SUPPLEMENTARY INFORMATION

TITLE:

NMR and MS urinary metabolic phenotyping in kidney diseases is fit-for-purpose in the presence of a protease inhibitor

AUTHORSHIP:

Claire L. Boulangé[1†], Ilse M. Rood[2], Joram M. Posma[3], John C. Lindon[1,3], Elaine Holmes[1,3], Jack F.M. Wetzels[2], Jeroen K.J. Deegens[2] , Manuja R. Kaluarachchi[1]*

1. Metabometrix Ltd, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ UK.
2. Department of Nephrology, Radboud University Medical Center, Nijmegen, Geert Grooteplein Zuid 8, PO box 9101, 6500 HB, Nijmegen, The Netherlands.
3. Imperial College London, Division of Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Sir Alexander Fleming Building, South Kensington, London SW7 2AZ UK.

* To whom correspondence should be addressed: Dr Manuja R. Kaluarachchi, email: manuja.kaluarachchi@imperial.ac.uk

† Present address: Phenome Centre, Singapore, Nanyang Technological University, Lee Kong Chian School of Medicine, Experimental Medicine Building, 59 Nanyang Drive, Singapore 636921

MATERIAL AND METHODS:

Protein precipitation: a volume of 400 µl of a urine sample was homogenised with 900 µl of cold deuterated methanol in Eppendorf tubes and stored at -20^{o}C overnight. After centrifugation for 10 minutes at 13,000 g, the supernatants were collected in different Eppendorfs tubes. Protein pellets and supernatants were dried-down in a speed vac overnight. The dried fraction from the supernatant was reconstituted in 400 µl of water, from which 100 µl was used for protein measurement and 300 µl for NMR analysis. Dried protein pellets were re-suspended with 40 µl of water prior to protein measurement.

Protein measurement: Urinary protein was quantified using the Pierce 660 nm protein assay kit (Thermo, Rockford, USA). Protein concentration was calculated by using a standard curve of bovine serum albumin (BSA) at different concentrations (dynamic range 0-20 mg/ml). For each urine sample, protein concentration was assessed in an intact aliquot, in an aliquot diluted with water (1:10) and from a re-suspended-protein pellet (sample concentrated 10 times) to overcome the dynamic range limitations of the assay. A volume of 10 µl of a urine sample or BSA standard was combined and mixed in 150 µl of protein reagent in a well of a 96-well plate. After 10 minutes of incubation (for the colour reaction), the absorption at 660 nm was measured. Each measurement was performed in triplicate on the same 96 well plate.

RESULTS:

Comparison of ^{1}H NMR spectra processed with median fold change (MFC) or total area normalisation

The volume and concentration of urine are not controlled and depend on the individual's hydration status. Therefore, a normalisation procedure is a critical step to partially account

for dilution differences in order to achieve a high quality and reproducible dataset. In the present study, the MFC normalisation and total area normalisation was compared in the ^{1}H NMR dataset before or after PI peak removal (**Supplementary Figure 4**).

When the MFC normalisation was applied before or after PI peak removal, no change in the comparison between iMN and NC or PI – vs PI + samples were observed. (Supplementary Figure 4E, 4F, 4G, 4H), which means that the MFC normalisation method is not affected by the presence of contaminant peaks. On the other hand, when the ^{1}H NMR data were normalised based on total area normalisation before removing the PI peaks, a natural separation between NC and NC+PI and between iMN and iMN+PI samples was found by PCA (Supplementary Figure 4A). Direct comparison of the NMR spectra highlighted an increase of peak area in urine samples containing PI when the spectra were normalised to total area (Supplementary Figure 4B). This common normalisation methods is based on the adjustment of each NMR data point intensity by normalisation to the total data point area in the whole spectrum. . However, when large peaks from proteins, or contaminants such as PI, are present in the urine, this normalisation method introduces intensity artefacts and may increase false positives in biomarker discovery. With this in mind, the total area normalisation of this data set must be achieved after careful removal of PI contaminant peaks (Supplementary Figure 4C, 4D). Alternatively, the MFC normalisation calculated the median of the log fold change of the peak intensities between a specific sample and a target sample and adjust the log fold change to approximately 0. The targeted sample is typically the median spectrum in a set of experiment. This normalisation method is based on the statement that metabolites peaks uniquely affected by dilution confer the same fold change. The calculated median fold changes corresponds to the most probable dilution factor for each ^{1}H NMR spectrum. This method is known to be robust even with the

presence of by high intensity resonances in the spectra. In our study, there is no significant difference in PCA score plot when the data have been MFC normalised either prior to or after contaminant peak removal (Figure S4 E, G), indicating that the method is not affected by the presence of PI peaks. .. In general, the MFC normalisation is the most suitable method to correct for dilution effects in ^{1}H NMR urine profiles of iMN patients, which may contain large peaks from proteins, drugs and their metabolites, as well as additives from sample preparation. The MFC-normalised ^{1}H NMR data after PI peak removal was therefore selected prior to carrying out the statistical analysis.

SUPPLEMENTARY FIGURES AND TABLES LEGENDS:

Table S1: Published list of the chemicals present in the protease inhibitor from Roche.

Table S2: Cross validation statistics of pairwise OPLS-DA models comparing NC and iMN samples with and without PI. The table presents cross validation (Q^2Y), goodness of fit (R^2Y) and CV-ANOVA p-values from cross-validated Y-variable residuals for all metabolomics datasets (^{1}HNMR and RP and HILIC UPLC-MS). OC: orthogonal component; PC: predictive component. NC: normal controls, iMN: idiopathic membranous nephropathy, PI: protease inhibitor.

Figure S1: ^{1}H NMR 1D standard spectrum from 0.3 to 9.5 ppm (A) and from 7.4 to 8.4 ppm (B) of a urine sample from a iMN patient (in purple) without PI and the spectrum of the same sample after storage with PI (in red). ^{1}H NMR 1D standard spectrum from 0.3 to 9.5 ppm of PI roche mini standard. Presence of mannitol, AEBSF derivative and protein background are observed in urine sample from iMM patient with PI and in the spectrum of the PI Roche mini preparation itself.

Figure S2: STOCSY (Statistical Total Correlation SpectroscopY) plot of ^{1}H NMR spectra showing the NMR variables associated with the PI related contaminant peak of the AEBSF derivative at 7.49 ppm (A, C) and mannitol at 3.87 ppm (B, D).

Figure S3: Base peak intensity chromatograms from the HILIC UPLC-MS (ESI+) (A) and RP UPLC-MS (ESI+) (B) analysis of a urine sample from a iMN patient with PI (in red), the same sample without PI(in purple) and PI itself in water (green). Fragmentation mass spectra of AEBSF (C) and leupeptin hemisulfate (D). iMN: idiopathic membranous nephropathy, PI: protease inhibitor.

Figure S4: Effect of total area or median fold change normalisation on ^{1}H NMR urine profiles before or after removal of PI contaminant peaks on group clustering and peak shape. PCA score plots based on NMR spectra from the iMN, iMN+PI, NC and NC+PI groups, using total area normalisation prior to PI peak removal (A) or after PI peak removal (C), or MFC normalised data prior to PI peak removal (E) and after PI peak removal (G). Comparison of the NMR peak shape of hippurate at 7.59 ppm in the spectra of urine from the iMN, iMN+PI, NC and NC+PI groups using total area normalised data prior to PI peak removal (B) or after PI peak removal (D), or MFC normalised data prior to PI peak removal (F) and after PI peak removal (H). NC: normal controls, iMN: idiopathic membranous nephropathy, PI: protease inhibitor

Figure S5: ^{1}H NMR 1D standard spectra of iMN (green), iMN+PI (red), NC (black) and NC+PI (blue) urine samples before (A-C) and after protein precipitation (D-F) . Expansion of NMR spectra at 7.5-7.9 ppm (A,D), at 3.9-4.2 ppm (B,E), and at 2.1-2.42 ppm (C,F). Peak intensities are in arbitrary units.

Figure S6: Urinary proteins measured by a protein assay kit (in μg/ml) before and after protein precipitation in urine from subjects in the NC and NC+PI, (A) iMN and iMN+PI (B). NC: normal controls, iMN: idiopathic membranous nephropathy, PI: protease inhibitor.

Figure S7: RP UPLC-MS (ESI+) features from samples containing PI predominantly relate to the components of the PI. The scatter plot shows position (m/z and retention time) of the 2199 MS features across the RP UPLC MS (ESI+) analysis that influence (WMW p-value<0.05) the discrimination between NC/iMN and NC+PI/iMN+PI. A total of 2145 features are possibly related to the trypsin inhibitor, 11 to leupeptin hemisulfate and 9 to AEBSF (N:

number of features) contained in the PI (A). Comparison of total ion current (TIC) chromatogram of samples containing PI (in red) and without PI (in purple) (B). MS spectrum extracted at a retention time of 6.26 minutes displaying the ions associated with the trypsin inhibitor. The trypsin inhibitor is responsible for the large signal observed in the TIC chromatogram of the sample with PI (C).

SUPPLEMENTARY TABLES:

Table S1: List of the chemicals present in the protease inhibitor from Roche

Chemical in the PI	cas-no	Molecular mass	detected	Detected m/z (species) in MS
Disodium dihydrogen ethylenediaminetetra-acetate (EDTA)	139-33-3	336.21	^{1}H NMR	Not detected
4-(2- Amino-ethyl)benzenesulfonyl fluoride hydrochloride	30827-99-7	239.69	HILIC(ESI+) RP(ESI+) ^{1}H NMR	HILIC and RP: 204.050 $[M+H^+]^+$ (with the loss of HCL) RP: 228.050 $[M+Na^+]^+$ RP: 187.023 $[M-NH_2+2H^+]^+$
Leupeptin hemisulfate	103476-89-7	475.75	RP(ESI+) HILIC(ESI+/-):	RP/HILIC: 427.303 $[M+H^+]^+$ RP: 499.285 $[M+Na^+]^+$ RP: 409.293 $[M-OH^-]^+$ RP:389.083 $[M-C_2H_2OH^-]^+$
Trypsin inhibitor, pancreatic basic	9087-70-1	6511.44	RP(ESI+)	Multiple ions detected in the RP data at regular intervals, typical of a peptide

Table S2: Cross validation statistics of pairwise OPLS-DA models comparing different NC and iMN samples with and without PI.

dataset	statistics	NC vs NC+PI	iMN vs iMN+PI+	NC/iMN vs NC+PI/iMN+PI	NC vs iMN	NC+PI vs iMN+ PI	NC/NC+PI vs iMN /iMN+PI
^{1}HNMR data (1D standard experiment)							
Original matrix (without protein precipitation)	Q^2Y	0.73	-0.04	0.44	0.88	0.87	0.79
	R^2Y	0.99	0.73	0.8	0.95	0.94	0.88
	CV-ANOVA *p* values	0.12	1	0.05	0.0006	0.0007	5.2×10^{-5}
	PC+OC	1+1	1+1	1+1	1+0	1+0	1+1
Matrix after protein precipitation	Q^2Y	0.59	-0.33	0.25	0.39	0.37	0.79
	R^2Y	0.99	0.89	0.90	0.91	0.92	0.99
	CV-ANOVA *p* values	0.27	1	0.32	0.17	0.19	5.1×10^{-5}
	PC+OC	1+1	1+1	1+1	1+0	1+0	1+1
UPLC-MS data							
HILIC (ESI+)	Q^2Y	0.39	-0.74	-0.20	0.76	0.77	0.81
	R^2Y	0.93	0.84	0.85	0.98	0.98	0.99
	CV-ANOVA *p* values	0.56	1	1	0.024	0.019	0.045
	PC+OC	1+1	1+1	1+1	1+1	1+1	1+1
HILIC (ESI-)	Q^2Y	0	0	-0.27	0.71	0.76	0.91
	R^2Y	0.	0	0.83	0.98	0.99	0.99
	CV-ANOVA *p* values	0.57	0.93	1	0.045	0.024	7.4×10^{-10}

	PC+OC	0	0	1+1	1+1	1+1	1+1
RP (ESI+)	Q^2Y	0.96	0.44	0.78	0.30	0.37	0.59
	R^2Y	0.99	0.82	0.87	0.98	0.95	0.92
	CV-ANOVA *p* values	0.0017	0.32	7.61×10^{-5}	0.70	0.59	0.0072
	PC+OC	1+1	1+1	1+1	1+1	1+1	1+1
RP (ESI+exc)	Q^2Y	0.78	0.11	0.51	0.42	0.40	0.76
	R^2Y	0.99	0.85	0.87	0.98	0.98	0.96
	CV-ANOVA *p* values	0.082	0.922	0.022	0.518	0.565	1.659×10^{-4}
	PC+OC	1+1	1+1	1+1	1+1	1+1	1+1
RP (ESI-)	Q^2Y	0.29	-0.67	-0.25	0.50	0.47	0.86
	R^2Y	0.99	0.87	0.86	0.98	0.98	0.98
	CV-ANOVA *p* values	0.66	1	1	0.39	0.43	2.91×10^{-6}
	PC+OC	1+1	1+1	1+1	1+1	1+1	1+1

(Q^2Y), goodness of fit (R^2Y) and CV-ANOVA p-values from cross-validated Y-variable residuals for all metabonomics datasets (^{1}H NMR and RP and HILIC UPLC-MS). OC: orthogonal component; PC: predictive component. NC: normal controls, iMN: idiopathic membranous nephropathy, PI: protease inhibitor.

SUPPLEMENTARY FIGURES:

Figure S1

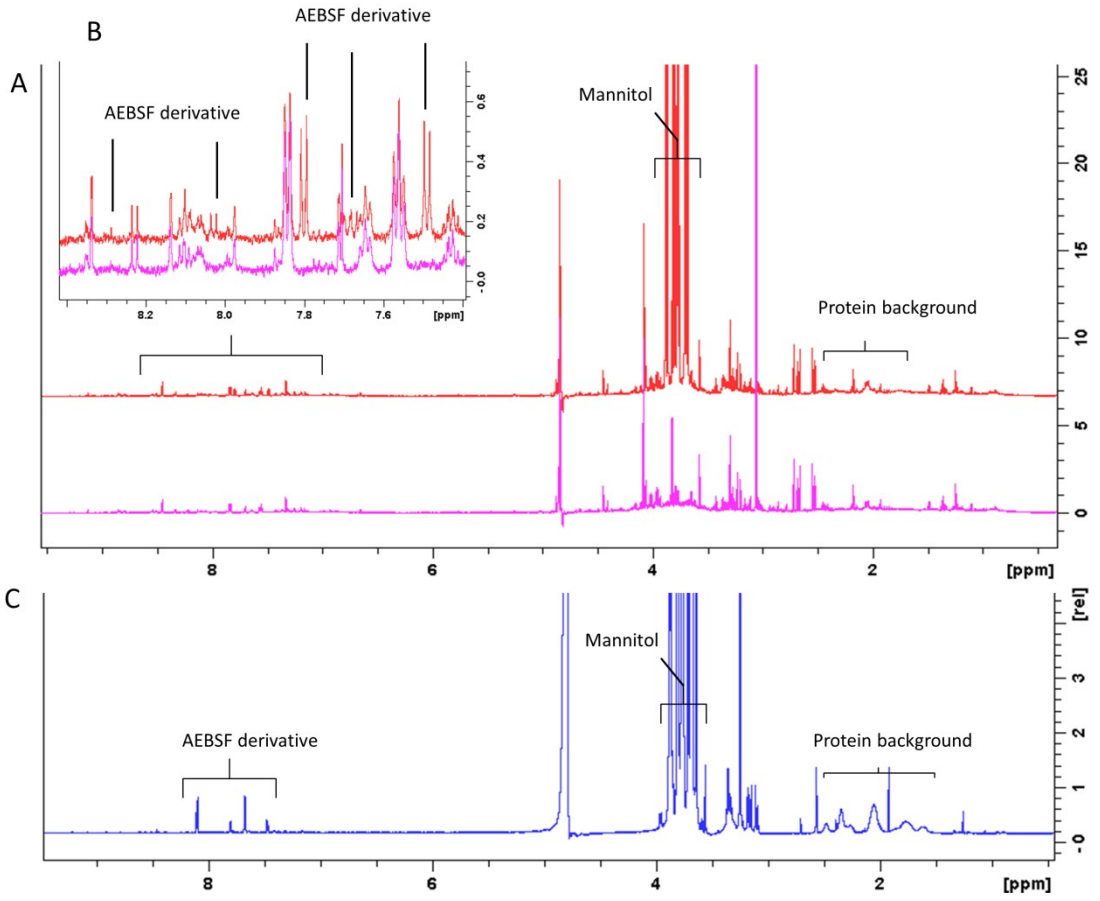


Figure S2:

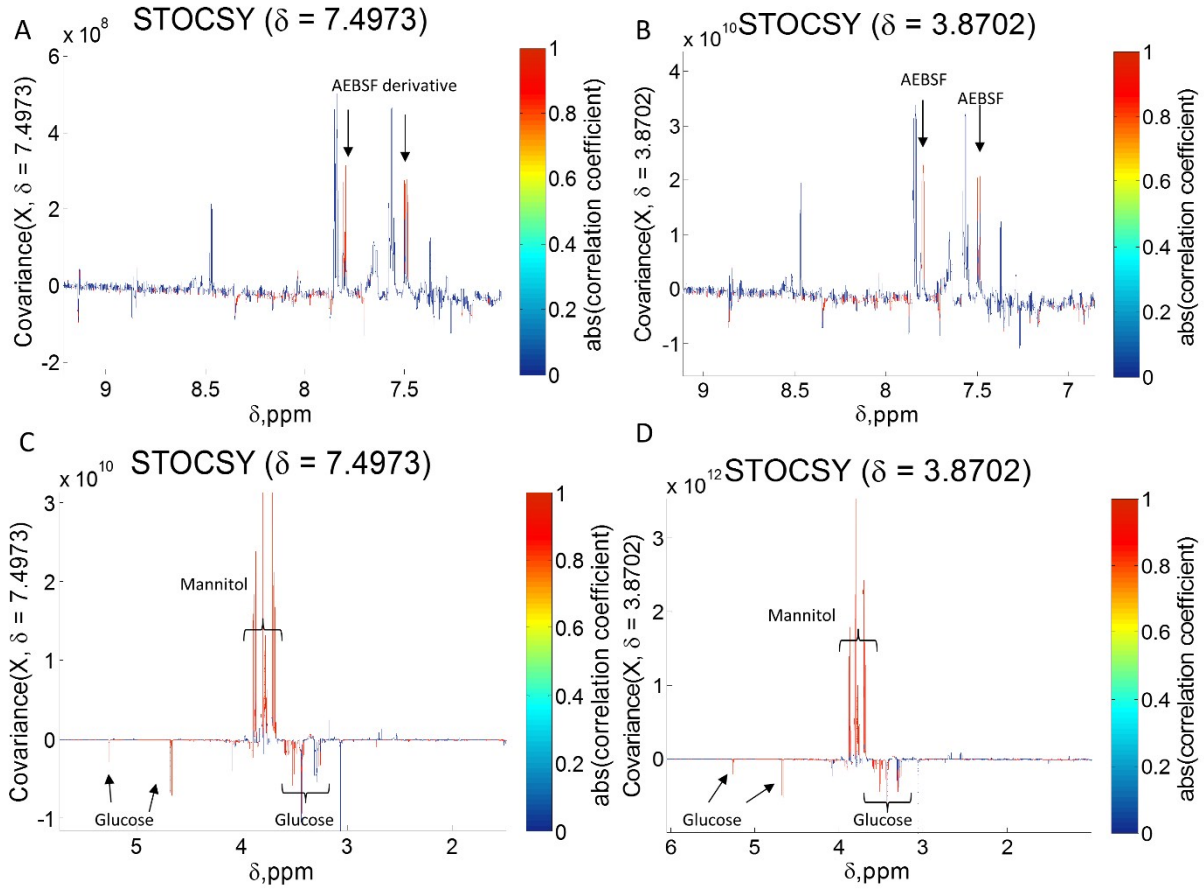

A
STOCSY (δ = 7.4973)
x 10^8
AEBSF derivative
Covariance(X, δ = 7.4973)
δ,ppm
abs(correlation coefficient)
B
STOCSY (δ = 3.8702)
x 10^10
AEBSF
AEBSF
Covariance(X, δ = 3.8702)
δ,ppm
abs(correlation coefficient)
C
STOCSY (δ = 7.4973)
x 10^10
Mannitol
Glucose
Glucose
Covariance(X, δ = 7.4973)
δ,ppm
abs(correlation coefficient)
D
STOCSY (δ = 3.8702)
x 10^12
Mannitol
Glucose
Glucose
Covariance(X, δ = 3.8702)
δ,ppm
abs(correlation coefficient)

Figure S3:

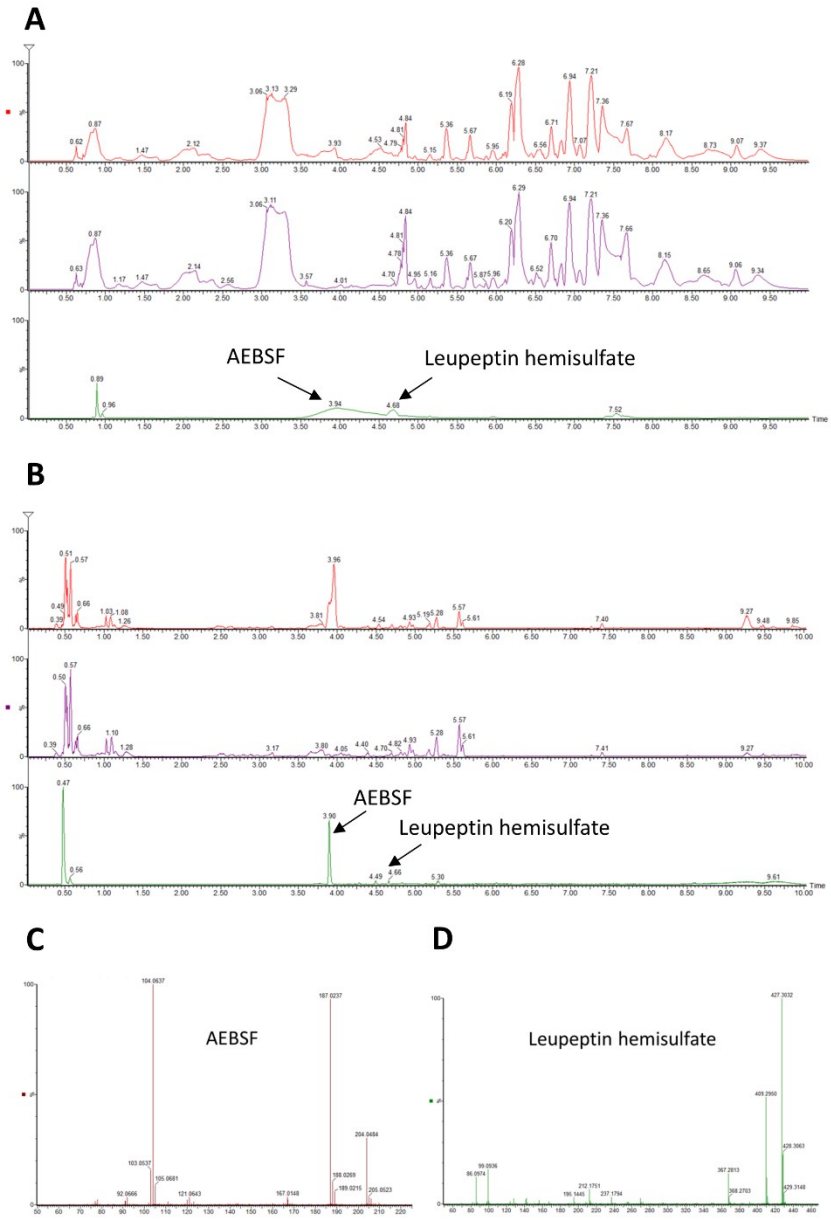

A
AEBSF
Leupeptin hemisulfate
B
AEBSF
Leupeptin hemisulfate
C
AEBSF
D
Leupeptin hemisulfate

Figure S4:

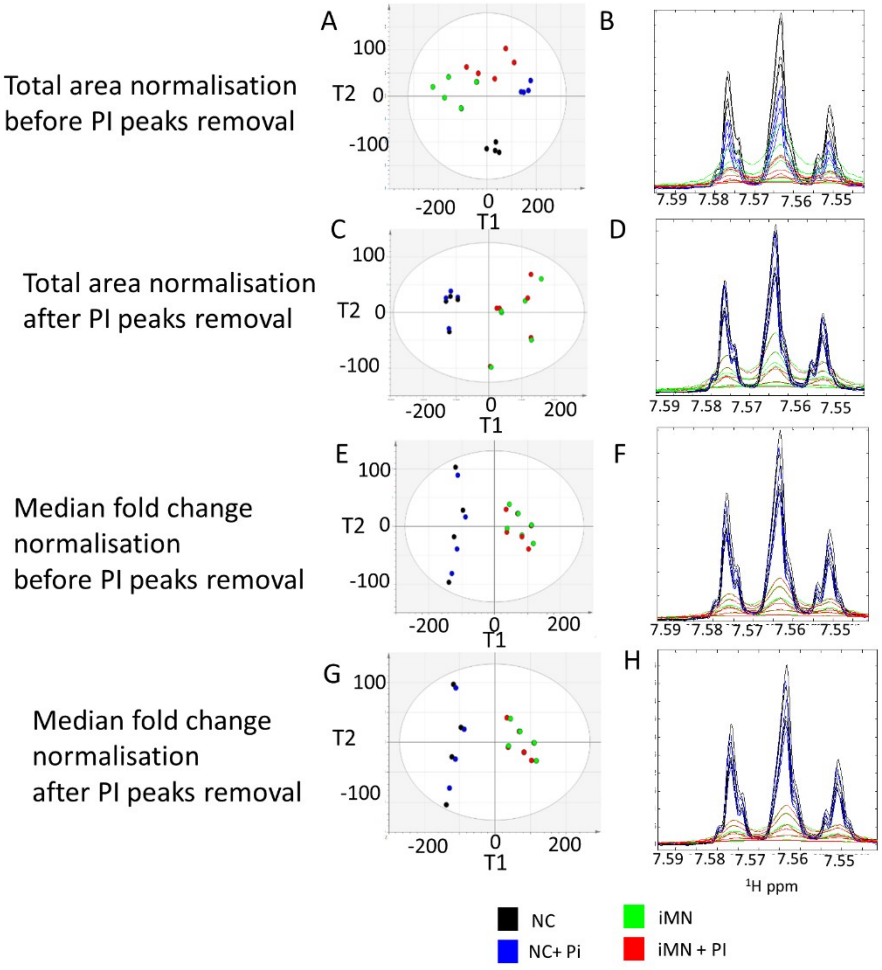

Total area normalisation
before PI peaks removal
Total area normalisation
after PI peaks removal
Median fold change
normalisation
before PI peaks removal
Median fold change
normalisation
after PI peaks removal
A
B
C
D
E
F
G
H
T1
T2
100
0
-100
-200
200
7.59 7.58 7.57 7.56 7.55
¹H ppm
NC
NC+ Pi
iMN
iMN + PI

Figure S5:

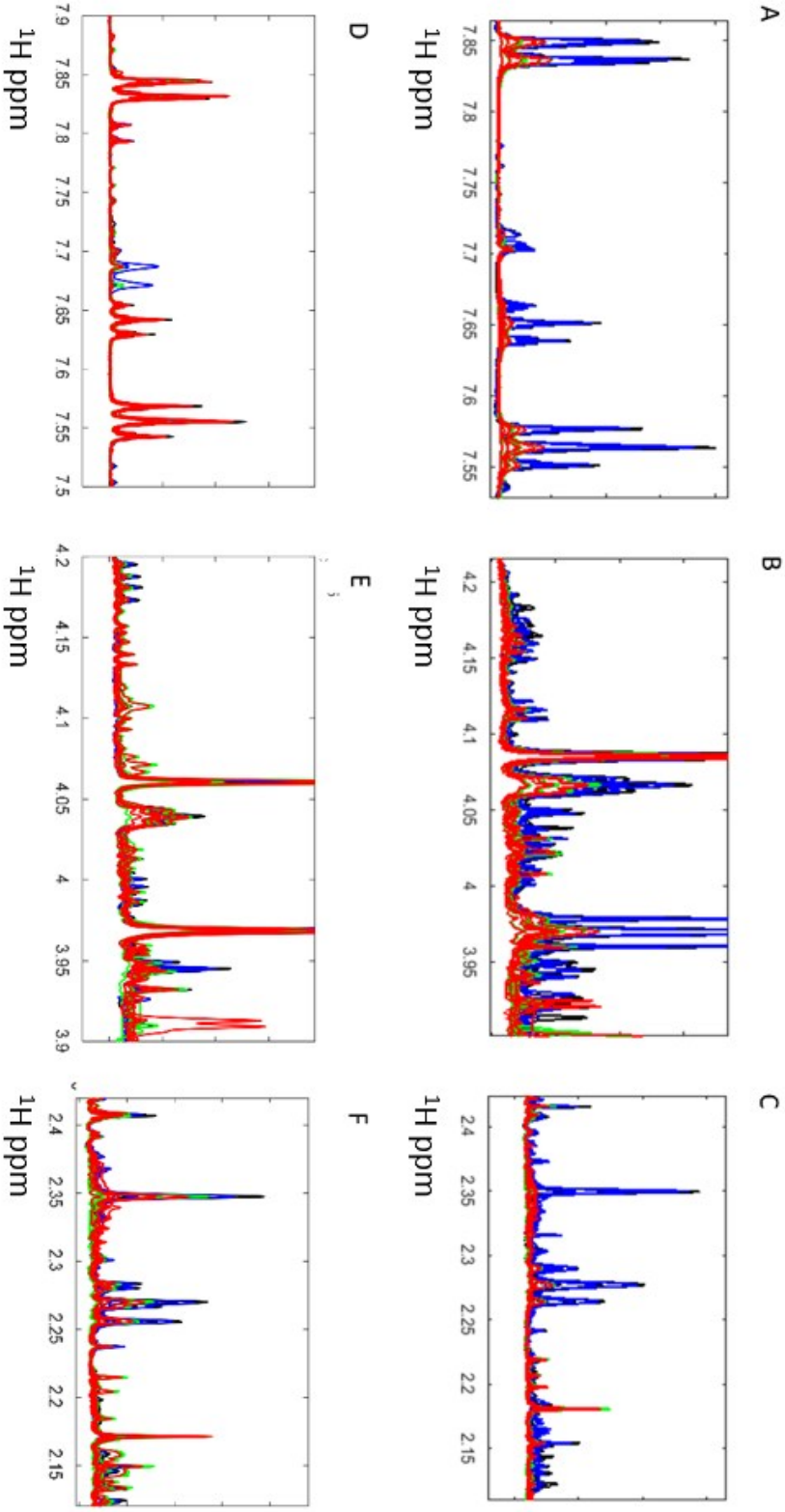


Figure S6:

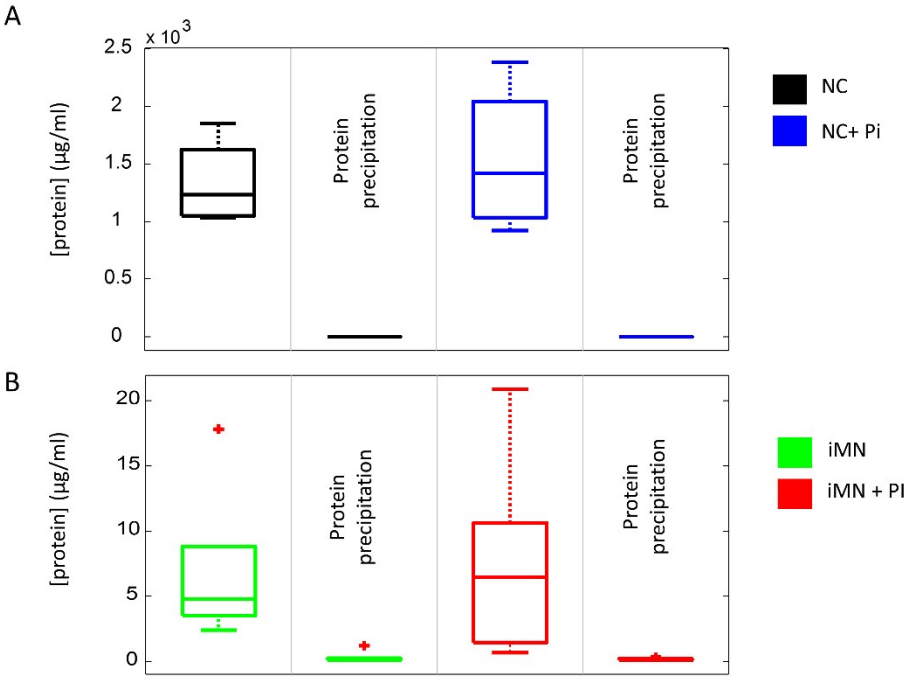
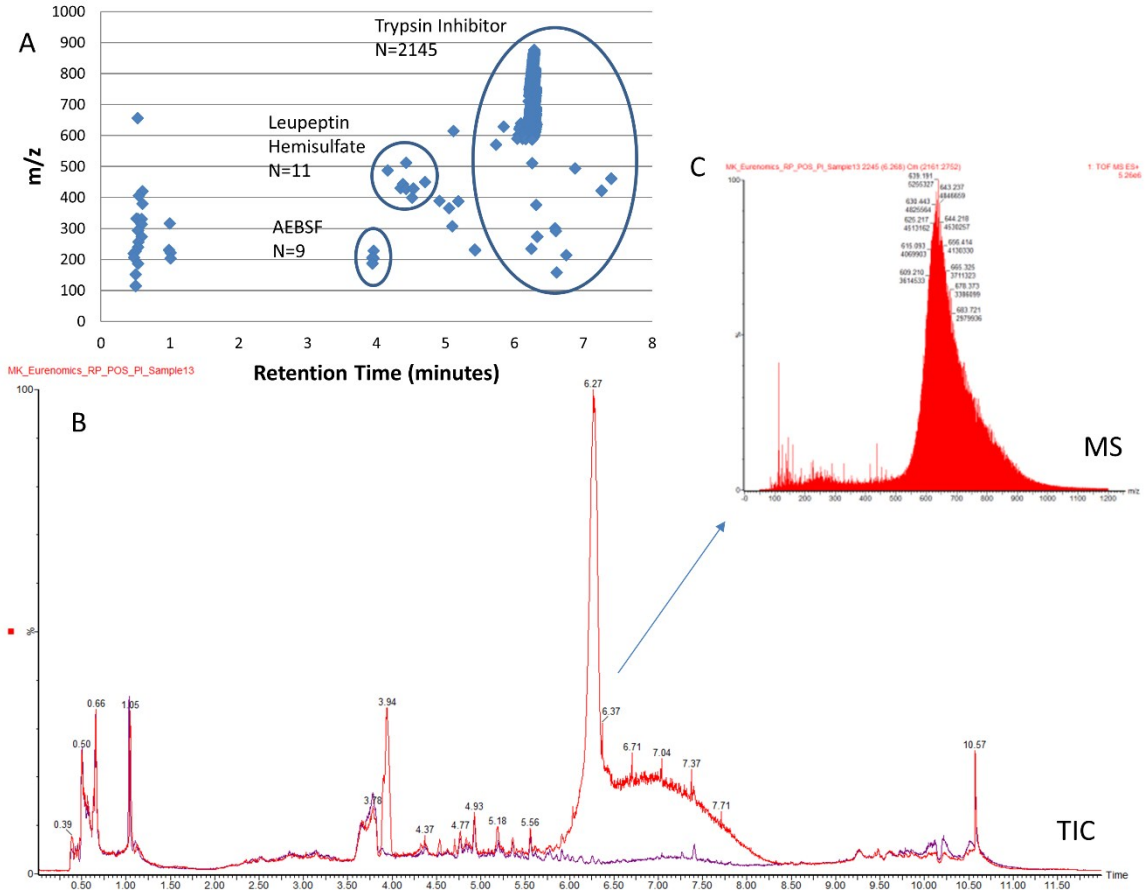

A
x 10^3
2.5
2
1.5
1
0.5
0
[protein] (µg/ml)
Protein precipitation
Protein precipitation
NC
NC+ Pi
B
20
15
10
5
0
[protein] (µg/ml)
Protein precipitation
Protein precipitation
iMN
iMN + PI

Figure S7:


A
Trypsin Inhibitor
N=2145
Leupeptin
Hemisulfate
N=11
AEBSF
N=9
m/z
Retention Time (minutes)
C
MS
B
TIC
Time