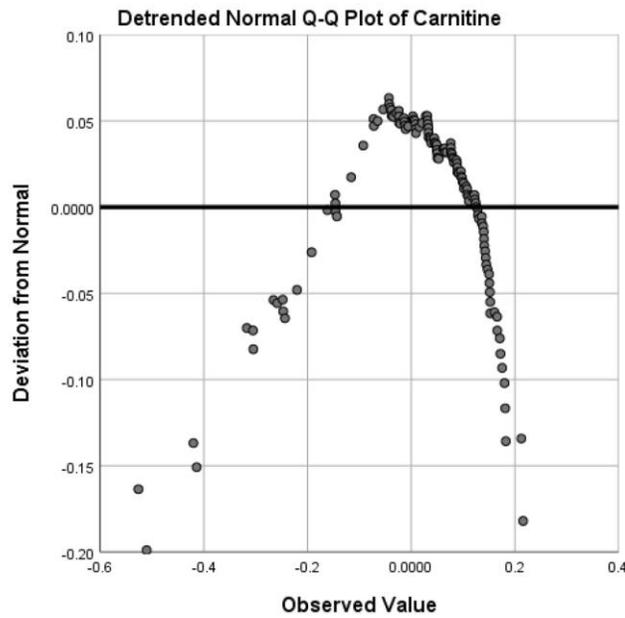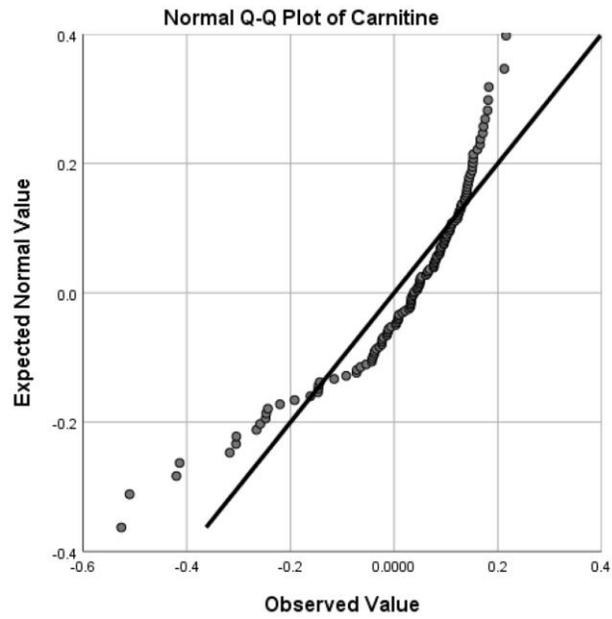
Supplementary Figure 47. The quantile-quantile (Q-Q) plot of normalized Carnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
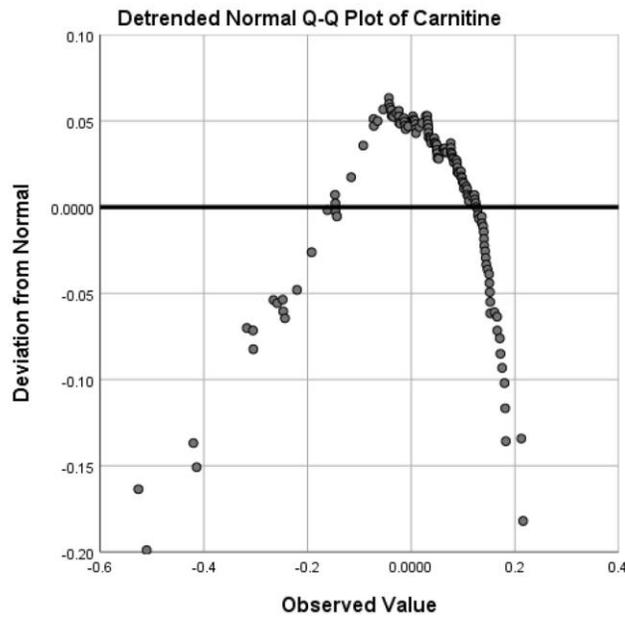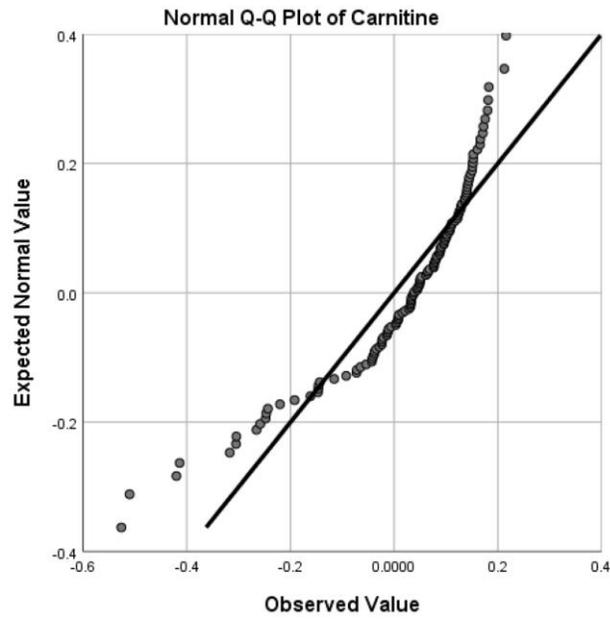
Supplementary Figure 48. The quantile-quantile (Q-Q) plot of normalized Glutamicacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
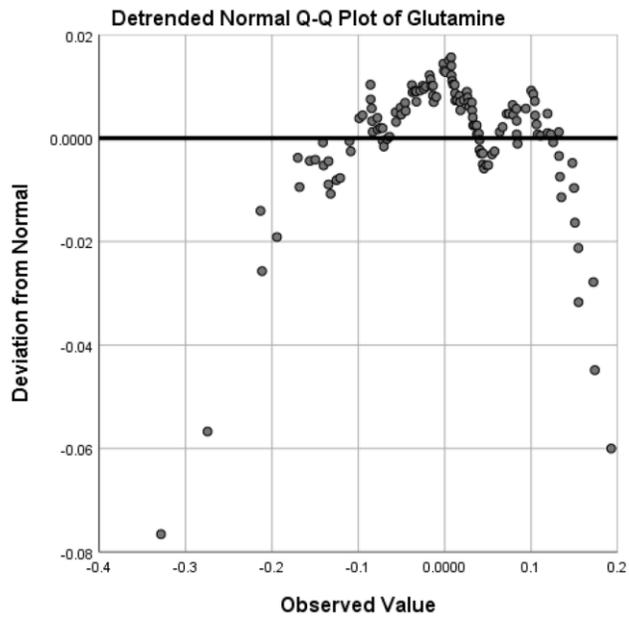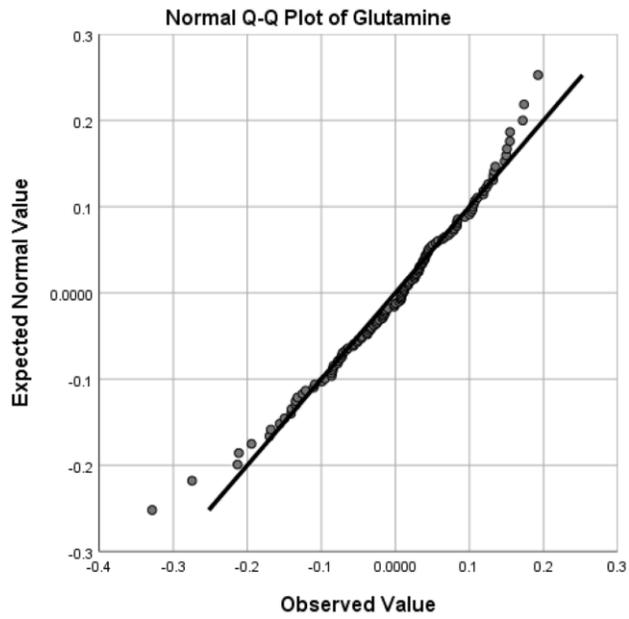
Supplementary Figure 49. The quantile-quantile (Q-Q) plot of normalized Glutamine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
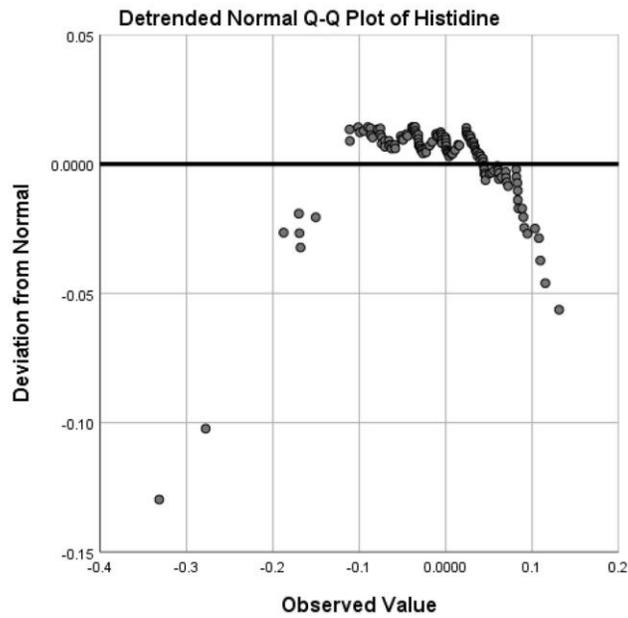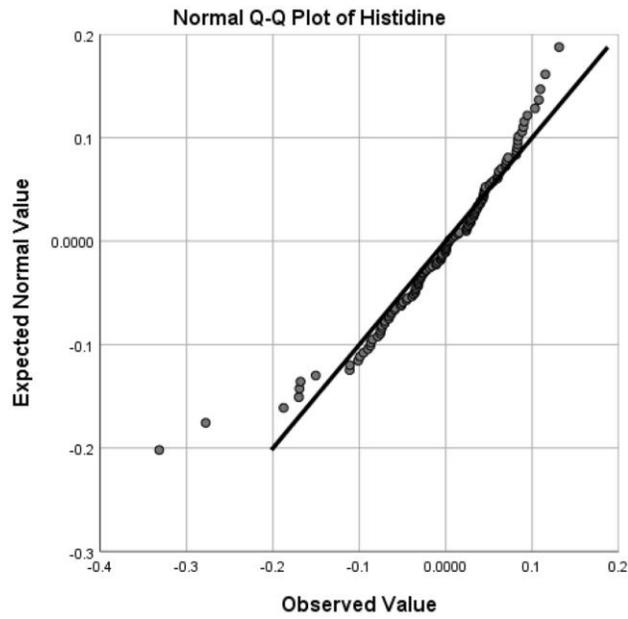
Supplementary Figure 50. The quantile-quantile (Q-Q) plot of normalized Histidine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Linoleylcarnitine

Detrended Normal Q-Q Plot of Linoleylcarnitine
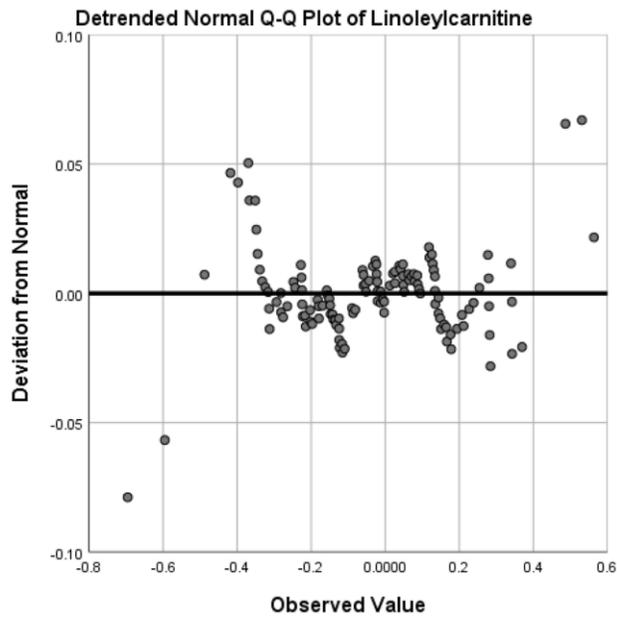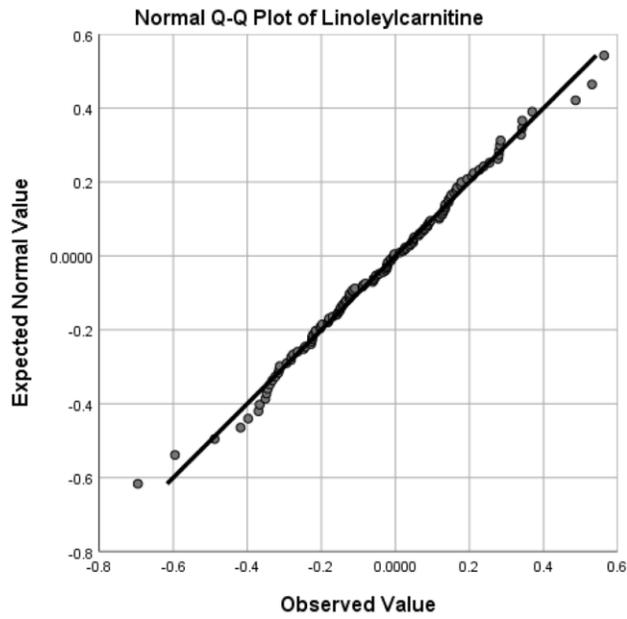
Supplementary Figure 51. The quantile-quantile (Q-Q) plot of normalized Linoleylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
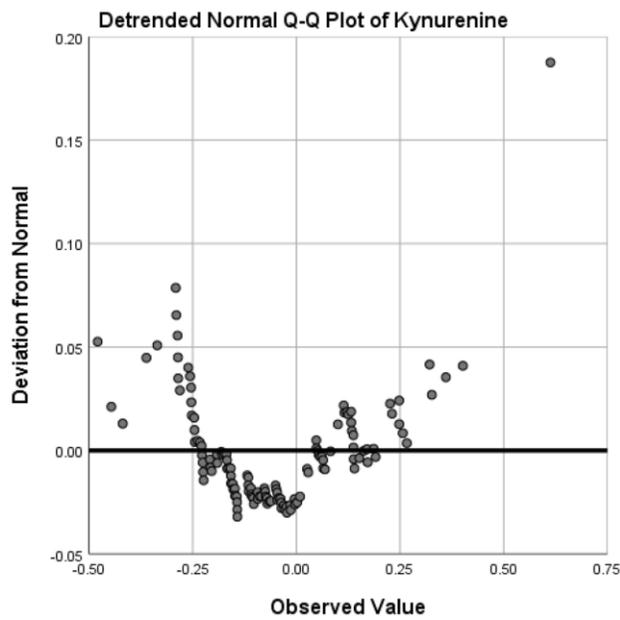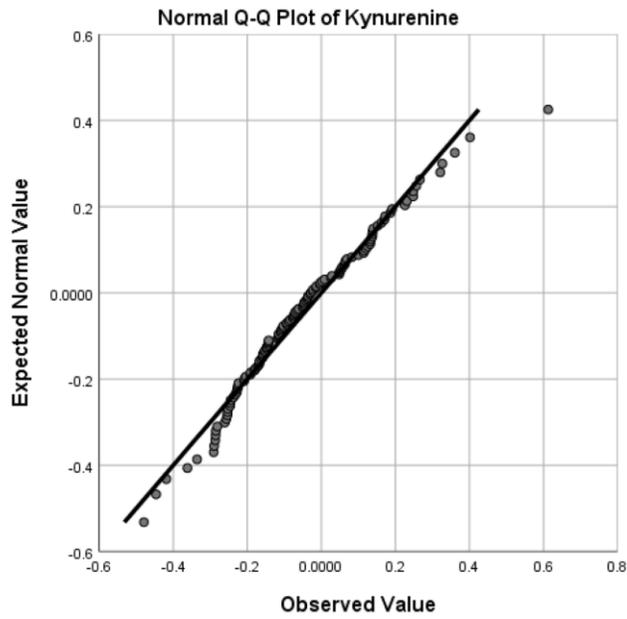
Supplementary Figure 52. The quantile-quantile (Q-Q) plot of normalized Kynurenine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
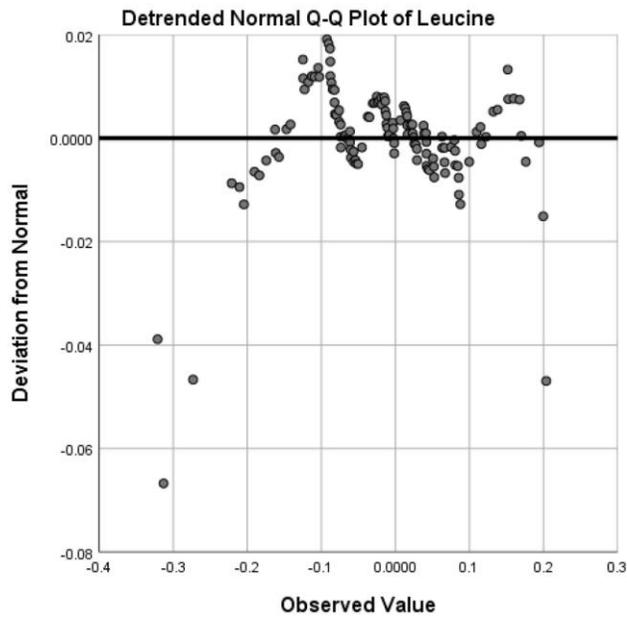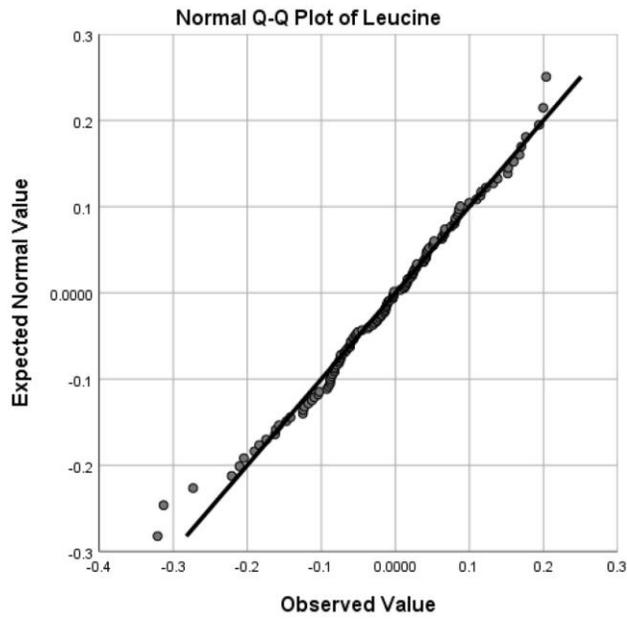
Supplementary Figure 53. The quantile-quantile (Q-Q) plot of normalized Leucine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
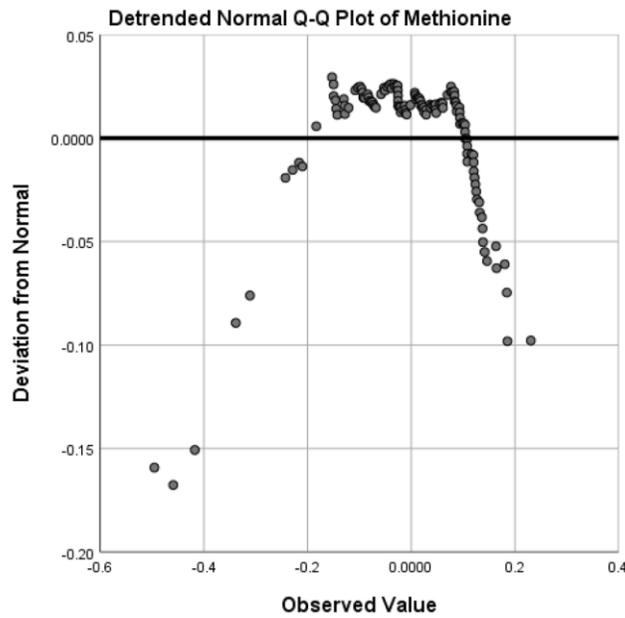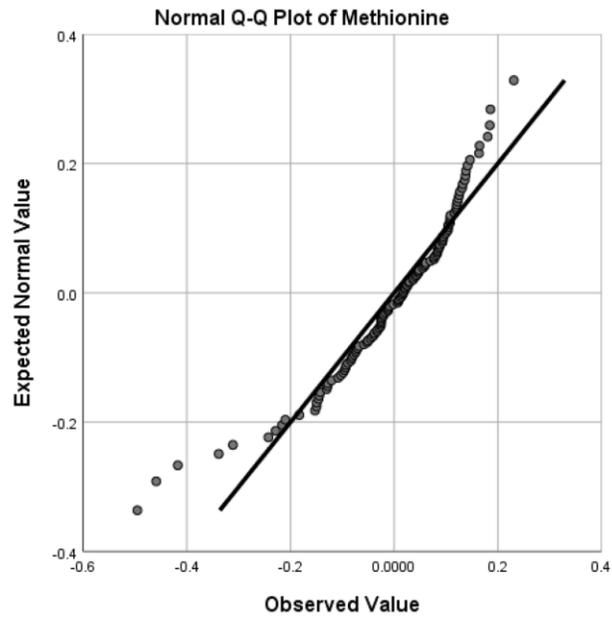
Supplementary Figure 54. The quantile-quantile (Q-Q) plot of normalized Methionine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
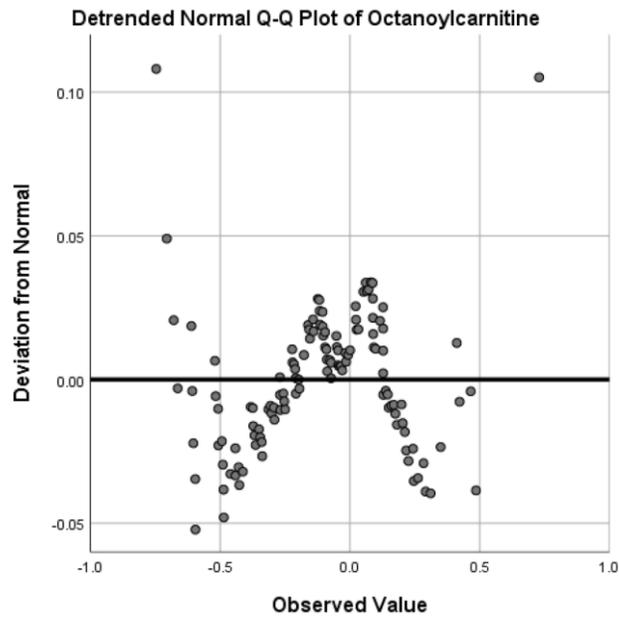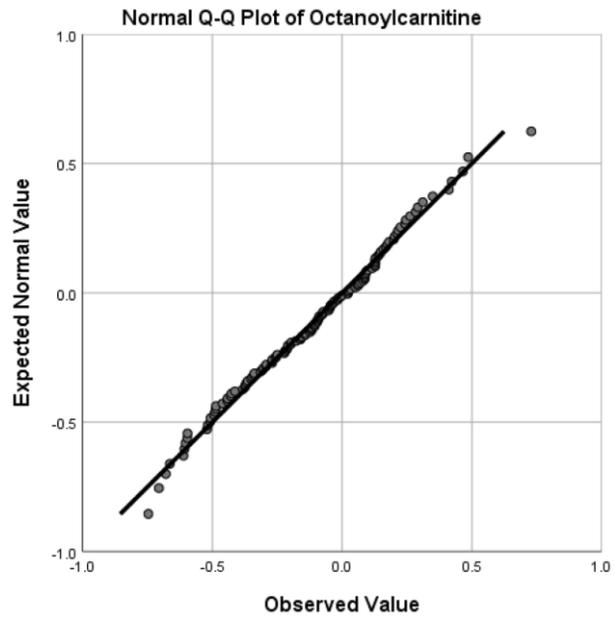
Supplementary Figure 55. The quantile-quantile (Q-Q) plot of normalized Octanoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
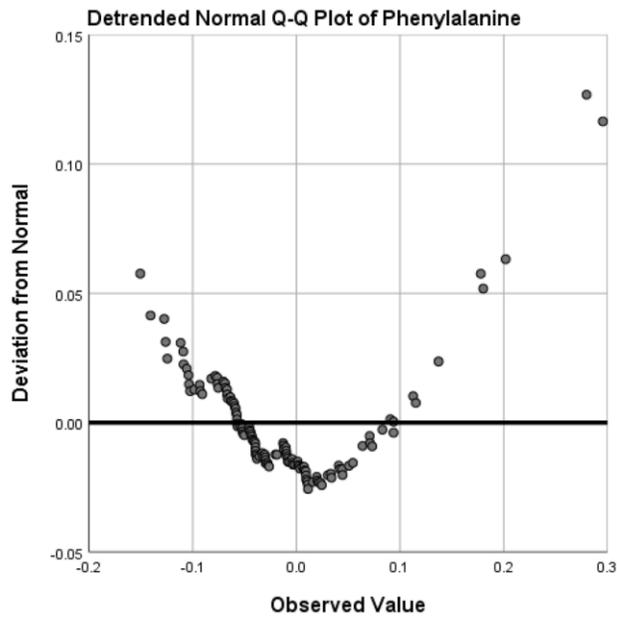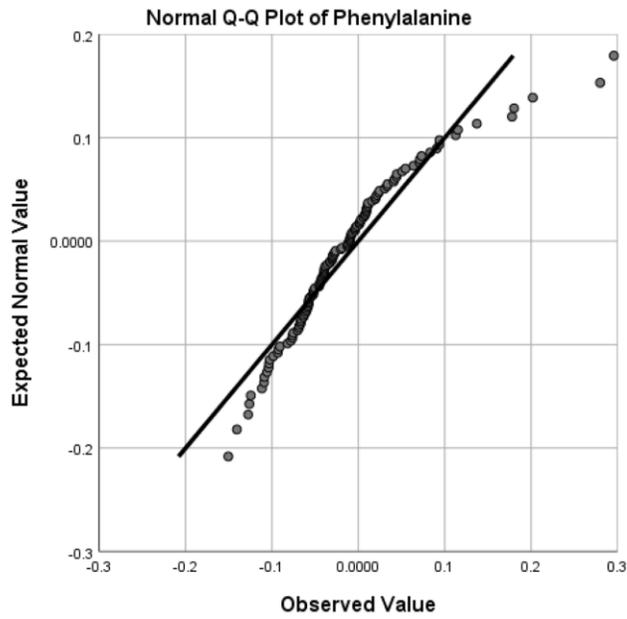
Supplementary Figure 56. The quantile-quantile (Q-Q) plot of normalized Phenylalanine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
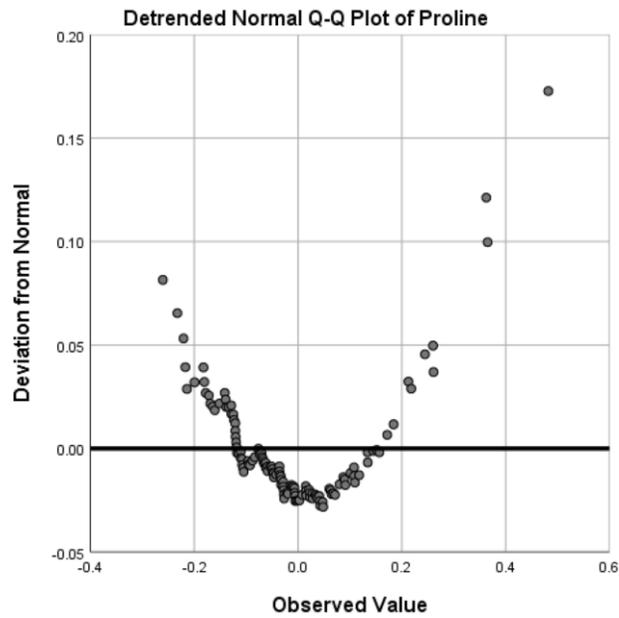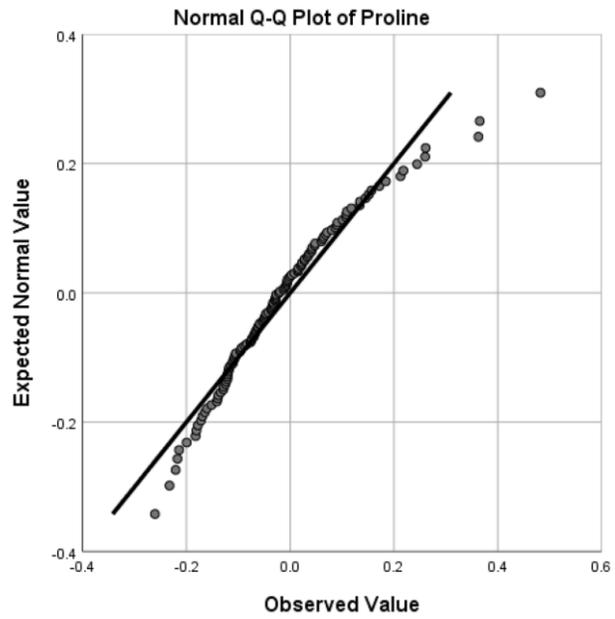
Supplementary Figure 57. The quantile-quantile (Q-Q) plot of normalized Proline which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
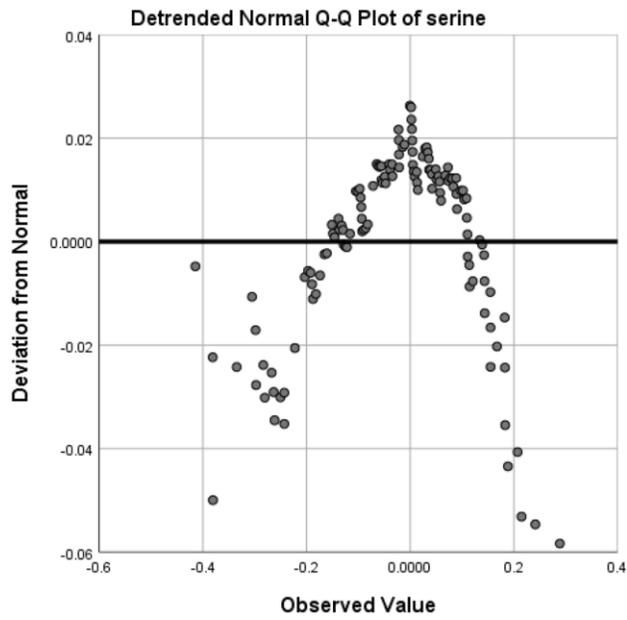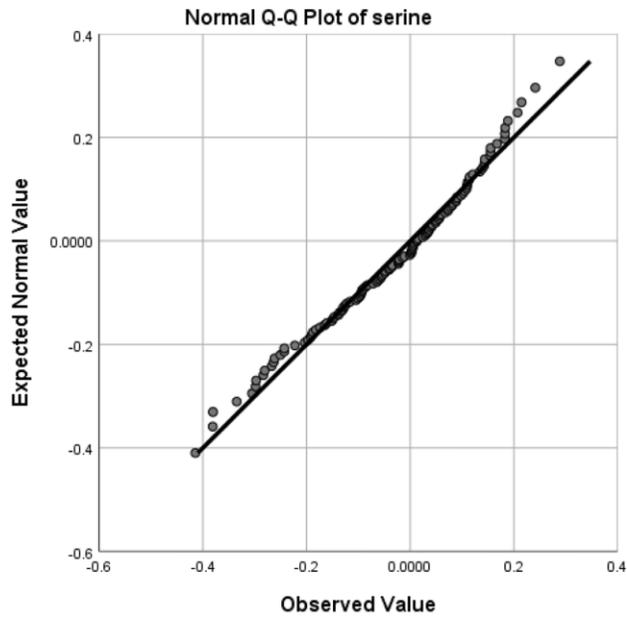
Supplementary Figure 58. The quantile-quantile (Q-Q) plot of normalized serine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
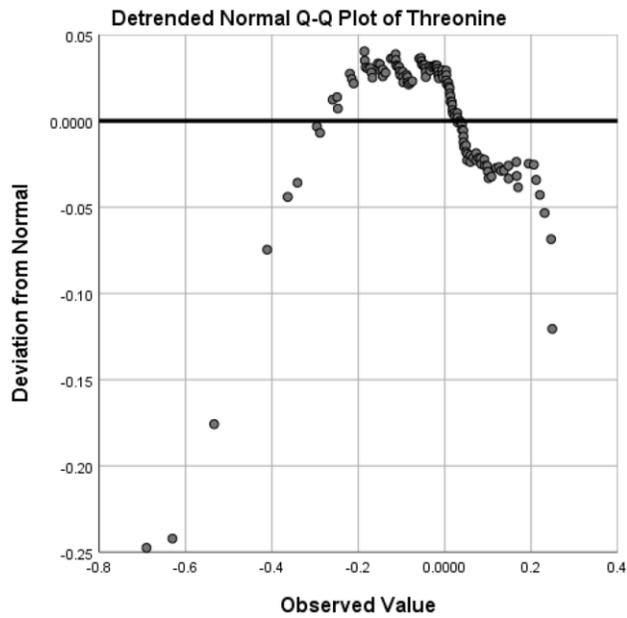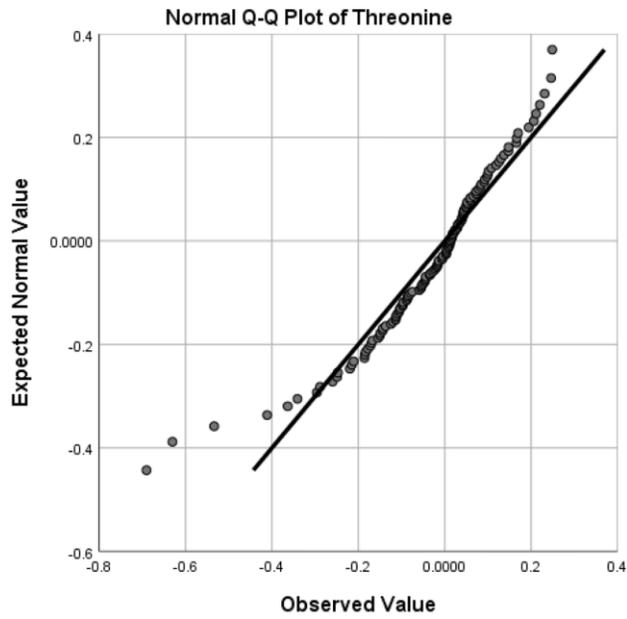
Supplementary Figure 59. The quantile-quantile (Q-Q) plot of normalized Threonine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
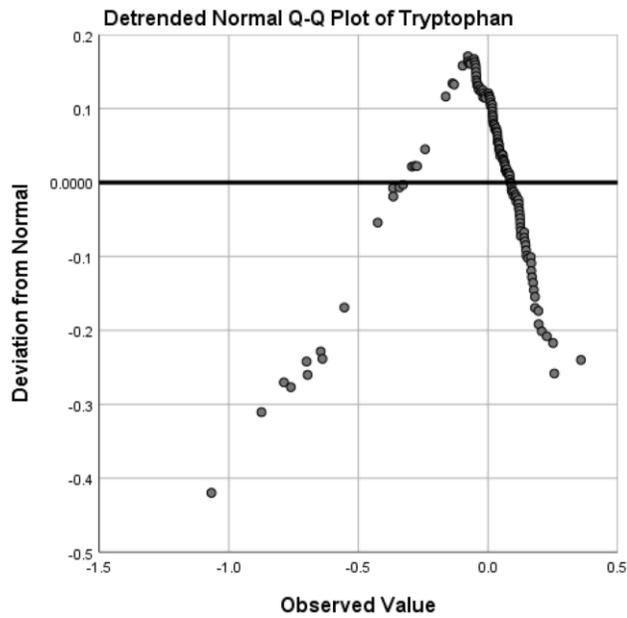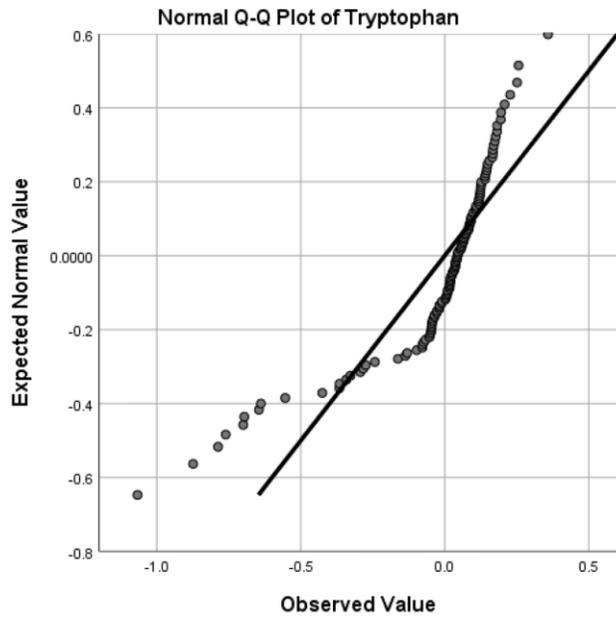
Supplementary Figure 60. The quantile-quantile (Q-Q) plot of normalized Tryptophan which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
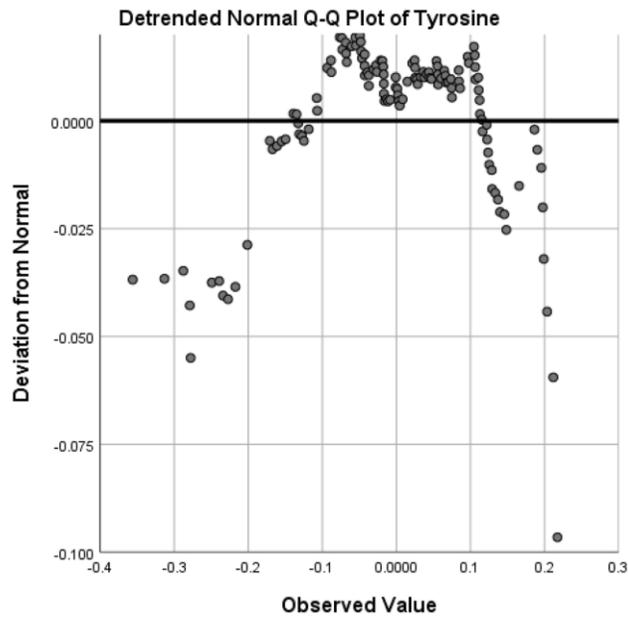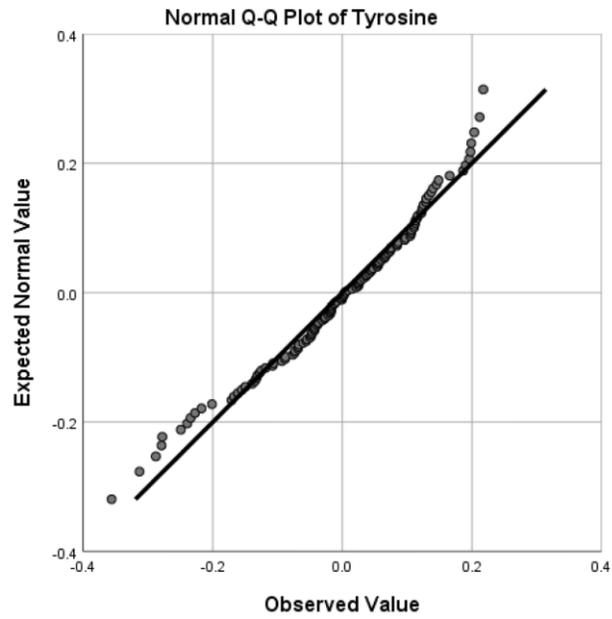
Supplementary Figure 61. The quantile-quantile (Q-Q) plot of normalized Tyrosine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
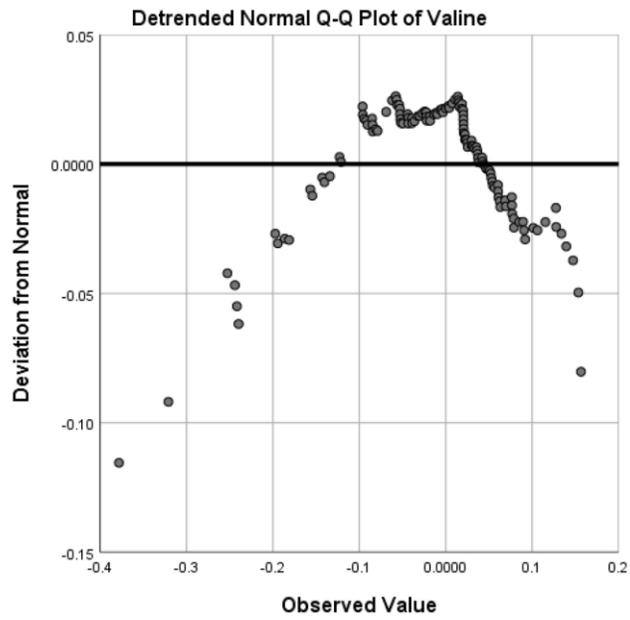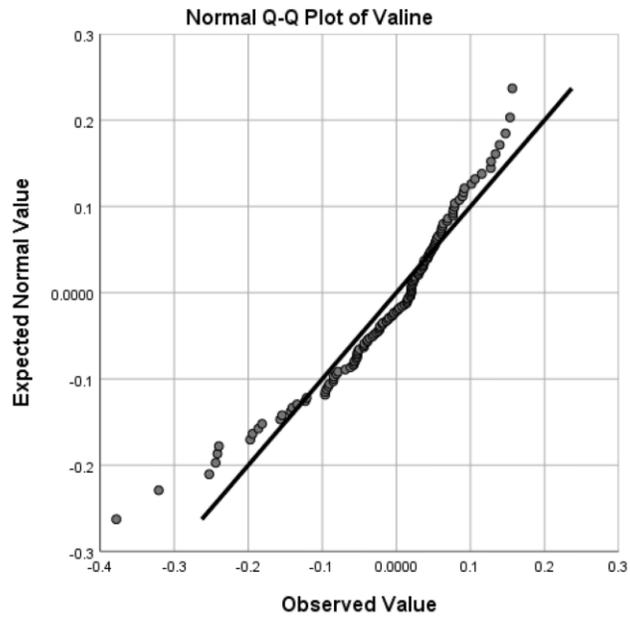
Supplementary Figure 62. The quantile-quantile (Q-Q) plot of normalized Valine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
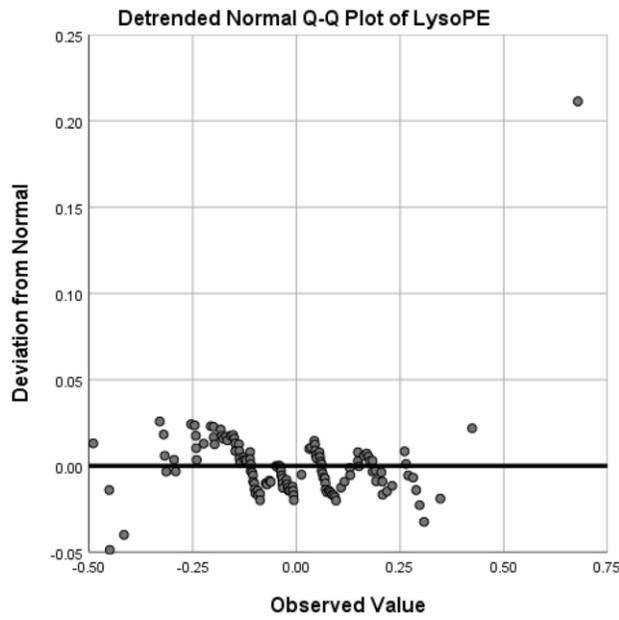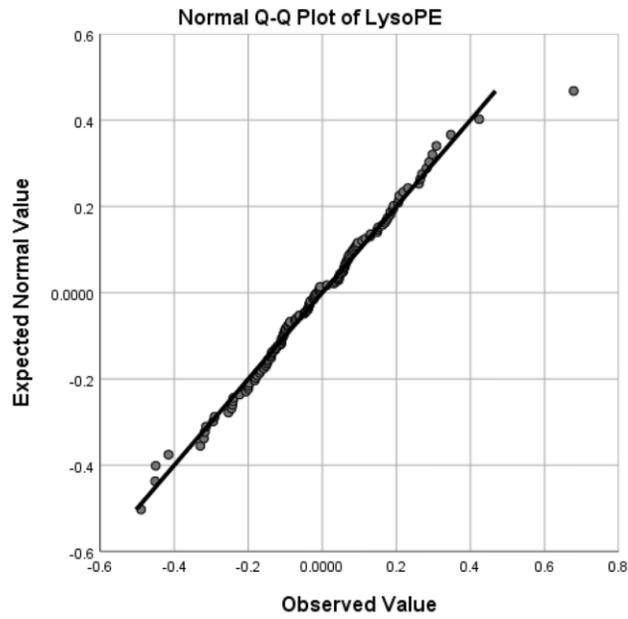
Supplementary Figure 63. The quantile-quantile (Q-Q) plot of normalized LysoPE which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of MannosylLtryptophan

Detrended Normal Q-Q Plot of MannosylLtryptophan
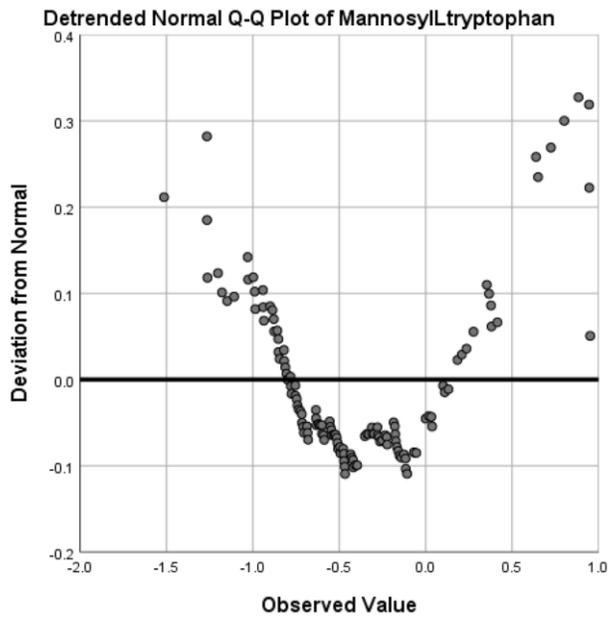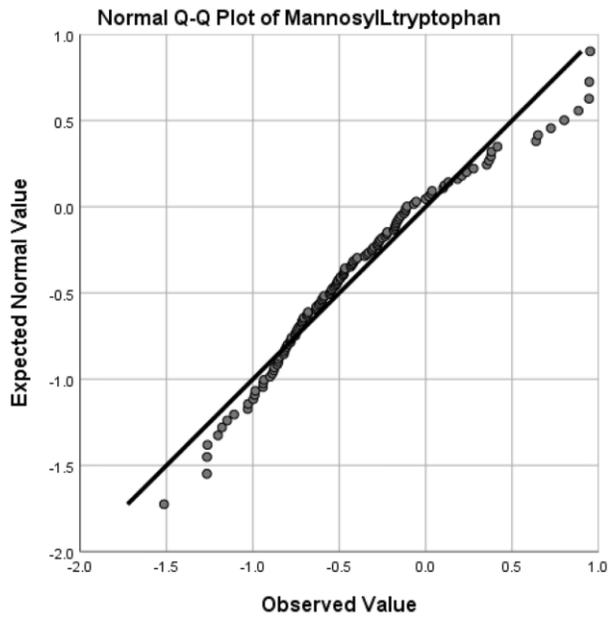
Supplementary Figure 64. The quantile-quantile (Q-Q) plot of normalized MannosylLtryptophan which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Acetylcarnosine

Detrended Normal Q-Q Plot of Acetylcarnosine
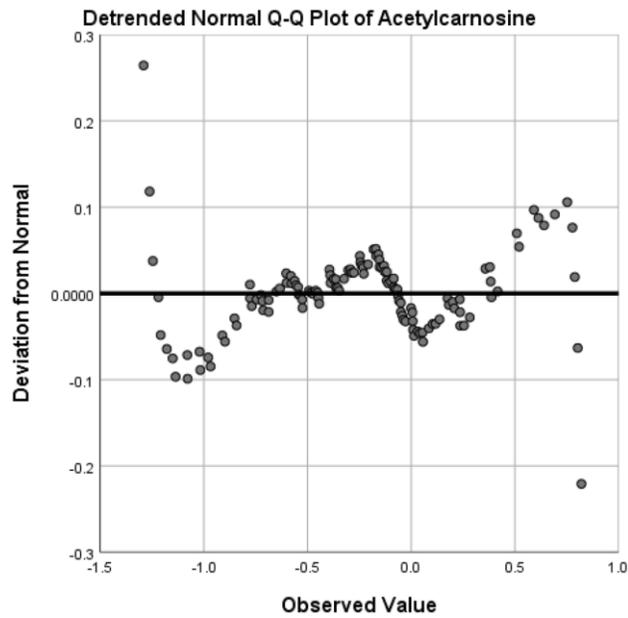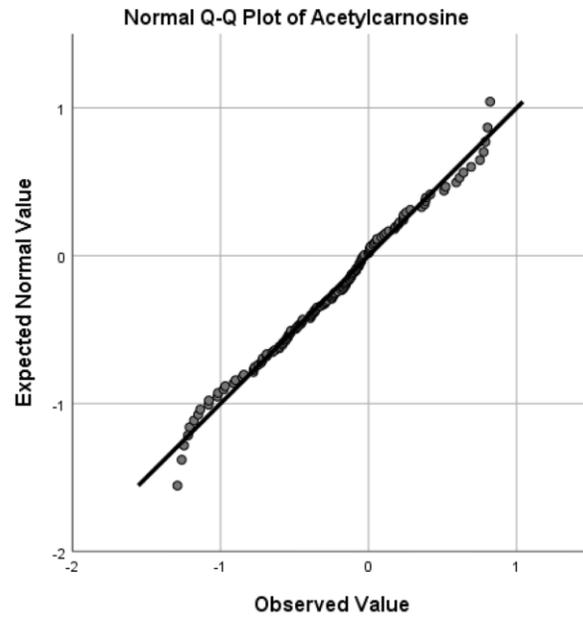
Supplementary Figure 65. The quantile-quantile (Q-Q) plot of normalized Acetylcarnosine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
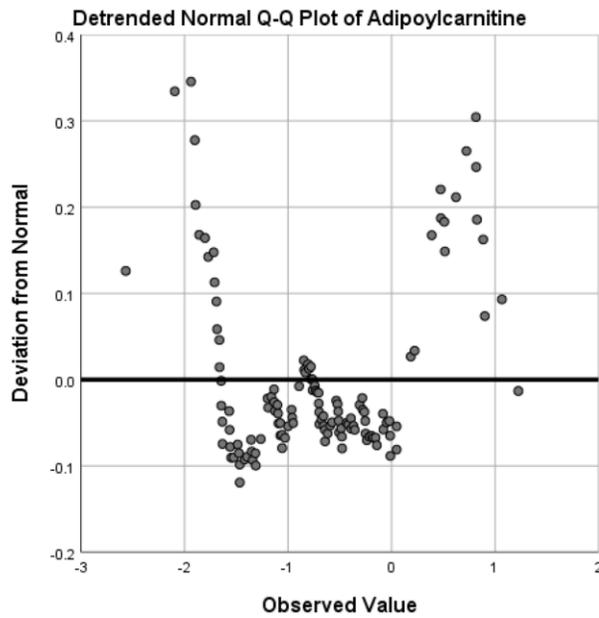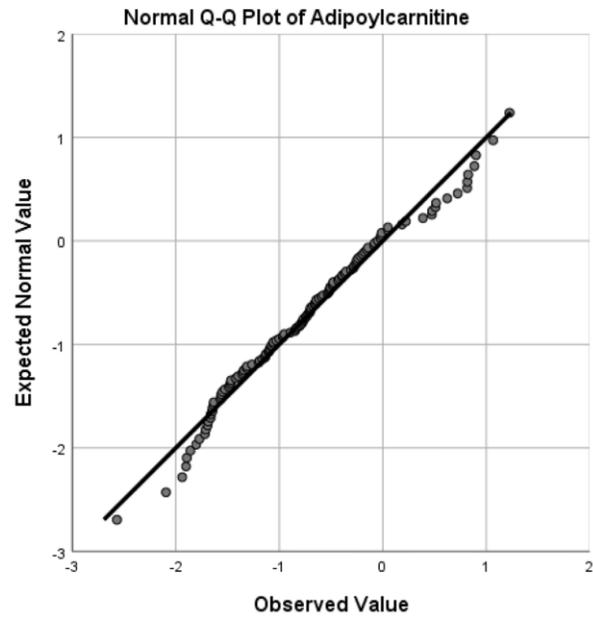
Supplementary Figure 66. The quantile-quantile (Q-Q) plot of normalized Adipoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
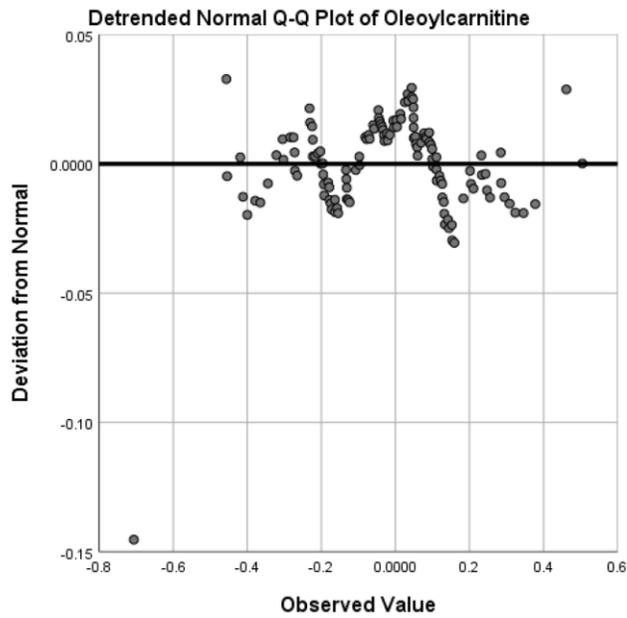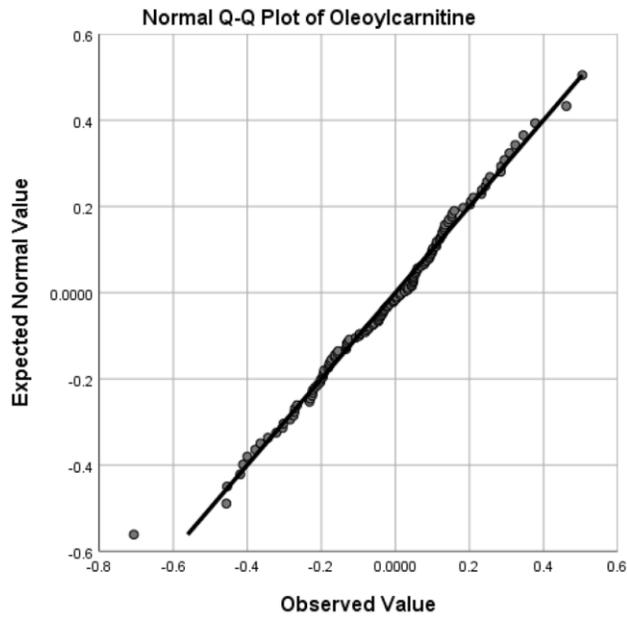
Supplementary Figure 67. The quantile-quantile (Q-Q) plot of normalized Oleoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Ornithine

Detrended Normal Q-Q Plot of Ornithine
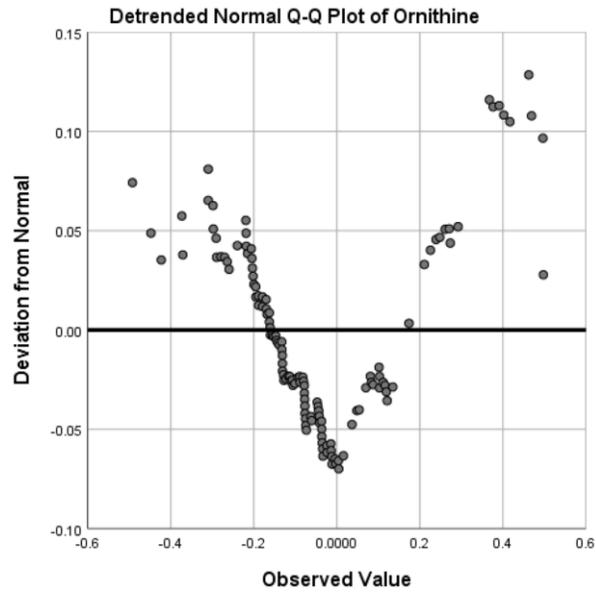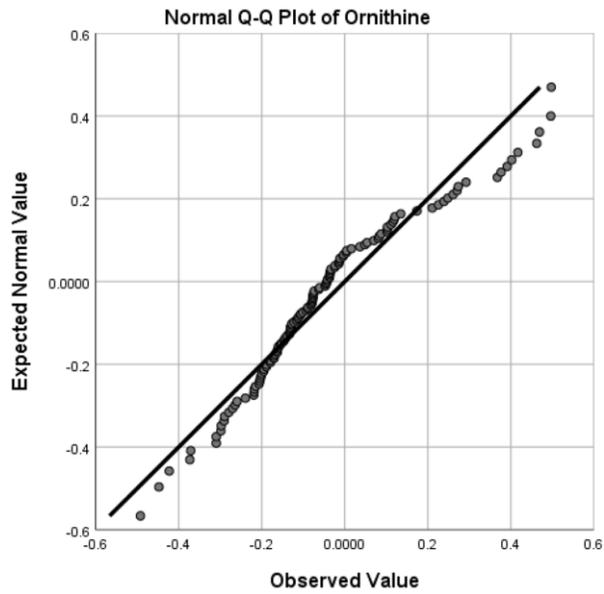
Supplementary Figure 68. The quantile-quantile (Q-Q) plot of normalized Ornithine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
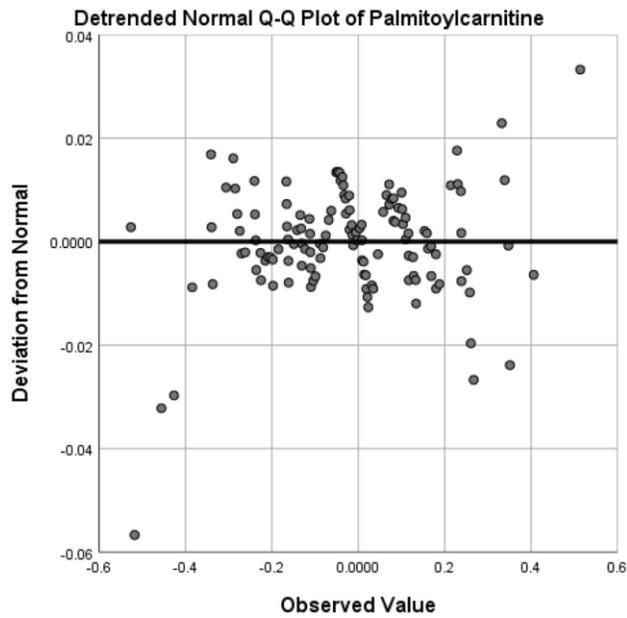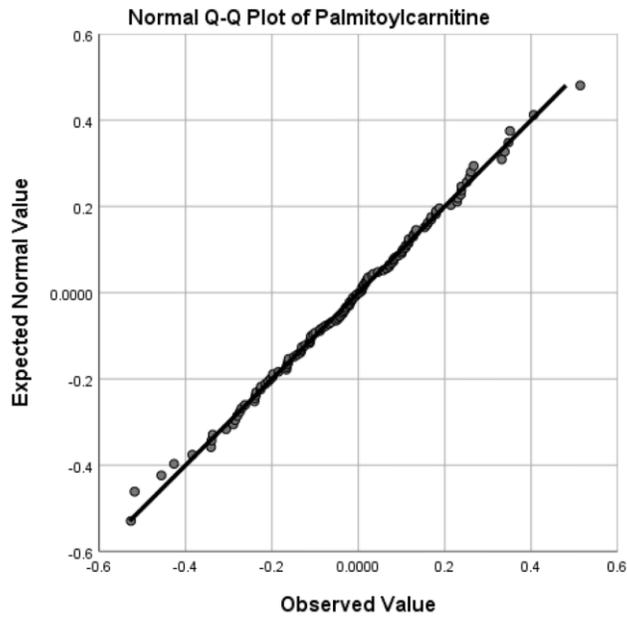
Supplementary Figure 69. The quantile-quantile (Q-Q) plot of normalized Palmitoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
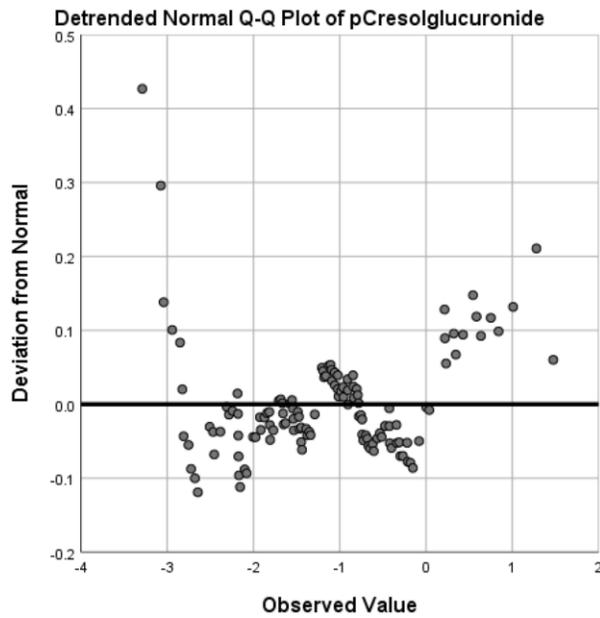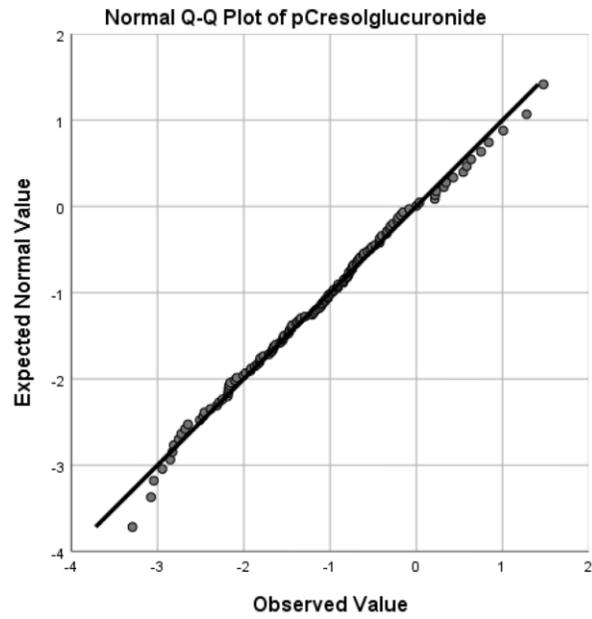
Supplementary Figure 70. The quantile-quantile (Q-Q) plot of normalized pCresolglucuronide which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
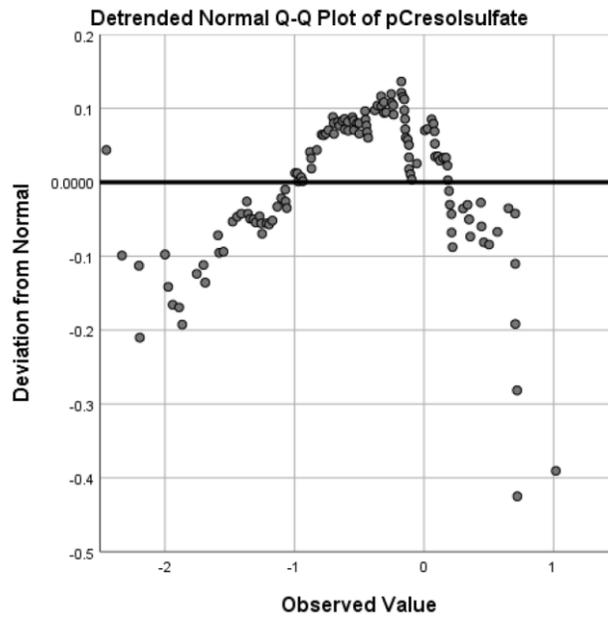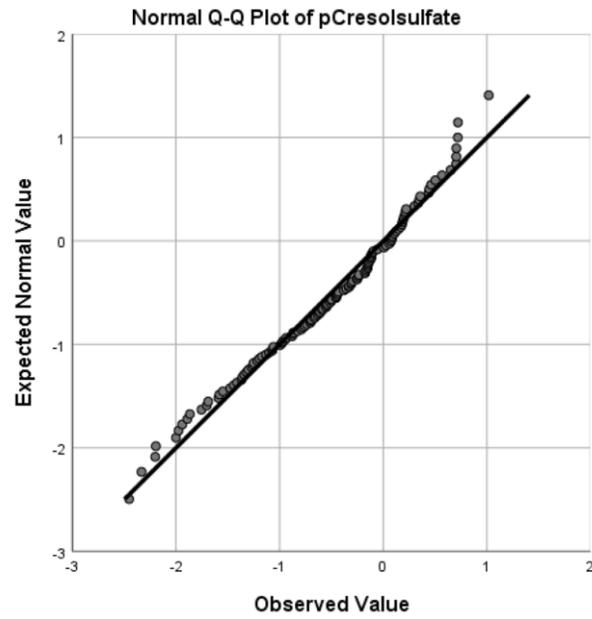
Supplementary Figure 71. The quantile-quantile (Q-Q) plot of normalized pCresolsulfate which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
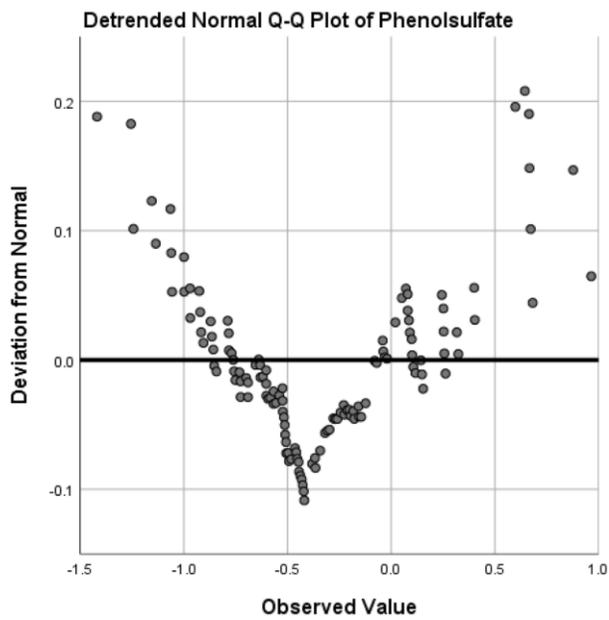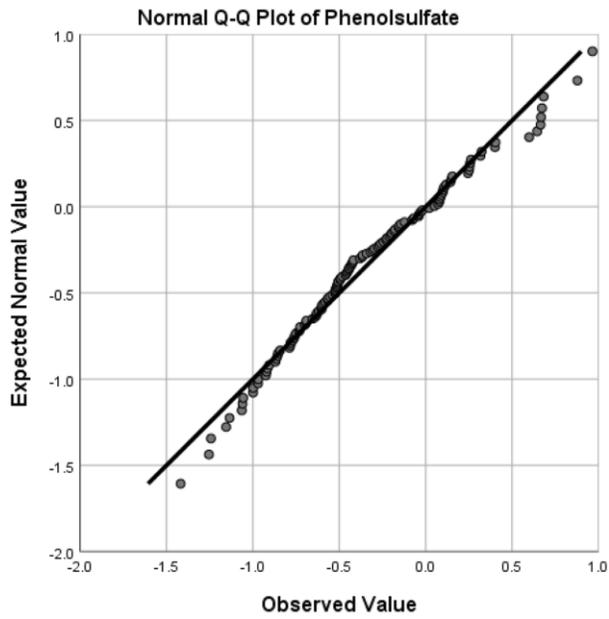
Supplementary Figure 72. The quantile-quantile (Q-Q) plot of normalized Phenolsulfate which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
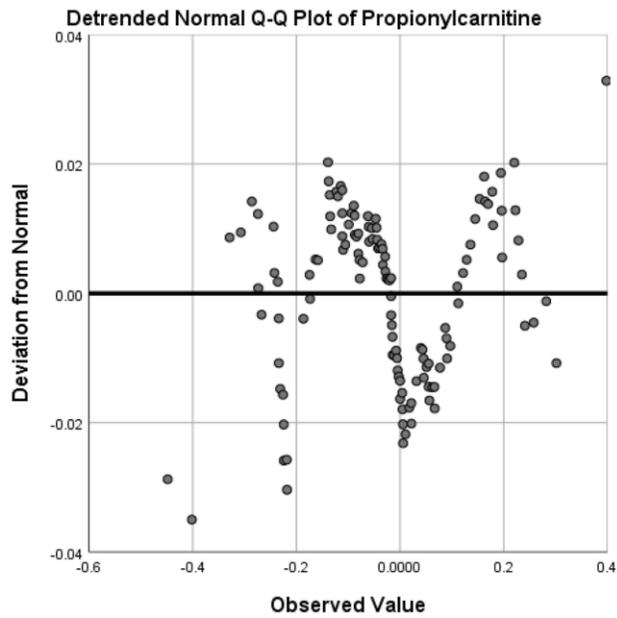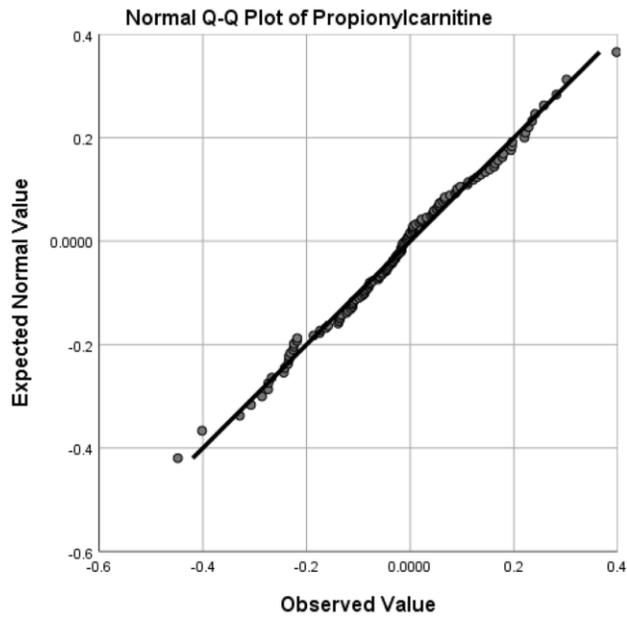
Supplementary Figure 73. The quantile-quantile (Q-Q) plot of normalized Propionylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
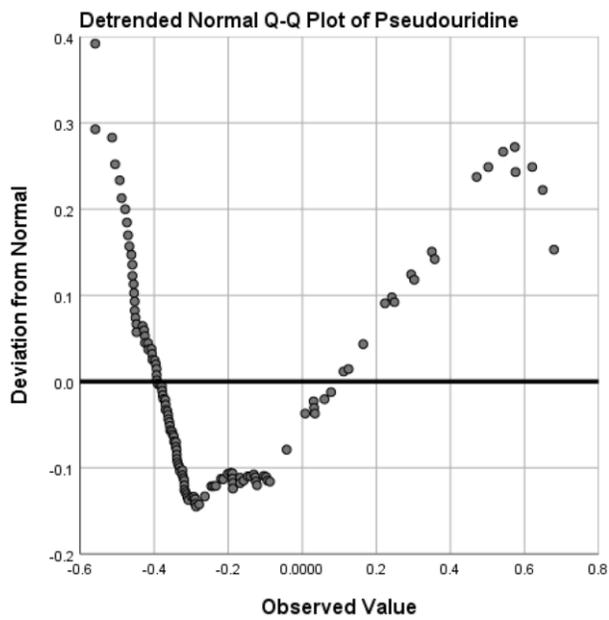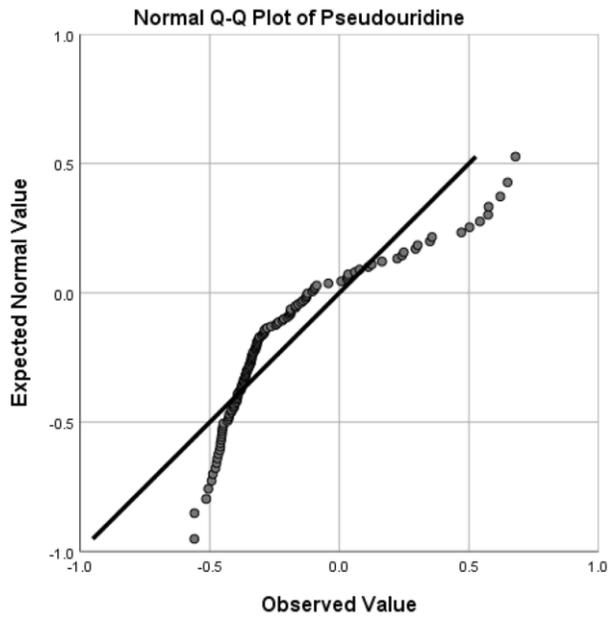
Supplementary Figure 74. The quantile-quantile (Q-Q) plot of normalized Pseudouridine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
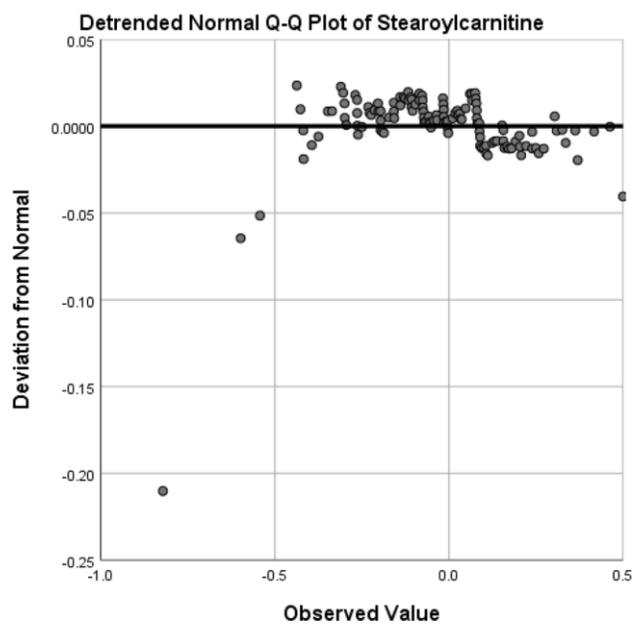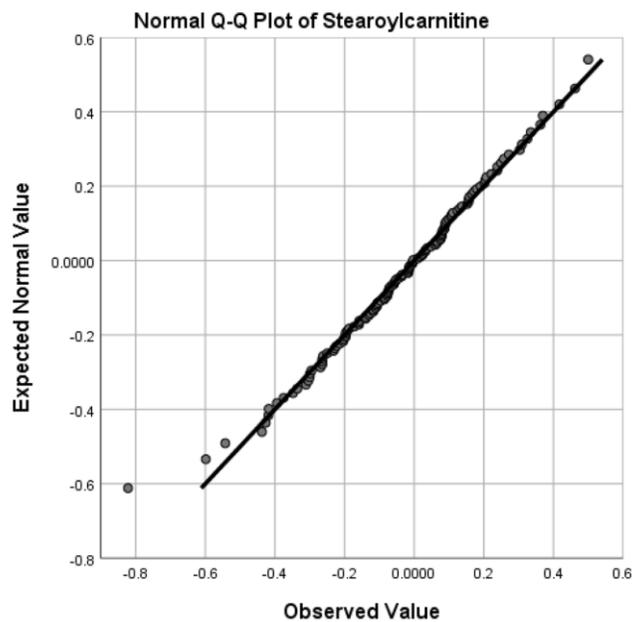
Supplementary Figure 75. The quantile-quantile (Q-Q) plot of normalized Stearoylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
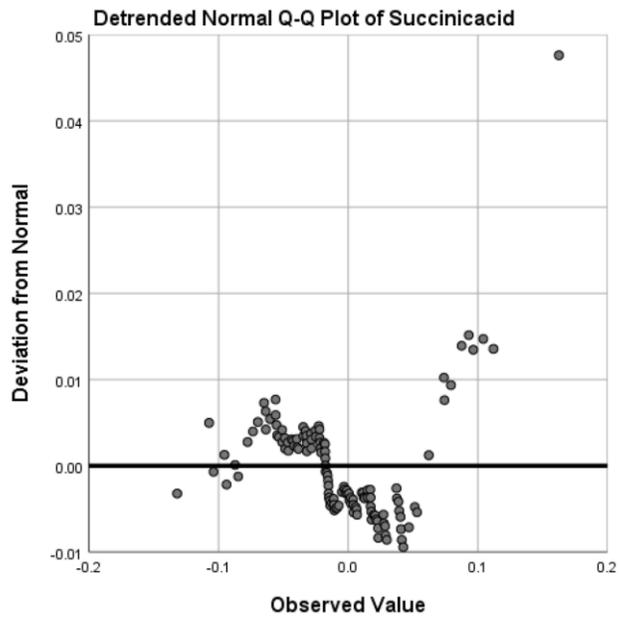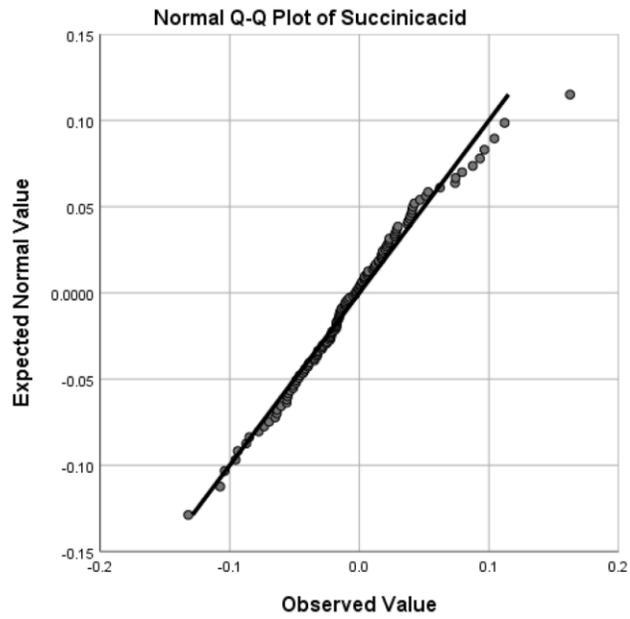
Supplementary Figure 76. The quantile-quantile (Q-Q) plot of normalized Succinicacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
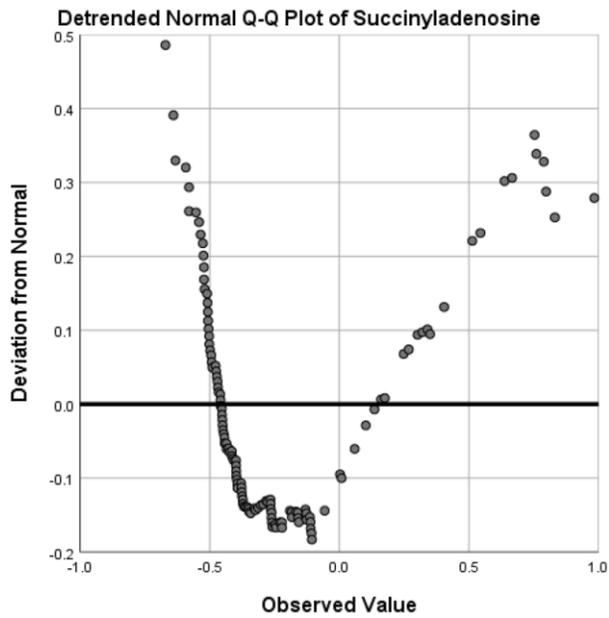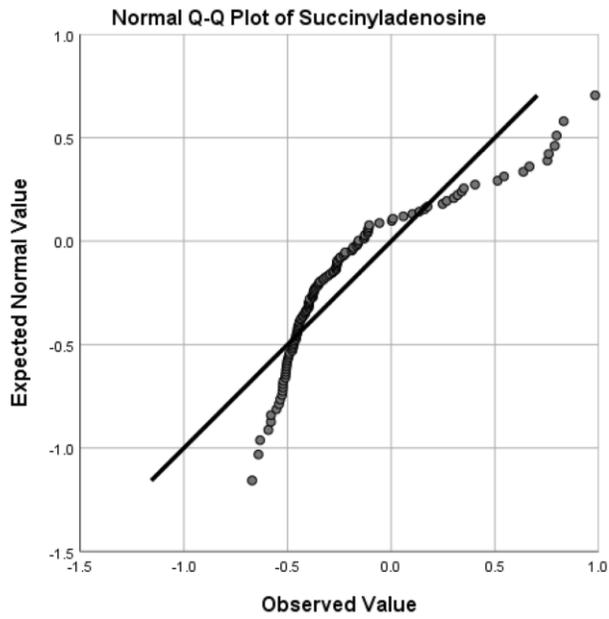
Supplementary Figure 77. The quantile-quantile (Q-Q) plot of normalized Succinyladenosine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
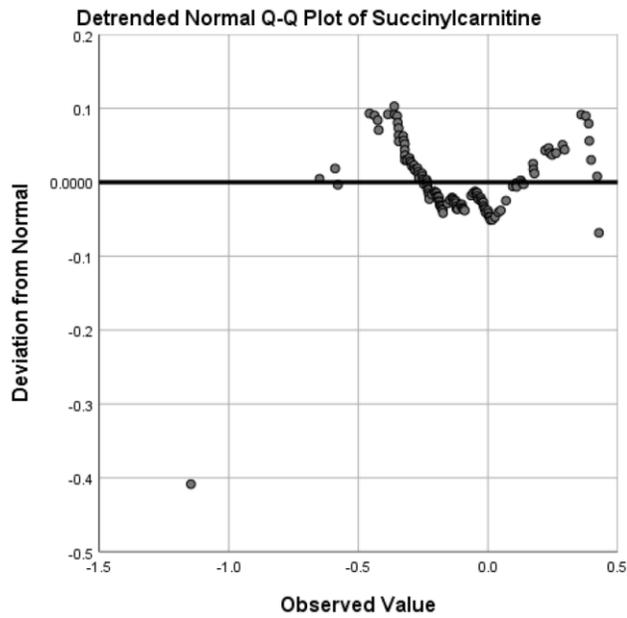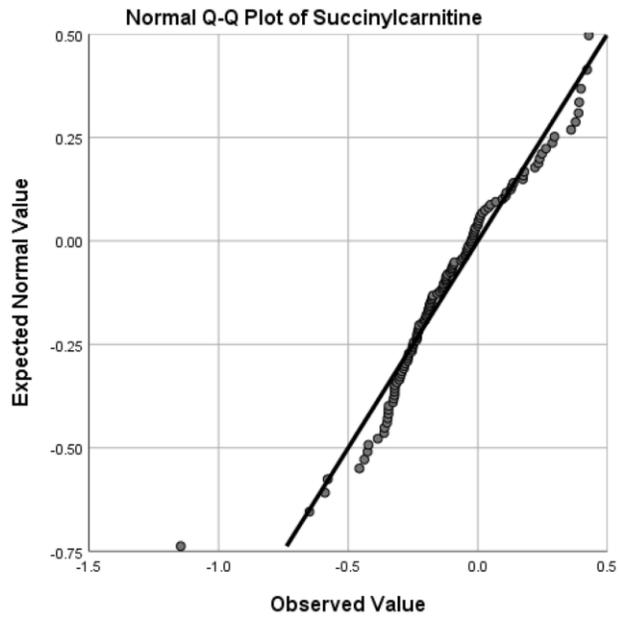
Supplementary Figure 78. The quantile-quantile (Q-Q) plot of normalized Succinylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
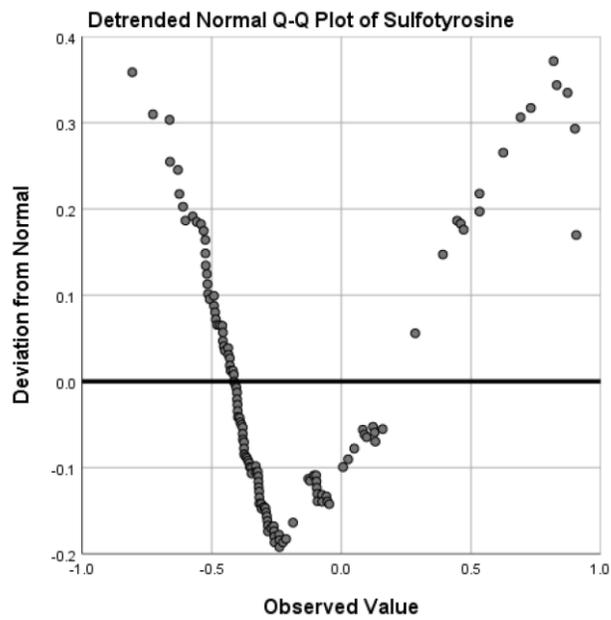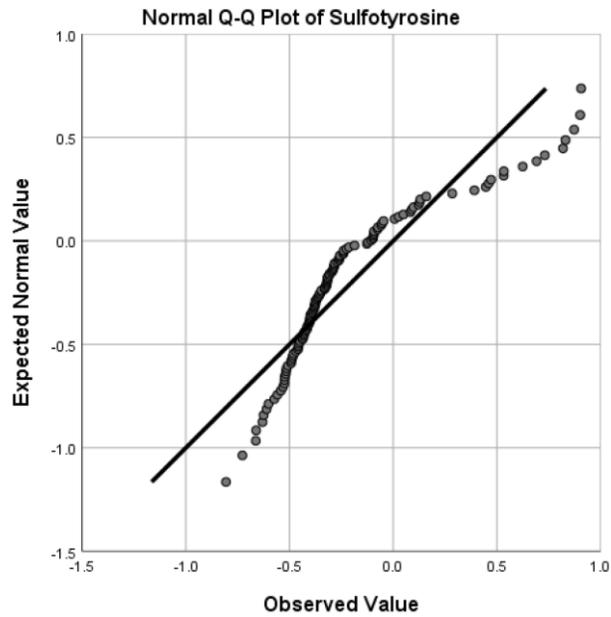
Supplementary Figure 79. The quantile-quantile (Q-Q) plot of normalized Sulfotyrosine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
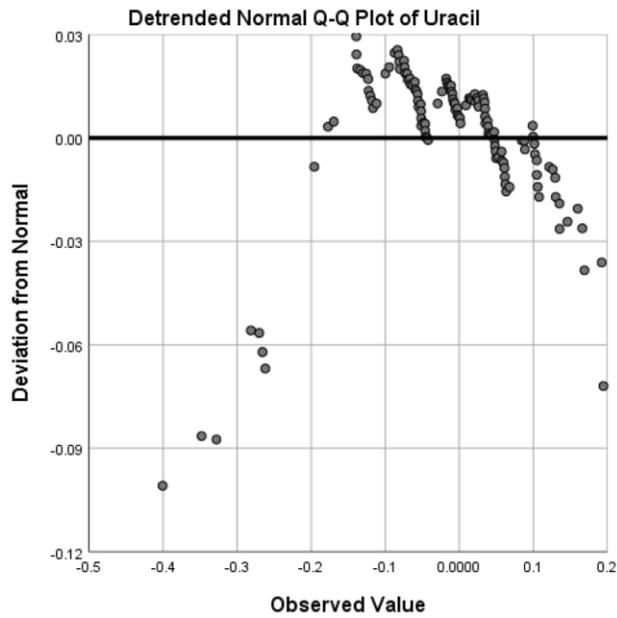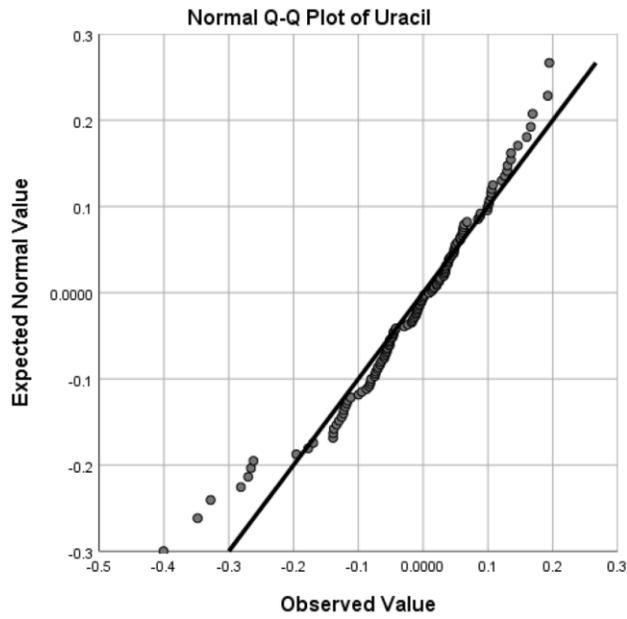
Supplementary Figure 80. The quantile-quantile (Q-Q) plot of normalized Uracil which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Uricacid

Detrended Normal Q-Q Plot of Uricacid
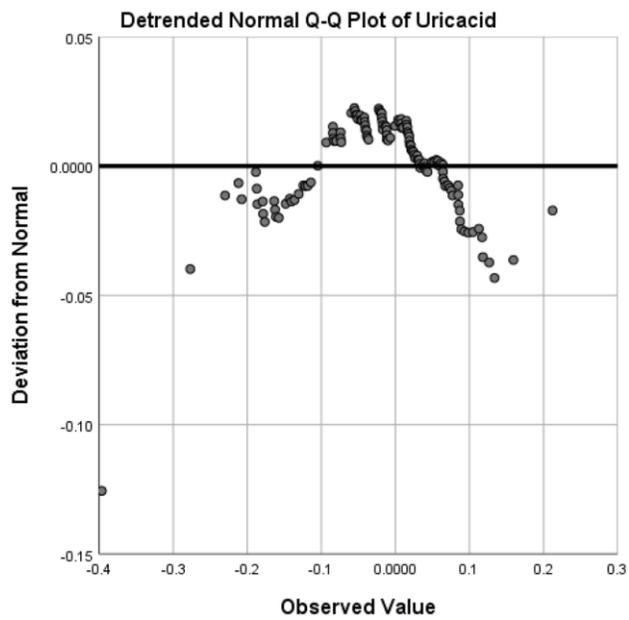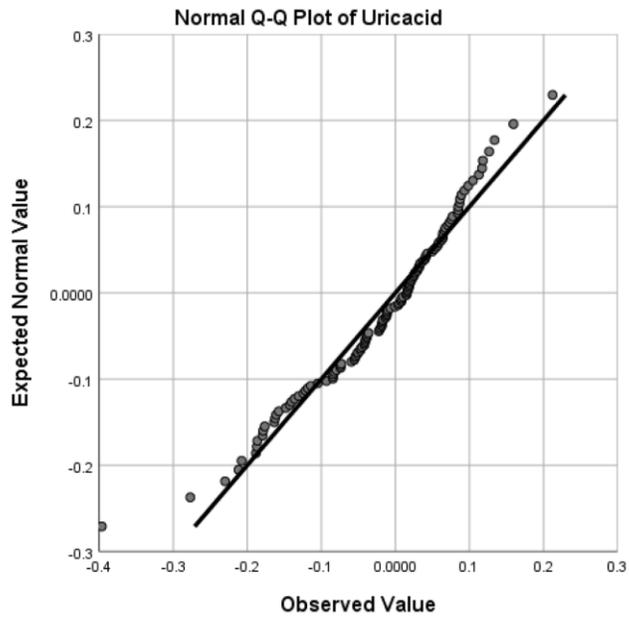
Supplementary Figure 81. The quantile-quantile (Q-Q) plot of normalized Uricacid which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).

Normal Q-Q Plot of Uridine

Detrended Normal Q-Q Plot of Uridine
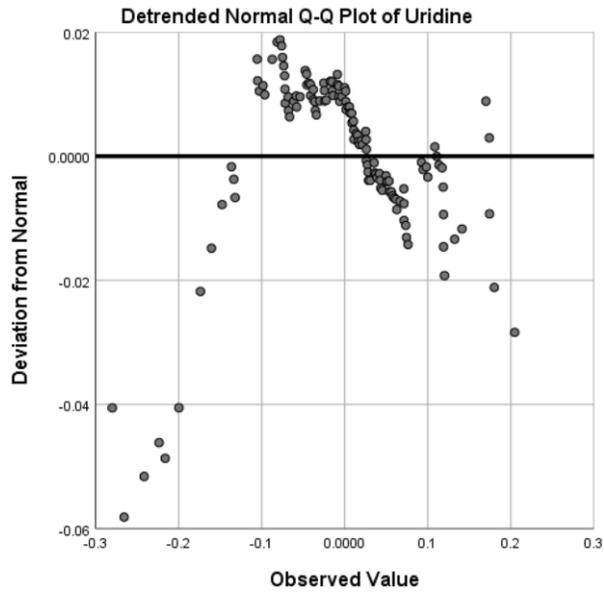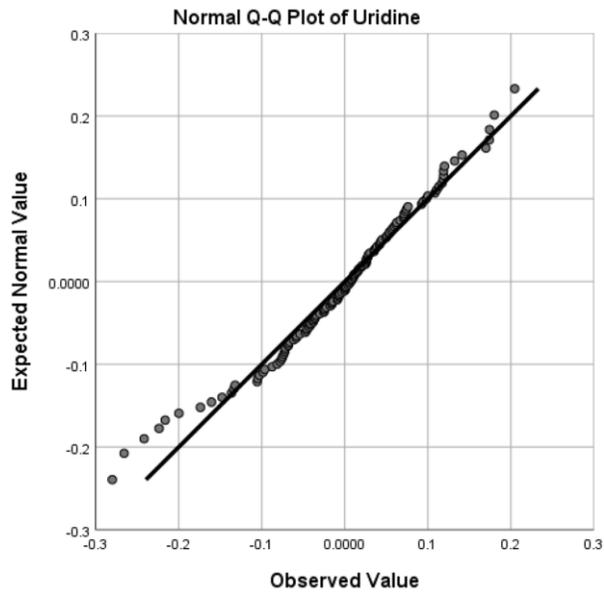
Supplementary Figure 82. The quantile-quantile (Q-Q) plot of normalized Uridine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
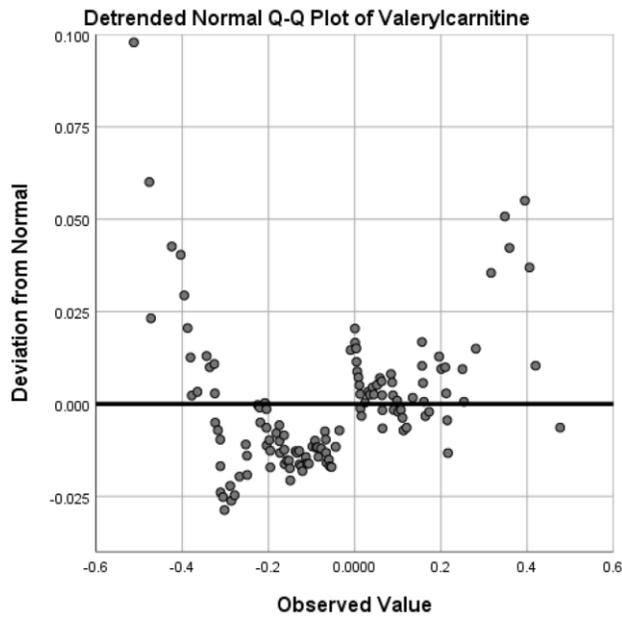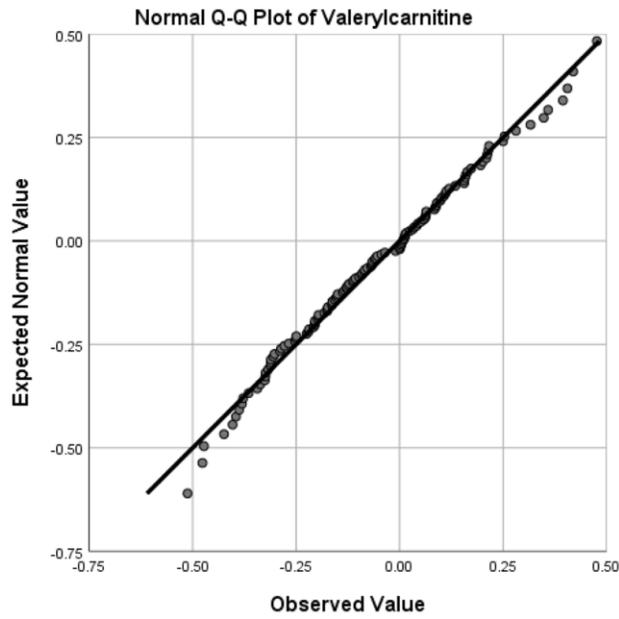
Supplementary Figure 83. The quantile-quantile (Q-Q) plot of normalized Valerylcarnitine which demonstrated the consistency of normalized value with expected normal value (upper panel) and the deviation of normalized values from expected normal value (lower panel).
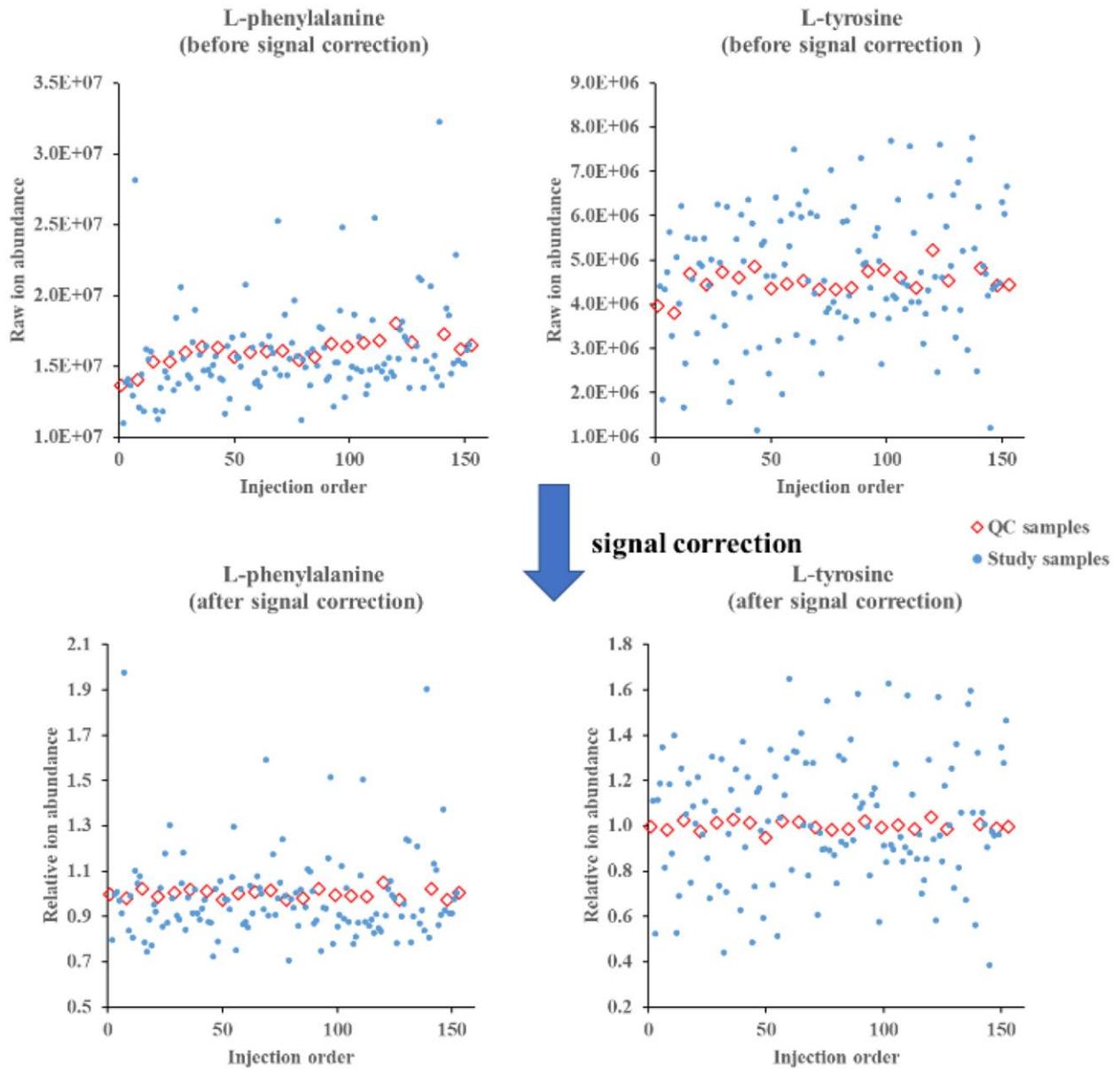
Supplementary Figure 84. Comparison of raw and relative ion abundances of L-tyrosine and L-phenylalanine in the study samples and QC samples before and after signal correction of data acquired by UPLC-Orbitrap-MS.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 2, 5, 6, 10 and 13.

Supplementary Table 1. Prevalence of microvascular and macrovascular complications in the patients with type 2 diabetes (*n* = 164) in each cohort.

	Diabetic retinopathy	Diabetic peripheral neuropathy	Macrovascular complication	At least one complication
Discovery Cohort (%)	21	38	43	58
Validation Cohort (%)	54	63	54	84
Overall (%)	32	46	46	67

Supplementary Table 2. Clinical features of participants in the discovery and validation cohorts.

Supplementary Table 3. The stability of relative ion abundance of internal standards in serum of subject samples and QC samples and overall features in QC samples.

Internal standards in serum of subject samples and QC samples							
Internal standards	**Dataset**	**Adduct**	**Retention time (min)**	**Detected m/z**	**Theoretical m/z**	**CV (%) of subject samples**	**CV (%) of quality control samples**
L-Tryptophan-(indole-D$_5$)	Discovery set	(M-H)−	2.95	208.1137	208.1140	18.04	2.16
	Validation set		2.96	208.1142	208.1140	27.08	1.61
	Discovery set	(M+H)+	2.95	210.1285	210.1285	4.24	2.43
	Validation set		2.96	210.1279	210.1285	12.14	4.28
Cholic acid D$_4$	Discovery set	(M-H)−	5.73	411.3054	411.3054	3.31	2.74
		(M+H-3H$_2$O)+	5.74	359.2878	359.2882	4.10	1.99
	Validation set	(M-H)−	5.72	411.3059	411.3054	3.44	1.31
		(M+H-3H$_2$O)+	5.73	359.2870	359.2882	9.82	2.84
***cis*-10-nonadecenoic acid**	Discovery set	(M-H)−	13.68	295.2642	295.2643	29.61	15.28
	Validation set	(M-H)−	13.85	295.2647	295.2643	25.67	11.37

Overall features in QC samples							
Dataset	**ESI mode**	**QC samples (*n*)**	**Total features**	**Number of features with RSD ≤ 30% in QC samples**	**Number of features with RSD ≤ 20% in QC samples**	**Percentage in all features (RSD ≤ 30%)**	**Percentage in all features (RSD ≤ 20%)**
Discovery set	−	22	5234	4316	3596	82.46	68.70
	+	25	4701	3164	2368	67.30	50.37
Validation set	−	12	8302	7249	6255	87.32	75.34
	+	12	12802	10279	8125	80.29	63.47

Supplementary Table 4. Identification details and stability of metabolites in QC samples using UPLC-Orbitrap-MS.

Identified metabolites	Retention time (min)	Detected m/z	Theoretical m/z	Mass error (ppm)	Adduct	Molecular formula	Confirmation/Supplier*	CV of QC (%) Discovery set	CV of QC (%) Validation set
L-Arginine	0.67	173.1042	173.1044	0.6	(M-H)−	C6H14N4O2	Sigma Aldrich	10.17	8.58
L-Ornithine	0.67	131.0823	131.0826	0	(M-H)−	C5H12N2O2	Sigma Aldrich	11.08	5.23
Choline	0.78	104.1071	104.1070	−2.9	(M)+	C5H14NO	Acros Organics	4.67	5.48
L-Glutamine	0.79	145.0616	145.0619	0.0	(M-H)−	C5H10N2O3	Sigma Aldrich	1.08	1.25
L-Citrulline	0.80	174.0882	174.0884	0.6	(M-H)−	C6H13N3O3	Sigma Aldrich	2.26	3.29
D-Glucose	0.80	215.0328	215.0328	1.4	(M+Cl)−	C6H12O6	International	1.14	0.68

							laboratory		
L-Carnitine	0.80	162.1125	162.1125	−3.1	(M+H)⁺	C7H15NO3	CIL	1.50	3.63
L−Glutamic acid	0.80	148.0604	148.0604	−2.7	(M+H)⁺	C5H9NO4	Sigma Aldrich	3.80	4.09
L−Threonine	0.80	120.0655	120.0655	−2.5	(M+H)⁺	C4H9NO3	Sigma Aldrich	6.47	5.57
Arabinose isomer	0.80	195.0513	195.0510	1.5	(M+FA−H)⁻	C5H10O5	Santa Cruz	1.46	1.88
Betaine	0.81	118.0863	118.0863	−3.4	(M+H)⁺	C5H11NO2	Sigma Aldrich	1.70	3.06
Creatinine	0.81	114.0662	114.0662	−2.6	(M+H)⁺	C4H7N3O	Acros Organics	2.24	3.93
2−Hydroxyethanesulfonate	0.82	124.9911	124.9914	0.0	(M−H)⁻	C2H6O4S	Sigma Aldrich	3.78	8.46
γ−Butyrobetaine	0.83	146.1175	146.1176	−3.4	(M+H)⁺	C7H15NO2	TRC	3.76	2.37
L−Proline	0.83	116.0707	116.0706	−2.6	(M+H)⁺	C5H9NO2	Sigma Aldrich	1.56	3.69
1,5−Anhydro−D−glucitol	0.85	199.0376	199.0379	1.0	(M+Cl)⁻	C6H12O5	TRC	2.09	1.82
N−Acetylcarnosine	0.89	269.1243	269.1243	−2.6	(M+H)⁺	C11H16N4O4	Santa Cruz	5.91	7.34
5−Methylthio−D−ribose	0.93	181.0529	181.0529	−2.8	(M+H)⁺	C6H12O4S	Online database (HMDB)	6.12	8.15
Pseudouridine	0.93	243.0622	243.0623	1.2	(M−H)⁻	C9H12N2O6	Supelco	1.78	3.65
L−Valine	0.94	118.0863	118.0863	−3.4	(M+H)⁺	C5H11NO2	Sigma Aldrich	1.89	4.29
L−Acetylcarnitine	0.95	204.1230	204.1230	−2.9	(M+H)⁺	C9H17NO4	CIL	2.35	6.67
L,L−TMAP isomer	0.96	229.1546	229.1547	−3.5	(M+H)⁺	C11H20N2O3	ChemPartner [3]	4.05	6.07
Uric acid	1.00	169.0356	169.0356	−3.0	(M+H)⁺	C5H4N4O3	Sigma Aldrich	1.78	4.69
L,L−TMAP	1.06	229.1546	229.1547	−3.5	(M+H)⁺	C11H20N2O3	ChemPartner [3]	2.31	5.44
L−Methionine	1.07	148.0435	148.0438	1.8	(M−H)⁻	C5H11NO2S	Sigma Aldrich	5.03	7.31
Citric acid	1.09	191.0195	191.0197	1.0	(M−H)⁻	C6H8O7	Sigma Aldrich	3.55	3.64
Hydroxybutyrylcarnitine	1.09	248.1491	248.1492	−3.6	(M+H)⁺	C11H21NO5	Online databases	11.84	8.58
Succinylcarnitine	1.09	262.1284	262.1285	−3.8	(M+H)⁺	C11H19NO6	Supelco	5.29	11.02
Uracil	1.09	113.0346	113.0346	−2.7	(M+H)⁺	C4H4N2O2	Wako	5.11	6.35
Uridine	1.09	243.0622	243.0623	1.2	(M−H)⁻	C9H12N2O6	Wako	1.69	1.14
L−Tyrosine	1.33	180.0664	180.0666	0.6	(M−H)⁻	C9H11NO3	Sigma Aldrich	2.11	1.06
Sulfotyrosine	1.46	260.0234	260.0234	1.9	(M−H)⁻	C9H11NO6S	Ref [4]	1.79	0.88
Inosine	1.50	267.0735	267.0735	1.9	(M−H)⁻	C10H12N4O5	Acros Organics	2.30	1.17
L−Leucine	1.56	132.1019	132.1019	−3.0	(M+H)⁺	C6H13NO2	Sigma Aldrich	1.81	6.27
4−Acetamidobutanoic acid	1.59	144.0664	144.0666	0.7	(M−H)⁻	C6H11NO3	Matrix Scientific	1.62	1.44
Propionylcarnitine	1.67	218.1387	218.1387	−3.2	(M+H)⁺	C10H19NO4	CIL	2.61	6.39
2−Hydroxybutyric acid	1.90	103.0398	103.0401	−1.0	(M−H)⁻	C4H8O3	Sigma Aldrich	1.85	1.35
2−(α−D−Mannopyranosyl)−L−tryptophan	2.21	367.1497	367.1500	−3.8	(M+H)⁺	C17H22N2O7	TRC	2.16	6.02
L−Kynurenine	2.42	209.0921	209.0921	−3.3	(M+H)⁺	C10H12N2O3	Sigma Aldrich	3.00	8.09
L−Phenylalanine	2.48	164.0714	164.0717	0.6	(M−H)⁻	C9H11NO2	Sigma Aldrich	2.02	2.36
Succinyladenosine	2.89	382.1005	382.1004	1.6	(M−H)⁻	C14H17N5O8	TRC	3.06	3.07
O−Adipoylcarnitine	2.90	290.1597	290.1598	−3.8	(M+H)⁺	C13H23NO6	Supelco	3.39	5.11
Butyrylcarnitine	2.92	232.1543	232.1543	−3.4	(M+H)⁺	C11H21NO4	CIL	2.62	5.65
L−β−aspartyl−L−leucine	2.93	247.1287	247.1288	−3.2	(M+H)⁺	C10H18N2O5	Online database (HMDB)	6.71	4.89
L−Tryptophan	2.95	203.0824	203.0826	1.0	(M−H)⁻	C11H12N2O2	Sigma Aldrich	2.21	1.26
Homovanillic acid sulfate	2.97	261.0073	261.0074	1.9	(M−H)⁻	C9H10O7S	Cayman Chemical	12.40	2.42
Kynurenic acid	2.99	190.0499	190.0499	−3.2	(M+H)⁺	C10H7NO3	Sigma Aldrich	14.71	4.95
2−(3−(sulfooxy)phenyl)acetic acid	3.00	230.9967	230.9969	1.3	(M−H)⁻	C8H8O6S	Online database (HMDB)	2.64	4.81
Valerylcarnitine	3.00	246.1699	246.1700	−3.7	(M+H)⁺	C12H23NO4	Cayman Chemical	5.96	7.49
Pyrocatechol sulfate	3.07	188.9865	188.9863	1.1	(M−H)⁻	C6H6O5S	Online database (HMDB)	2.82	1.57
α−N−Phenylacetyl−L−glut	3.07	263.1037	263.1037	1.9	(M−H)⁻	C13H16N2O4	Santa Cruz	1.87	1.41

amine

Phenol sulfate	3.10	172.9912	172.9914	1.2	(M−H)−	C6H6O4S	Online databases	2.32	0.99
Hexanoylcarnitine	3.11	260.1855	260.1856	−3.5	(M+H)+	C13H25NO4	Santa Cruz	28.21	8.88
Hippuric acid	3.14	178.0508	178.0510	0.6	(M−H)−	C9H9NO3	Acros Organics	2.03	0.91
Indoxyl sulfate	3.15	212.0022	212.0023	1.4	(M−H)−	C8H7NO4S	Sigma Aldrich	2.82	1.43
p−**Cresol glucuronide**	3.17	283.0823	283.0823	1.8	(M−H)−	C13H16O7	TRC	2.09	1.42
2−Octenoylcarnitine	3.32	286.2011	286.2013	−3.8	(M+H)+	C15H27NO4	Online databases	3.15	6.83
p−**Cresol sulfate**	3.36	187.0070	187.0071	1.1	(M−H)−	C7H8O4S	CIL	1.82	1.00
Indole−3−lactic acid	3.44	204.0664	204.0666	1.0	(M−H)−	C11H11NO3	Santa Cruz	2.25	1.28
3−Hydroxydecanoyl carnitine	3.62	332.2429	332.2431	−3.6	(M+H)+	C17H33NO5	Online databases	5.28	5.99
L−Octanoylcarnitine	3.65	288.2167	288.2169	−3.5	(M+H)+	C15H29NO4	CIL	2.63	5.72
3−Indoleacetic acid	3.78	176.0706	176.0706	−2.8	(M+H)+	C10H9NO2	Sigma Aldrich	3.54	5.14
Cortisol	4.03	363.2163	363.2166	−3.6	(M+H)+	C21H30O5	Sigma Aldrich	3.00	3.57
9−Decenoylcarnitine	4.13	314.2324	314.2326	−3.8	(M+H)+	C17H31NO4	Online databases	2.18	6.51
Bilirubin	4.30	585.2706	585.2708	−2.7	(M+H)+	C33H36N4O6	Acros Organics	4.95	5.13
Decanoylcarnitine	4.51	316.2480	316.2482	−3.8	(M+H)+	C17H33NO4	Sigma Aldrich	2.80	2.82
Dehydroepiandrosterone sulfate	4.73	367.1584	367.1585	1.1	(M−H)−	C19H28O5S	Cayman Chemical	3.63	1.64
3,5−Tetradecadiencarnitine	5.49	368.2793	368.2795	−3.5	(M+H)+	C21H37NO4	Online databases	5.88	6.73
Dodecanoylcarnitine	5.57	344.2793	344.2795	−3.5	(M+H)+	C19H37NO4	CIL	9.44	4.34
cis−**5−Tetradecenoylcarnitine**	6.39	370.2949	370.2952	−3.5	(M+H)+	C21H39NO4	Online databases	12.23	9.59
LysoPE(18:1(11Z)/0:0)	10.46	480.3083	480.3085	−2.9	(M+H)+	C23H46NO7P	Online databases	14.54	9.97
Stearoylcarnitine	10.52	428.3732	428.3734	−3.3	(M+H)+	C25H49NO4	CIL	16.91	8.10

Supplementary Table 5. Metabolites that had significant fold changes at different stages with respect to the healthy subjects in both discovery and validation sets.

Supplementary Table 6. Metabolites that had significant fold changes at different stages with respect to their later stage in both discovery and validation sets.

Supplementary Table 7. Spearman rank correlation of metabolites with total BSA-related renal volume and renal resistive index in validation set (absolute Spearman *R* > 0.40).

Metabolites	**Total BSA-related renal volume**		**Renal resistive index**
	Validation (Stage 0–4) (*n* = 58)	**Validation (Stage 1–4)** (*n* = 48)	**Validation (Stage 1–4)** (*n* = 48)
MDRD eGFR	0.542	0.638	−0.525
Bilirubin	0.539	0.553	−0.423
Stearoylcarnitine	0.430	0.469	−0.148
Ratio of tyrosine to phenylalanine	0.387	0.425	−0.468
L−Tryptophan	0.319	0.375	−0.569
L−Tyrosine	0.309	0.326	−0.434
SBP(mmHg)	−0.082	−0.129	0.596
N−Acetylcarnosine	−0.137	−0.228	0.434
Succinylcarnitine	−0.206	−0.359	0.401

Ratio of kynurenic acid to kynurenine	−0.234	−0.309	0.499
Citric acid	−0.245	−0.322	0.484
Phenol sulfate	−0.248	−0.318	0.451
Arabinose isomer	−0.264	−0.402	0.607
Indolelactic acid	−0.307	−0.409	0.448
3−hydroxydecanoyl carnitine	−0.310	−0.489	0.384
Indoleacetic acid	−0.320	−0.444	0.196
2−Octenoylcarnitine	−0.342	−0.486	0.239
MS−detected creatinine	−0.352	−0.473	0.459
Ratio of pseudouridine to uridine	−0.369	−0.528	0.544
L,L−TMAP isomer	−0.376	−0.560	0.495
O−Adipoylcarnitine	−0.390	−0.578	0.356
Urea (mmol/L)	−0.391	−0.535	0.554
Renal resistive index	−0.392	−0.392	−
2−(3−(sulfooxy)phenyl)acetic acid	−0.401	−0.516	0.407
2−Hydroxyethanesulfonate	−0.429	−0.488	0.503
Kynurenic acid	−0.432	−0.537	0.567
Homovanillic acid sulfate	−0.436	−0.547	0.455
4−Acetamidobutanoic acid	−0.440	−0.557	0.494
Sulfotyrosine	−0.441	−0.572	0.535
Serum Creatinine (mg/dL)	−0.443	−0.571	0.446
L,L−TMAP	−0.449	−0.596	0.517
Butyrylcarnitine	−0.457	−0.617	0.369
Pseudouridine	−0.466	−0.599	0.588
5−Methylthio−D−ribose	−0.475	−0.618	0.466
2−(α−D−Mannopyranosyl)−L−tryptophan	−0.480	−0.604	0.586
L−β−aspartyl−L−leucine	−0.487	−0.589	0.578
Serum cystatin C (mg/L)	−0.503	−0.614	0.544
Succinyladenosine	−0.543	−0.615	0.556
L−Kynurenine	−0.562	−0.683	0.472
Ratio of kynurenine to tryptophan	−0.563	−0.640	0.613

Supplementary Table 8. Average AUC of single biomarkers for differentiation between stages in diabetic patients using RF classification in both discovery and validation sets.

Classification	Dataset	RF	MS-detected serum creatinine	Pseudouridine	L,L-TMAP	2-(α-D-Mannopyranosyl)-L-tryptophan	Succinyladenosine
Stage 1a vs. Stage 1b-4	Dis	AUC	0.849 ± 0.051	0.915 ± 0.054	0.898 ± 0.046	0.888 ± 0.066	0.888 ± 0.058
		Sens	0.721 ± 0.118	0.933 ± 0.053	0.810 ± 0.066	0.832 ± 0.120	0.832 ± 0.059
		Spec	0.876 ± 0.120	0.688 ± 0.148	0.902 ± 0.132	0.753 ± 0.156	0.791 ± 0.158
	Val	AUC	0.879 ± 0.078	0.915 ± 0.054	0.810 ± 0.066	0.893 ± 0.065	0.923 ± 0.050
		Sens	0.920 ± 0.100	0.933 ± 0.053	0.902 ± 0.132	0.852 ± 0.074	0.866 ± 0.067
		Spec	0.674 ± 0.183	0.688 ± 0.148	0.939 ± 0.022	0.806 ± 0.252	0.824 ± 0.179
Stage 1a vs. Stage 1b-2	Dis	AUC	0.754 ± 0.080	0.867 ± 0.072	0.848 ± 0.057	0.811 ± 0.074	0.816± 0.065
		Sens	0.643 ± 0.102	0.881 ± 0.089	0.751 ± 0.078	0.840 ± 0.093	0.688 ± 0.122
		Spec	0.805 ± 0.162	0.696 ± 0.138	0.812 ± 0.176	0.704 ± 0.145	0.807 ± 0.167
	Val	AUC	0.763 ± 0.109	0.757 ± 0.078	0.774 ± 0.115	0.780 ± 0.099	0.857 ± 0.066
		Sens	0.909 ± 0.115	0.651 ± 0.142	0.864 ± 0.121	0.672 ± 0.141	0.819 ± 0.143
		Spec	0.636 ± 0.209	0.760 ± 0.243	0.494 ± 0.209	0.794 ± 0.234	0.730 ± 0.173

Comparison	Set	Metric					
		AUC	0.681 ± 0.094	0.741 ± 0.079	0.680 ± 0.078	0.707 ± 0.097	0.777 ± 0.069
	Dis	Sens	0.785 ± 0.177	0.679 ± 0.108	0.518 ± 0.127	0.687 ± 0.180	0.556 ± 0.121
Stage 1b vs. Stage 2		Spec	0.447 ± 0.187	0.772 ± 0.153	0.760 ± 0.185	0.598 ± 0.190	0.816 ± 0.170
		AUC	0.600 ± 0.170	0.982 ± 0.031	0.668 ± 0.156	0.870 ± 0.092	0.889 ± 0.078
	Val	Sens	0.480 ± 0.238	0.900 ± 0.100	0.478 ± 0.245	0.842 ± 0.258	0.766 ± 0.221
		Spec	0.684 ± 0.266	0.916 ± 0.150	0.778 ± 0.257	0.792 ± 0.152	0.802 ± 0.189
		AUC	0.918 ± 0.028	0.935 ± 0.035	0.906 ± 0.038	0.935 ± 0.026	0.899 ± 0.038
	Dis	Sens	0.724 ± 0.071	0.874 ± 0.058	0.782 ± 0.057	0.778 ± 0.064	0.787 ± 0.059
Stage 1 vs. Stage 2-4		Spec	0.907 ± 0.095	0.905 ± 0.083	0.939 ± 0.081	0.920 ± 0.098	0.902 ± 0.092
		AUC	0.927 ± 0.048	0.994 ± 0.010	0.928 ± 0.046	0.954 ± 0.034	0.971 ± 0.031
	Val	Sens	0.811 ± 0.081	0.966 ± 0.036	0.784 ± 0.085	0.939 ± 0.094	0.931 ± 0.079
		Spec	0.910 ± 0.145	0.976 ± 0.063	0.929 ± 0.141	0.857 ± 0.083	0.925 ± 0.106
		AUC	0.950 ± 0.025	0.982 ± 0.021	0.969 ± 0.020	0.974 ± 0.021	0.958 ± 0.028
	Dis	Sens	0.919 ± 0.079	0.904 ± 0.076	0.949 ± 0.057	0.923 ± 0.065	0.912 ± 0.075
Stage 1-2 vs. Stage 3-4		Spec	0.880 ± 0.080	0.949 ± 0.044	0.851 ± 0.074	0.898 ± 0.055	0.841 ± 0.077
		AUC	0.931 ± 0.039	0.951 ± 0.034	0.915 ± 0.046	0.939 ± 0.039	0.912 ± 0.044
	Val	Sens	0.932 ± 0.066	0.905 ± 0.096	0.845 ± 0.111	0.786 ± 0.133	0.899 ± 0.086
		Spec	0.817 ± 0.109	0.883 ± 0.090	0.858 ± 0.088	0.863 ± 0.136	0.843 ± 0.100
		AUC	0.989 ± 0.009	0.992 ± 0.020	0.982 ± 0.034	0.988 ± 0.024	0.993 ± 0.018
	Dis	Sens	0.967 ± 0.074	0.914 ± 0.096	0.934 ± 0.114	0.926 ± 0.098	0.942 ± 0.093
Stage 1-3 vs. Stage 4		Spec	0.975 ± 0.015	0.973 ± 0.027	0.950 ± 0.028	0.973 ± 0.028	0.983 ± 0.015
		AUC	0.976 ± 0.081	0.968 ± 0.061	0.984 ± 0.054	0.953 ± 0.055	0.955 ± 0.051
	Val	Sens	0.952 ± 0.163	0.882 ± 0.168	0.938 ± 0.135	0.906 ± 0.111	0.886 ± 0.124
		Spec	1.000 ± <0.001	0.983 ± 0.022	1.000 ± <0.001	0.995 ± 0.020	0.987 ± 0.039

Data were expressed as mean ± SD. Abbreviations: Dis: discovery; Val: validation; RF: random forest; Sens: sensitivity; Spec: specificity.

Supplementary Table 9. Multivariate linear regression analyses among biomarkers, sex and log (MDRD eGFR) trained with discovery set and tested with validation set among diabetic patients.

log (MDRD GFR)	Discovery set				Validation set
	β	*p*-value	R^{2*}	RMSE	$R^{2\dagger}$
Model 1			0.9583	0.0765	0.9709
log (MS-detected creatinine)	−1.25 (−1.30 to −1.20)	*p < 0.0001*			
sex	0.16 (0.13 to 0.19)	9.76E-18			
Model 2			0.9531	0.0812	0.9528
log (pseudouridine)	−1.22 (−1.27 to −1.17)	*p < 0.0001*			
sex	0.02 (−0.02 to 0.05)	0.3352			
Model 3			0.9375	0.0938	0.9569
log (L,L-TMAP)	−0.95 (−1.00 to −0.91)	*p < 0.0001*			
sex	0.06 (0.03 to 0.10)	0.0009			
Model 4			0.8908	0.1240	0.9116
log (succinyladenosine)	−0.94 (−1.01 to −0.88)	*p < 0.0001*			
Sex	0.01 (−0.03 to 0.06)	0.5662			
Model 5			0.8690	0.1358	0.9499
log (2-(α-D-mannopyranosyl)-L-tryptophan)	−0.67 (−0.72 to 0.62)	*p < 0.0001*			
sex	0.01 (−0.04 to 0.06)	0.6834			

β, unstandardized coefficient of linear regression. sex, female = 1 and male = 2. Discovery set, *n* = 106; validation set, *n* =56. RMSE, root mean square error. *R^2 was based on the predicted log (MDRD eGFR) against actual log (MDRD eGFR) using the

equation of the model and data of discovery set. [†]R^2 was based on the predicted log (MDRD eGFR) against actual log (MDRD eGFR) using the equation of the model of discovery set and data of validation set.

Supplementary Table 10. Description of 106 follow-up information.

Supplementary Table 11. Analysis of 106 follow-up patients by random forest.

ABCD group		AB group		CD group	
Index	**Mean AUC and Quartile of 100 Iterations**	**Index**	**Mean AUC and Quartile of 100 Iterations**	**Index**	**Mean AUC and Quartile of 100 Iterations**
Sex	0.5625 (0.5234–0.625)	Sex	0.6429 (0.5–0.7143)	Sex	0.5556 (0.5–0.6667)
Age	0.7031 (0.6445–0.7822)	Age	0.7245 (0.6327–0.8367)	Age	0.7407 (0.6327–0.8302)
eGFR	0.7822 (0.7031–0.8359)	eGFR	0.8367 (0.6939–0.9184)	SAdo	0.7809 (0.6975–0.858)
Pseu	0.8125 (0.7397–0.8867)	SAdo	0.7296 (0.6607–0.8393)	eGFR	0.7963 (0.6914–0.8704)
SAdo	0.8174 (0.748–0.8716)	Pseu	0.8622 (0.7245–0.9388)	Cr	0.8086 (0.7207–0.8781)
Cr	0.8438 (0.7925–0.8926)	UACR	0.898 (0.8367–0.9796)	ADT	0.8117 (0.713–0.9012)
ADT	0.8691 (0.8179–0.918)	Cr	0.9133 (0.8265–0.9796)	Pseu	0.8302 (0.7346–0.909)
UACR	0.8887 (0.8262–0.9185)	ADT	0.9184 (0.8673–0.9923)	UACR	0.8889 (0.7901–0.9213)
ADT+UACR	0.9443 (0.9141–0.9727)	ADT+UACR	0.9796 (0.9388–1)	SAdo+UACR	0.9012 (0.8194–0.9506)
UACR+ADT +SAdo+Cr	0.9482 (0.9248–0.9805)	ADT+SAdo+UACR	0.9796 (0.9388–1)	ADT+UACR	0.9136 (0.8395–0.963)
ADT+SAdo+ UACR	0.9502 (0.9062–0.9805)	UACR+ADT+Cr	1 (0.9592–1)	UACR+Pseu+ Cr	0.929 (0.8765–0.9753)
		UACR+ADT+Age+Sex	1 (0.9592–1)	UACR+Pseu	0.9352 (0.8765–0.9753)

Pseu:pseudouridine.

Supplementary Table 12. Pearson correlation of MDRD GFR with three another GFRs that were calculated by the three newly reported equations.

GFR calculation methods	Equations	Pearson correlation coefficients with MDRD GFR
MDRD Study equation[1]	Estimated GFR = 186 × (serum creatinine)$^{-1.154}$ × (age in years)$^{-0.203}$ × 0.742 (if female) × 1.210 (if African American)	–
CKD-EPI creatinine equation[2]	Estimated GFR = 141 × min (serum creatinine/κ, 1)α × max (serum creatinine/κ, 1)$^{-1.209}$ × 0.993$^{Age in year}$ × 1.018 (if female) × 1.159 (if black), where κ is 0.7 for females and 0.9 for males, α is −0.329 for females and −0.411 for males. min indicates the minimum of ratio of serum creatinine to κ or 1, and max indicates the maximum of ratio of serum creatinine to κ or 1.	Discovery (n = 128): 0.9523 Validation (n = 66): 0.9729
CKD-EPI cystatin C equation[2]	Estimated GFR = 133 × min(serum cystatin C/0.8, 1)$^{-0.499}$ × max(serum cystatin C/0.8, 1)$^{-1.328}$ × 0.996^{Age} × 0.932 (if female), where min indicates the minimum of ratio of serum cystatin C to 0.8 or 1, and max indicates the maximum of ratio of serum cystatin C to 0.8 or 1.	Validation (n = 58): 0.9468
CKD-EPI creatinine– cystatin C equation[2]	Estimated GFR = 135 × min (serum creatinine/κ, 1)$^{\alpha}$ × max (serum creatinine/κ, 1)$^{-0.601}$ × min (serum cystatin C/0.8, 1)$^{-0.375}$ × max(serum cystatin C/0.8, 1)$^{-0.711}$ × 0.995^{Age} × 0.969 (if female) × 1.08 (if black), where κ is 0.7 for females and 0.9 for males, and α is −0.207 for males and −0.248 for females.	Validation (n = 58): 0.9681

Supplementary Table 13. List of ID in online database for networking and pathway analysis.

Supplementary Table 14. Gradient diluted and corresponding concentrations of standards.

	100% (ng/ml)	90% (ng/ml)	80% (ng/ml)	60% (ng/ml)	50% (ng/ml)	40% (ng/ml)	20% (ng/ml)	10% (ng/ml)	8% (ng/ml)	6% (ng/ml)
2-(A-D-Mannopyranosyl-)L-Tryptophan	50.00	45.00	40.00	30.00	25.00	20.00	10.00	5.00	4.00	3.00
Succinyladenosine	58.00	52.20	46.40	34.80	29.00	23.20	11.60	5.80	4.64	3.48
pseudouridine	160.00	144.00	128.00	96.00	80.00	64.00	32.00	16.00	12.80	9.60

Supplementary Table 15. Recovery rate and precision of selected metabolites.

Metabolite	Recovery rate
2-(A-D-Mannopyranosyl-)L-Tryptophan	83%
Succinyladenosine	95%
pseudouridine	94%
Tryptophan-D5	95%

Supplementary Table 16. Q1/Q3 mass and MRM conditions for the selected metabolites.

Metabolite	Q1 mass (Da)	Q3 mass (Da)	Dwelling time (msec)	Declustering Potential (volts)	Collision energy (volts)
2-(A-D-Mannopyranosyl-)L-Tryptophan	367.1	247.1	100	152.15	19.3
Succinyladenosine	384.1	252	100	122.58	27.09
pseudouridine	245.1	191	100	108.61	19.97
Tryptophan-D5	210	192	100	61.85	16.69