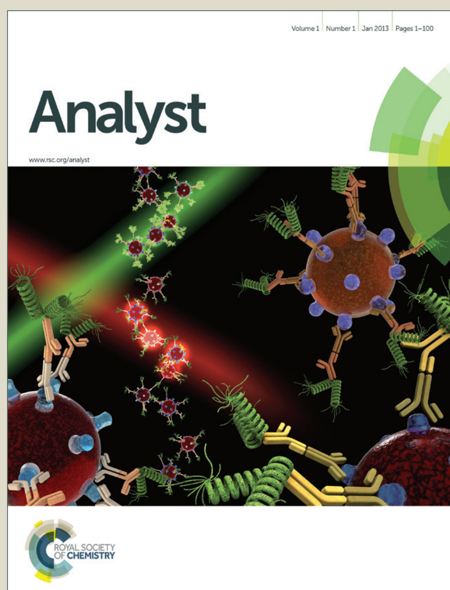
Analyst

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/analyst

A short review of applications of liquid chromatography mass spectrometry based metabolomics techniques to the analysis of human urine.

Tong Zhang and *David G. Watson

Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral Street, Glasgow, UK, G4 0RE.

*Corresponding author:

D. G. Watson

161 Cathedral Street,

Glasgow G4 0RE

Telephone: +441415482651

Fax: +441415526443

Email: d.g.watson@strath.ac.uk

Analyst Accepted Manuscript

Abstract

The applications of metabolomics as a methodology for providing better treatment and understanding human disease continue to expand rapidly. In this review, covering the last two years, the focus is on liquid chromatography-mass spectrometry (LC-MS) profiling of metabolites in urine. In LC-MS based metabolomics there are still problems with regard to: chromatographic separation, peak picking and alignment, metabolite identification, metabolite coverage, instrument sensitivity and data interpretation and in the case of urine sample normalisation. Progress has been made with regard to all of these issues during the period of the review. Of particular interest are the increasing use of orthogonal chromatographic methods for optimal metabolite coverage and the increasing adoption of receiver operator characteristic (ROC) curves for biomarker validation.

Introduction

The numbers of publications involving global profiling of low MW (<1000 amu) metabolites are increasing rapidly. In relation to human beings such studies focus on: diagnosis of disease, understanding disease, risk stratification, personalised medicine, monitoring the success of disease treatments and drug discovery. In human studies urine is a convenient bio-fluid for metabolomics studies for several reasons: it can be collected non-invasively; it is available in large quantities; collected continuously over a period it provides a complete metabolic profile unlike blood which provides a snapshot; it shows a different metabolome from blood; sample handling is simple since there is no need to remove protein. A major drawback of is the difficulty in standardising urinary metabolite concentrations which might vary significantly due to varying personal hydration status. In the past urine has been standardised to creatinine but this may not be completely reliable[1].

A comprehensive list of the metabolites in urine has been compiled. By using several analytical platforms it was possible to identify 449 and quantify 378 metabolites in urine [2]. The platforms detected (quantified) metabolites as follows: 209 (209) NMR, 179 (85) by GC-MS, 127 (127) by DFI/LC-MS/MS, 40 (40) by ICP-MS and 10 (10) by HPLC. Several previously unknown urinary metabolites were identified. A literature review led to the identification and annotation of another 2206 metabolites (http://www.urinemetabolome.ca). The methods used were complementary but for coverage, high

Analyst Accepted Manuscript

throughput and ease of use NMR was rated the highest. Within the two years covered by this review, there have been eleven reviews of the topic [3-13]. The key issues in LC-MS based metabolomics studies remain: instrument sensitivity and selectivity; chromatographic selectivity [14]; optimal data extraction which includes alignment and picking of real peaks, metabolite identification [15] and data interpretation. Two terms are used frequently in metabolomics reports. The term feature is used to describe a clear chromatographic peak to which an identity cannot be always assigned. A feature is not necessarily a metabolite but can be due to an adduct, fragment or isotope peak relating to another metabolite. The other term is putative identification or annotation which corresponds to MSI level 2 identification[16]. Most of the literature reviewed below utilised high resolution mass spectrometry and within this frame of reference putative identification or annotation indicates that a metabolite has an exact mass linked to a defined elemental composition that also matches a metabolite in a data base. In order to achieve MSI level 1 identification of a metabolite, an additional orthogonal method should be used such as comparison of its chromatographic retention time with that of an authentic standard or comparison of the MS/MS fragmentation pattern of the metabolite with that of an authentic standard or against a spectral database[16]. Table 1 summarises studies over the last two years which have carried out metabolomic profiling of urine in order to find biomarkers of health or disease[17-65] . As can be seen from the table there is no standard approach and sometimes complete details of the methodology are not given. Some key references are discussed further below.

Chromatography Methods

Chromatographic separation is important even when high resolution mass spectrometry detection, which gives a defined elemental composition, is used since isomers have to be separated in order to achieve definitive identification of a unique metabolite. The majority of studies shown in table 1 have utilised reversed phase chromatography (RPC) alone. For best metabolite coverage orthogonal methods, which use hydrophilic interaction chromatography (HILIC) and RPC, are required. Thus far only a limited range of HILIC columns are available in small particle UPLC format but in chromatography selectivity is much more significant than efficiency[14]. The different chromatographic methods used in metabolomics have been reviewed [14]. Zhang et al analysed human urine plus 173 standards on a C18 column, a ZICHILIC column, a

Analyst Accepted Manuscript

ZICpHILIC column and a Cogent Diamond Hydride column. The numbers of putatively identified metabolites in urine were as follows: C18 column 564, ZICpHILIC column 789, ZICHILIC column 824 and Cogent column 570 [66]. The ZICpHILIC column gave the best overall coverage; the C18 column produced the lowest number (8%) of unique metabolites detected by that method alone. Roux et al analysed urine on a C18 UPLC column or a pentafluorophenyl (PFP) column. Most of the polar compounds in urine eluted at or close to the void volume of the C18 column whereas the PFP column was able to separate polar isomeric compounds. Putative identification of 384 metabolites was made and 192 were matched against reference compounds [67]. Kloos et al also made the similar observations when comparing orthogonal methods[68] . Many compounds, apart from dietary xenobiotics and their metabolites, in urine are very polar and thus will not retain strongly on a C18 column. Early eluting compounds are more likely to be subject to ion suppression which compromises quantitative estimation and in addition isomers of polar compounds are not separated on RP columns.

Optimisation of Mass Spectrometer Performance

The mass spectrometers used in the various studies are summarised in table 1. The majority of studies utilised time of flight (TOF) instruments, some used Orbitrap Fourier transform instruments and a few used targeted tandem MS/MS methods. A quantitative targeted metabolomics approach monitoring 134 urinary metabolites was used to study T-cell mediated rejection (TCMR) after kidney transplantation and the method was as effective in assessing the risk of rejection as an invasive microarray analysis [41]. Targeted methods usually follow from biomarker discovery methods based on high resolution methods using TOF or Orbitrap analysis. In these cases the primary identification of metabolites relies on an accurate masses being obtained and matched against a database such as the Human Metabolome Data Base (HMDB)[69] for corresponding masses. In human metabolomics studies if the mass deviation of the proposed mass is < 3 ppm from that of a known, non-xenobiotic, compound it is likely that, based on the elements C,H,N,O,S,P, the only competing metabolite ids produced will result from isomers[15]. Older TOF instruments tend to produce poorer mass accuracy than that of Orbitraps of a similar age. For instance a recent study reported between laboratory mass accuracy of ±12 ppm for a variety UPLC TOF platforms[70] . Juo et al used sodium formate cluster ions post-acquisition to improve retrospectively to improve mass accuracy on a TOF

Analyst Accepted Manuscript

instrument so that it routinely had a deviation of < 4ppm [71]. The more recently released TOF systems can produce the same mass accuracy as the older generation Orbitraps[72] which routinely work to sub 3 ppm accuracy.

Instrument Sensitivity and Metabolite Coverage

By far the most commonly used ionisation technique in LC-MS is ESI. However, there are many metabolites which are either not ionised under ESI conditions or have poor ionisation efficiencies and thus are not readily detected. In addition many compounds such as hormones are present at low levels and thus are difficult to detect with untargeted methods. It was found that post column infusion of 2-(2-methoxyethoxy) ethanol improved ESI–MS sensitivity in negative ion mode up to 60 fold [73]. Analyte response in negative ion mode was investigated and it was concluded that the best additive for sensitivity was 1 mM acetic acid [74]. A series of papers have looked at stable isotope tagging to improve method precision and sensitivity. The general approach is to tag a pooled sample of urine with a stable isotope labelled tag and then match individual samples tagged with unlabelled tag against the pooled standard thus giving improved precision. In addition, selective derivatisation can be used to assist in characterising new metabolites. Liu et al used a stable isotope labelled tag in combination with precursor ion scan to identify or putatively identify 103 thiols in urine [75]. Several other papers have utilised the peak pairs generated from tagging with labelled and unlabelled reagent to improve sensitivity, via improved ionisation efficiency, and improved specificity, which is conferred by the co-elution of labelled and unlabelled metabolites [76,77] [78]. A software tool was developed for the analysis of biological samples tagged with unlabelled and labelled dansyl chloride (www.mycompoundid.org/IsoMS)[79]

Feature Selection

Features are extracted from high resolution mass spectrometric data with a certain mass and retention time width e.g. ±10 ppm and ±0.5min. This generates extracted ion chromatograms containing peaks which can be linked to the accurate mass responsible for generating the peak. The initial problem the software has to deal with is to align chromatographic peaks efficiently since there are always slight variations in retention time from run to run and this can result in some peaks with the same exact mass and very close retention times being collected as different features. This can compromise the statistics applied post-

Analyst Accepted Manuscript

extraction and result in, for instance, missing values. Table 1 summarises the software packages used for data extraction in the studies covered by this review. Many of these are provided by vendors but there are popular packages such as XCMS[80] and MZMine[81] which are freely available. Chen et al compared three popular software packages for their effectiveness in peak finding, filtering, alignment, de-noising, and normalization. None of the data extraction packages picked exactly the same set of marker compounds and thus they concluded data pre-processing should be carried out with more than one package for optimal results [82]. Zheng et al defined a reliability index to reflect the linear correlation between metabolite concentration and peak area and found that it was improved 9.5 times for a standard mixture and 14.5 times for human urine data with the optimised parameter settings in XCMS being calculated by experimental design approaches[83]. There is no perfect data extraction software as evidenced by the continual refinement of packages by vendors. It is important to roughly check the raw data, particularly when key marker compounds are being proposed, avoiding the false positive outcomes generated by poor chromatographic signals. Some examples of how inappropriate data extraction can occur as a result of poor chromatography have recently been presented[84]. Optimal setting of MS acquisition parameters can have a bearing on the performance of the data extraction software. Mattarucchi et al found that feature selection was improved on an Acquity UPLC system coupled to a Q-TOF Premier mass spectrometer (Waters) as follows: metabolic profiles were acquired in triplicate with a scan time of 0.6 s and the samples were run in random order, which is standard practise. The application of this strategy reduced the number missing values in data sets [85] A procedure using pseudocolor plots was used analysis of LC/MS data providing an alternative approach to traditional untargeted metabolomics workflow and eliminating alignment and pre-processing of spectra [86].

Data Normalisation

The variation in the biological sample is much greater than instrumental variation although this may become significant over very long runs and is compensated for by randomising sample order and using pooled samples to correct data [3,9,12, 23, 87]. Urine strength varies and several parameters have been used to normalise it including creatinine concentration, osmolality and specific gravity. Creatinine normalisation has traditionally been the most popular but it is not always reliable because many exogenous factors (e.g.

Analyst Accepted Manuscript

1 2 150 disease states) could affect the excretion of creatinine[1]. As shown in table 1 nearly 50% of the studies do
3 4 151 not state that the data were normalised and only ten used creatinine and in most cases details of the
5 6 152 creatinine normalisation process were not given. Creatinine can be determined from its mass spectrometry
7 8 153 signal but one has to be careful that its signal is not outside the linear range of the instrument since it is
9 10 154 much more abundant than the majority of compounds in urine. Thus getting an accurate measure of it
11 12 155 might entail doubling up sample runs so that a more concentrated sample is run for maximum coverage
13 14 156 and a diluted sample is run for creatinine determination. An alternative is to use the spectrophotometric
15 16 157 methods which have been used for years in clinical practice but are not necessarily completely specific for
17 18 158 creatinine. Another popular method for sample normalisation is to use MS Total Useful Signals (MSTUS)
19 20 159 and this was used by nearly half the studies shown in table 1. This technique proposed by Warrack et al[88]
21 22 160 carries out subtraction of all the background signals in the mass spectrometer present in blank runs and
23 24 161 retains the useful signals associated with the samples. However, there does seem to be a problem with this
25 26 162 in that abundant metabolites such as creatinine tend to fluctuate throughout the day. Even the use of 24
27 28 163 hour samples is not fool proof because levels creatinine, metabolites in the purine pathway[89],
29 30 164 acylcarnitines[90] and many others[91] depend on the level of physical activity day to day. Thus for instance a
31 32 165 relatively active control group and a more sedentary patient group could have quite different profiles
33 34 166 resulting from different levels of physical activity. Application of OPLS-DA to urine samples from prostate
35 36 167 cancer patients and controls gave the best discrimination when MSTUS or creatinine were used as
37 38 168 normalisation methods compared to osmolality or no normalisation[23]. A new normalisation strategy was
39 40 169 demonstrated by Wu et al. where the total concentration of metabolites was determined by a LC-UV
41 42 43 170 measurement of each urine sample following dansylation labelling and subsequently was utilised to adjust
44 45 171 the sample loading amount in a LC-HRMS analysis for standardisation of urinary metabolite
46 47 172 concentrations[92]. In a comparative study of different normalisation methods OPLS-DA was applied to reveal
48 49 173 the discriminant MS features between cohorts with high and low/zero dietary intake of certain food.
50 51 174 Normalisation to specific gravity prior to LC-HRMS analysis showed the most significant improvement in the
52 53 175 number of total discriminant MS features recovered in comparison to that obtained without
54 55 176 normalisation[93].
56 57 58 59 60

Analyst Accepted Manuscript

Statistical Methods

As detailed in table 1 a range different statistical methods are used in modelling of data sets and good coverage of these methods has been provided by several recent reviews[94-97]. The simplest statistics used to discriminate control and affected groups is based on univariate statistics with determination of difference based on a t test, a Wilcoxon rank test or ANOVA. Unsupervised multivariate methods such as principle components analysis (PCA) enable classification of samples sets which, in the case of metabolomics studies, may be based on data reduction for a large number of variables into a few key components . Supervised multivariate methods such as PLS-DA and OPLS-DA have been widely employed in biomarker discovery study. It is important to bear in mind that these methods can suffer from over-fitting caused by allowing too many variables relative to the number of observations. . Therefore an internal (a subset from within the sample set used for generating the model) or external (using a new sample set) cross-validation (CV) should be carried out in order to evaluate the model reliability. Such models provide a useful method for targeting significant biomarkers and once a set of candidate biomarkers has been compiled it is preferable to construct a ROC curve. ROC curves, which have been widely used in clinical chemistry, are generated by using a non-parametric statistical method and provide a method for checking the sensitivity and specificity of a set of biomarkers for diagnostic applications. ROC curves are increasingly being used to check the robustness of biomarkers discovered in metabolomics studies [17-19,22,23,31,38-40,55.] and their applications have been recently addressed in a tutorial paper[96]. Problems with the application of univariate statistics to metabolomics data sets arise from multiple sampling. Thus when comparing data sets for treatments and controls, which contain hundreds metabolites, it is probable that some metabolites will be significantly different by chance. Correction for the possibility of false positives using the Bonferroni correction or the Benjamin-Hochberg false discovery rate is not widely carried out as judged from the papers we have reviewed. This may in part be because it is difficult to define the number of variables which are being employed from the thousands of metabolites available in theory. If several biochemically related metabolites are altered between a treatment and its control then this can improve confidence in the robustness of an observation. For instance several metabolites in one pathway might be up or down regulated but in many studies this is not the case. Zhang et al applied a range of chemometric methods to

Analyst Accepted Manuscript

204 LCMS data in order find biomarkers for patients with hepatitis C virus and map them onto Kegg pathways
205 but on close examination of this paper there is no strong association with a particular pathway[42]. Perhaps
206 by far the best way to avoid false positives is to repeat the experiment using the same sample size drawn
207 from the population of interest[96].
208

209 **Metabolite Identification**

210 LC-MS methods can be classified as either targeted or untargeted. Targeted methods are generally based
211 on tandem mass spectrometry using low resolution trap instruments or triple quadrupole instruments.
212 Targeted methods use a set of authentic standards to standardise both chromatographic retention times
213 and the fragments produced in the collision cell of the mass spectrometer. One can consider the
214 compounds being monitored by these methods to be characterised to MSI level 1[16]. Targeted methods are
215 usually also designed to produce quantitative data. The work flow in high resolution mass spectrometry
216 methods based on HRMS (e.g. TOF or Orbitrap) methods is more complex. Using such methods it is more
217 likely that potential biomarkers will be observed for which authentic standards are not available or which
218 are unknown compounds. The first pass in global metabolomics screens based on HRMS generates a list of
219 features and then the accurate masses of these features can be searched against a database containing the
220 accurate masses of metabolites. This generates a list of metabolites which are present in sample. These
221 metabolites can be considered to be putatively identified at this stage to MSI level 2[16] since there may be
222 several isomers for a particular elemental composition. Usually mature global screening methodologies will
223 have lists of retention times for common metabolites and this information can be incorporated into the
224 database search so that both exact mass and chromatographic retention time are matched to a metabolite
225 in order to give MSI level 1 confidence whenever a complete match is achieved. Sometimes in addition to
226 this MS/MS or MS^n fragmentation of a metabolite is carried out and is matched against that of a
227 corresponding standard which increases confidence of identification still further. Authentic standards are
228 not always available for a metabolite or an important marker feature may be completely unknown. In this
229 case MS/MS is the only option for achieving MSI level 1 identification and may only allow partial
230 characterisation. In a study by Dai et al software was produced for capturing ion pairs produced by

Analyst Accepted Manuscript

predictable neutral losses (glucuronide, sulphate etc.) following ion source fragmentation and was used to explore unknown metabolites in the urinary metabolome. In the samples studied phase II conjugation reactions were severely impaired in patients with liver cirrhosis [42]. One method for characterising unknowns is to predict metabolites. In silico prediction of the metabolism of 75 green tea components was carried and predicted 27245 metabolites and led to the identification of 74 known metabolites and 26 new metabolites of green tea in urine [32]. Some good examples of the use of MS/MS or MS^2 for the identification of unknown compounds are covered by several papers within this review [23, 34, 45,67].

Analyst Accepted Manuscript

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Table 1 Summary of urinary metabolomics studies screening for biomarkers from 2012-2014. Small <30 Medium 30-100 Large >100. MSTUS=MS Total Useful Signals, TIC total ion current, N/A not applied. PCA principal components analysis, (O)PLS-DA (orthogonal) partial least squares discriminant analysis, UVA univariate analysis, ROC receiver operator characteristic.

Application	Sample size	Analytical platform/Separation techniques	Data processing software	Normalisation	Statistical Methods
Urogenital Cancer					
Kidney cancer[17]	Medium	UPLC-IT/RP	not mentioned	MSTUS	UVA
Kidney and Bladder cancer[18]	Medium	HPLC-QTOF/RP+HILIC	Profile Analysis (Bruker)	MSTUS	PCA, OPLS-DA, UVA and ROC
Bladder cancer[19]	Large	HPLC-QTOF/RP	MZMine 2	MSTUS	OPLS-DA, UVA and ROC
Bladder cancer[20]	Medium	HPLC-QTOF/RP	XCMS	13C-labeled Universal Metabolome Standard (UMS)	PCA, OPLS-DA and ROC
Bladder cancer[21]	Medium	HPLC-QTOF/RP + CE-QTOF	MassHunter (Agilent)	Creatinine	PCA, OPLS-DA and UVA
Bladder, Kidney and Prostate cancer[22]	Large	HPLC-QQQ/RP	targeted analysis of nucleosides	Creatinine	PCA, PLS-DA and UVA
Prostate cancer[23]	Medium	HPLC-Orbitrap/RP+HILIC	MZMine 2	Creatinine, MSTUS and osmolality	PCA, OPLS-DA, UVA and ROC
Cervical cancer[24]	Medium	HPLC-QTOF/RP	MarkerVeiw (AB Sciex), XCMS and MZMine 2	Creatinine	PCA, OPLS-DA, UVA and ROC
Cervical cancer[25]	Medium	HPLC-QTOF/RP	MassHunter (Agilent)	TIC	PLS and UVA
Ovarian cancer[26]	Large	UPLC-QTOF/RP	XCMS	MSTUS	PCA, PLS-DA and UVA
Ovarian cancer[27]	Medium	UPLC-QTOF/RP+HILIC	MarkerLynx (Waters)	MSTUS	PCA/PLS-DA/OSC-DA
Bladder cancer[28]	Medium	UPLC-QTOF/RP	MassHunter (Agilent)	Creatinine	OPLS-DA

Analyst Accepted Manuscript

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Non-urogenital Cancer					
Liver cancer[29]	Medium	UPLC-QTOF/RP	MarkerLynx (Waters)	N/A	OPLS-DA
Liver cancer[30]	Small	UPLC-QTOF/RP	XCMS online	N/A	PCA and UVA
Stomach cancer[31]	Medium	HPLC-QTOF/RP	MassHunter (Agilent)	N/A	PCA
Lung cancer[32]	Medium	UPLC-QTOF/RP	MarkerLynx (Waters)	MSTUS and creatinine	OPLS-DA, UVA and ROC
Other Disease					
Asthma[33]	Medium	UPLC-QTOF/RP	MarkerLynx (Waters)	MSTUS	OPLS-DA
Xanthinuria[34]	Medium	HPLC-Orbitrap/HILIC	SIEVE (Thermo)	N/A	UVA
Idiopathic nephrotic syndrome[35]	Small	UPLC-QTOF/RP	Comet (Nonliner Dynamics)	TIC	OPLS-DA
Rheumatoid arthritis[36]	Medium	HPLC-Orbitrap	SIEVE (Thermo)	N/A	Nonlinear Principal Component Analysis (NPCA) and PLS-DA
Jaundice syndrome[37]	Medium	UPLC-QTOF/RP	MarkerLynx (Waters)	TIC	OPLS-DA and HCA
Metabolic syndrome[38]	Medium	UPLC-QTOF/RP	Profile Analysis (Bruker)	Creatine	PCA, OPLS-DA and UVA
Liver-stagnation/spleen-deficiency syndrome [39]	Medium	UPLC-QTOF/RP	MarkerLynx (Waters)	N/A	PCA, OPLS-DA, HCA, ROC and UVA
Male infertility[40]	Large	HPLC-QTOF/RP	Profile Analysis (Bruker)	MSTUS	OPLS-DA, ROC and UVA
T-cell mediated rejection[41]	Medium	Targeted analysis QTRAP	Quantitative method	Creatinine	PLS-DA, ROC, Wilcoxon rank test
Hepatitis C[42]	Small	HPLC-QTOF/RP	EZ Info Software (Waters)	N/A	PCA/OPLS/Mann Whitney
Liver cirrhosis[43]	Medium	UPLC-Orbitrap/RP	MS Finder software	MSTUS	PCA/t test/Non parametric test
Toxicity/metabolism					

Analyst Accepted Manuscript

Dioxin exposure[44]	Small	UPLC-QTOF/RP	MarkerLynx (Waters)	MSTUS	OPLS-DA and UVA
Procainamide[45]	Small	UPLC-QTOF/RP	MarkerLynx (Waters)	N/A	OPLS-DA and UVA
Renal clearance rate[46]	Small	UPLC-Orbitrap/RP	MZMine 2	N/A	UVA
Radiation[47]	Medium	UPLC-TOF/RP	MarkerLynx (Waters)	Creatinine	PCA and UVA
Sports					
Doping control[48]	Medium	UPLC-QTOF/RP	Profile Analysis (Bruker)	N/A	PCA, OPLS-DA and UVA
Tetrahydrocannabinol[49]	Small	UPLC-TOF/RP	MarkerLynx (Waters)	N/A	ANOVA/PLSDA/OPLS
Treatment					
Chinese herbal formula[50]	Medium	UPLC-QTOF/RP	MarkerLynx (Waters)	MSTUS	PCA, OPLS-DA and UVA
Chinese herbal formula [51]	Medium	HPLC-Orbitrap/RP	SIEVE (Thermo)	N/A	OPLS-DA and UVA
Standard therapy for TB[52]	Medium	HPLC-QTOF/RP	MassHunter (Agilent)	Creatinine	PCA and UVA
Chinese herbal formula[53]	Small	UPLC-QTOF/RP	MarkerLynx (Waters)	TIC	PCA, OPLS-DA and UVA
Schizophrenia risperidone[54]	Medium	UPLC-QTOF/RP	MarkerLynx (Waters)	TIC	OPLS-DA and UVA
Food and nutrition					
Cocoa[55]	Small	HPLC-QTOF/RP	MarkerView (AB Sciex)	N/A	OPLS-DA and UVA
Coffee[56]	Medium	UPLC-QTOF/RP	XCMS	N/A	PLS-DA, ROC and UVA
Fruits and vegetable[57] consumption vs basal	Medium	HPLC-Orbitrap/RP	msInspect	N/A	PCA and UVA
Citrus juice[58]	Small	HPLC-QTOF/RP	Profile Analysis (Bruker)	Largest peak	PCA
Citrus fruit[59]	Large	UPLC-QTOF/RP	MarkerLynx (Waters)	N/A	PCA, HCA, PLS-DA and UVA
Aronia-citrus juice[60]	Medium	HPLC-QTOF/RP	MarkerView (AB Sciex)	N/A	OPLS-DA and UVA

Analyst Accepted Manuscript

Pu-erh Tea[61]	Small	UPLC-QTOF/RP	MarkerLynx (Waters)	N/A	PCA, OPLS-DA and UVA
Goji Tea[62]	Small	HPLC-QTOF/RP	Profile Analysis (Bruker)	N/A	PCA, PLS-DA and UVA
Green tea[63]	Small	HPLC-Orbitrap/RP	MAGMa	N/A	N/A
vitamin E capsules[64]	Small	UPLC-QTOF/RP	MarkerLynx (Waters)	Internal standard	PCA, OPLS-DA and UVA
Dietary pattern[65]	Large	UPLC-QTOF/RP	Mzmine 2	MSTUS	PCA, PLS-DA and UVA

Analyst Accepted Manuscript

Conclusion

It is apparent that LC-MS profiling of metabolites urine has the potential to improve diagnosis of treatment of disease. It is less invasive to collect than plasma and is available in larger volumes than plasma so that for trace metabolites a concentration step is an option. Instrumental methods in clinical biochemistry have a mixed history and have found some niche applications such as in monitoring in-born errors of metabolism. However, where a limited range of biomarkers were being screened in rare conditions instruments were often deemed too expensive buy and run. By using metabolomics methods markers for many diseases can be screened for with same platform. If targeted MS methods are used run times for biomarker screening may be reduced to a minute or two per sample. In addition dipstick tests for new biomarkers discovered by metabolomics screens might be produced. Thus clinical metabolomics has a promising future.

References

1. D. Ryan, K. Robards, P. D. Prenzler and M. Kendall, *Anal Chim Acta*, 2011, 684, 17-29.
2. S. Bouatra, F. Aziat, R. Mandal, A. C. Guo, M. R. Wilson, C. Knox, T. C. Bjorndahl, R. Krishnamurthy, F. Saleem, P. Liu, Z. T. Dame, J. Poelzer, J. Huynh, F. S. Yallou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring and D. S. Wishart, *Plos One*, 2013, 8.
3. H. G. Gika, G. A. Theodoridis, R. S. Plumb and I. D. Wilson, *J Pharm.Biomed*, 2014, 87, 12-25.
4. H. G. Gika, I. D. Wilson and G. A. Theodoridis, *J Chromatogr B*, 2014, 966, 1-6.
5. S. Naz, M. Vallejo, A. Garcia and C. Barbas, *J Chromatogr A*, 2014, 1353, 99-105.
6. R. Ramautar and G. J. de Jong, *Bioanalysis*, 2014, 6, 1011-1026.
7. L. Denoroy, L. Zimmer, B. Renaud and S. Parrot, *J Chromatogr B*, 2013, 927, 37-53.
8. G. A. Theodoridis, F. Michopoulos, H. G. Gika, R. S. Plumb and I. D. Wilson, *Chromatographic Methods in Metabolomics*, 2013, DOI: Book_Doi 10.1039/9781849737272, 64-86.
9. W. B. Dunn, I. D. Wilson, A. W. Nicholls and D. Broadhurst, *Bioanalysis*, 2012, 4, 2249-2264.
10. M. A. Fernandez-Peralbo and M. D. L. de Castro, *TrAC*, 2012, 41, 75-85.
11. D. Rojo, C. Barbas and F. J. Ruperez, *Bioanalysis*, 2012, 4, 1235-1243.
12. G. A. Theodoridis, H. G. Gika, E. J. Want and I. D. Wilson, *Anal Chim Acta*, 2012, 711, 7-16.
13. A. H. Zhang, H. Sun, X. H. Wu and X. J. Wang, *Clin Chim Acta*, 2012, 414, 65-69.

Analyst Accepted Manuscript

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

14. T. Zhang and D.G. Watson. *Current Metabolomics* 2013, 1, 58-83.

15. D.G. Watson. *Comp Struct Biotech J.* 2013, 4.

16. L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W.-M. Fan, O. Fiehn, R. Goodacre and J. L. Griffin, *Metabolomics*, 2007, 3, 211-221.

17. S. Ganti, S. L. Taylor, K. Kim, C. L. Hoppel, L. N. Guo, J. Yang, C. Evans and R. H. Weiss, *Int J Cancer*, 2012, 130, 2791-2800.

18. Z. Z. Huang, Y. J. Chen, W. Hang, Y. Gao, L. Lin, D. Y. Li, J. C. Xing and X. M. Yan, *Metabolomics*, 2013, 9, 119-129.

19. X. Jin, S. J. Yun, P. Jeong, I. Y. Kim, W. J. Kim and S. Park, *Oncotarget*, 2014, 5, 1635-1645.

20. J. Peng, Y. T. Chen, C. L. Chen and L. Li, *Anal Chem*, 2014, 86, 6540-6547.

21. J. V. Alberice, A. F. S. Amaral, E. G. Armitage, J. A. Lorente, F. Algaba, E. Carrilho, M. Marquez, A. Garcia, N. Malats and C. Barbas, *J Chromatogr A*, 2013, 1318, 163-170.

22. W. Struck, D. Siluk, A. Yumba-Mpanga, M. Markuszewski, R. Kaliszan and M. J. Markuszewski, *J Chromatogr A*, 2013, 1283, 122-131.

23. T. Zhang, D. G. Watson, L. J. Wang, M. Abbas, L. Murdoch, L. Bashford, I. Ahmad, N. Y. Lam, A. C. F. Ng and H. Y. Leung, *Plos One*, 2013, 8.

24. Y. H. Chen, J. Xu, R. P. Zhang, G. Q. Shen, Y. M. Song, J. H. Sun, J. M. He, Q. M. Zhan and Z. Abliz, *Analyst*, 2013, 138, 2669-2677.

25. Q. Liang, Q. Yu, H. Wu, Y. Z. Zhu and A. H. Zhang, *Anal Methods-Uk*, 2014, 6, 3937-3942.

26. J. Chen, L. N. Zhou, X. Y. Zhang, X. Lu, R. Cao, C. J. Xu and G. W. Xu, *Electrophoresis*, 2012, 33, 3361-3369.

27. J. Peng, Y. T. Chen, C. L. Chen and L. Li, *Anal Chem*, 2014, 86, 6540-6547.

28. J. V. Alberice, A. F. S. Amaral, E. G. Armitage, J. A. Lorente, F. Algaba, E. Carrilho, M. Marquez, A. Garcia, N. Malats and C. Barbas, *J Chromatogr A*, 2013, 1318, 163-170.

29. A. H. Zhang, H. Sun, G. L. Yan, Y. Han, Y. Ye and X. J. Wang, *Clin Chim Acta*, 2013, 418, 86-90.

30. A. H. Zhang, P. Wang, H. Sun, G. L. Yan, Y. Han and X. J. Wang, *Mol Biosyst*, 2013, 9, 2259-2265.

31. Y. Zhang, H. Ren, Y. Jiang, Y.-f. Gao and S.-y. Liu, *Chinese Med J-Peking*, 2013, 126, 1930-1933.

Analyst Accepted Manuscript

32. Q. Wu, Y. Wang, X. Gu, J. Y. Zhou, H. P. Zhang, W. Lv, Z. Chen and C. Yan, *J Sep Sci*, 2014, 37, 1728-1735.

33. E. Mattarucchi, E. Baraldi and C. Guillou, *Biomed Chromatogr*, 2012, 26, 89-94.

34. H. Peretz, D. G. Watson, G. Blackburn, T. Zhang, A. Lagziel, M. Shtauber-Naamati, T. Morad, E. Keren-Tardai, V. Greenshpun, S. Usher, H. Shalev, D. Landau and D. Levartovsky, *Metabolomics*, 2012, 8, 951-959.

35. M. Sedic, L. A. Gethings, J. P. C. Vissers, J. P. Shockcor, S. McDonald, O. Vasieva, M. Lemac, J. I. Langridge, D. Batinic and S. K. Pavelic, *Biochem Bioph Res Co*, 2014, 452, 21-26.

36. H. A. van Wietmarschen, W. D. Dai, A. J. van der Kooij, T. H. Reijmers, Y. Schroen, M. Wang, Z. L. Xu, X. C. Wang, H. W. Kong, G. W. Xu, T. Hankemeier, J. J. Meulman and J. van der Greef, *Plos One*, 2012, 7.

37. X. J. Wang, A. H. Zhang, Y. Han, P. Wang, H. Sun, G. C. Song, T. W. Dong, Y. Yuan, X. X. Yuan, M. Zhang, N. Xie, H. Zhang, H. Dong and W. Dong, *Mol Cell Proteomics*, 2012, 11, 370-380.

38. Z. R. Yu, Y. Ning, H. Yu and N. J. Tang, *J Huazhong U Sci-Med*, 2014, 34, 276-283.

39. J. Fan, J. Hong, J. D. Hu and J. L. Chen, *Gastroent Res Pract*, 2012, DOI: Artn 474907

40. J. Zhang, Z. Z. Huang, M. J. Chen, Y. K. Xia, F. L. Martin, W. Hang and H. Q. Shen, *Fertil Steril*, 2014, 102, 44-+.

41. T. D. Blydt-Hansen, A. Sharma, I. W. Gibson, R. Mandal and D. S. Wishart, *Am J Transplant*, 2014, 14, 2339-2349.

42. A. H. Zhang, H. Sun, Y. Han, G. L. Yan, Y. Yuan, G. C. Song, X. X. Yuan, N. Xie and X. J. Wang, *Anal Chem*, 2013, 85, 7606-7612.

43. W. D. Dai, P. Y. Yin, Z. D. Zeng, H. W. Kong, H. W. Tong, Z. L. Xu, X. Lu, R. Lehmann and G. W. Xu, *Anal Chem*, 2014, 86, 9146-9153.

44. F. Jeanneret, J. Boccard, F. Badoud, O. Sorg, D. Tonoli, D. Pelclova, S. Vlckova, D. N. Rutledge, C. F. Samer, D. Hochstrasser, J. H. Saurat and S. Rudaz, *Toxicol Lett*, 2014, 230, 234-243.

45. F. Li, A. D. Patterson, C. C. Hofer, K. W. Krausz, F. J. Gonzalez and J. R. Idle, *Biochem Pharmacol*, 2011, 81, 1043-1053.

Analyst Accepted Manuscript

46. T. L. Sirich, P. A. Aronov, N. S. Plummer, T. H. Hostetter and T. W. Meyer, *Kidney Int*, 2013, 84, 585-590.
47. E. C. Laiakis, T. D. Mak, S. Anizan, S. A. Amundson, C. A. Barker, S. L. Wolden, D. J. Brenner and A. J. Fornace, *Radiat Res*, 2014, 181, 350-361.
48. A. Kiss, M. Lucio, A. Fildier, C. Buisson, P. Schmitt-Kopplin and C. Cren-Olive, *Plos One*, 2013, 8.
49. A. Kiss, C. Bordes, C. Buisson, F. Lasne, P. Lanteri and C. Cren-Olive, *Anal Bioanal Chem*, 2014, 406, 1209-1219.
50. P. Liu, J. A. Duan, P. J. Wang, D. W. Qian, J. M. Guo, E. X. Shang, S. L. Su, Y. P. Tang and Z. X. Tang, *Mol Biosyst*, 2013, 9, 77-87.
51. C. J. Lu, J. W. Deng, L. Li, D. M. Wang and G. Z. Li, *Bba-Proteins Proteom*, 2014, 1844, 280-288.
52. S. Mahapatra, A. M. Hess, J. L. Johnson, K. D. Eisenach, M. A. DeGroote, P. Gitta, M. L. Joloba, G. Kaplan, G. Walzl, W. H. Boom and J. T. Belisle, *Bmc Infect Dis*, 2014, 14.
53. S. Su, J. Duan, P. Wang, P. Liu, J. Guo, E. Shang, D. Qian, Y. Tang and Z. Tang, *J Proteome Res*, 2013, 12, 852-865.
54. H. L. Cai, H. D. Li, X. Z. Yan, B. Sun, Q. Zhang, M. Yan, W. Y. Zhang, P. Jiang, R. H. Zhu, Y. P. Liu, P. F. Fang, P. Xu, H. Y. Yuan, X. H. Zhang, L. Hu, W. Yang and H. S. Ye, *J Proteome Res*, 2012, 11, 4338-4350.
55. R. Llorach, M. Urpi-Sarda, S. Tulipani, M. Garcia-Aloy, M. Monagas and C. Andres-Lacueva, *Mol Nutr Food Res*, 2013, 57, 962-973.
56. A. Rothwell, Y. Fillatre, J. F. Martin, B. Lyan, E. Pujos-Guillot, L. Fezeu, S. Hercberg, B. Comte, P. Galan, M. Touvier and C. Manach, *Plos One*, 2014, 9.
57. D. H. May, S. L. Navarro, I. Ruczinski, J. Hogan, Y. Ogata, Y. Schwarz, L. Levy, T. Holzman, M. W. McIntosh and J. W. Lampe, *Brit J Nutr*, 2013, 110, 1760-1770.
58. JS. Medina, F. Ferreres, C. Garcia-Viguera, M. N. Horcajada, J. Orduna, M. Saviron, G. Zurek, J. M. Martinez-Sanz, J. I. Gil and A. Gil-Izquierdo, *Food Chem*, 2013, 136, 938-946.
59. E. Pujos-Guillot, J. Hubert, J.-F. o. Martin, B. Lyan, M. Quintana, S. Claude, B. Chabanas, J. A. Rothwell, C. Bennetau-Pelissero and A. Scalbert, *J Proteome Res*, 2013, 12, 1645-1659.

Analyst Accepted Manuscript

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

60. R. Llorach, S. Medina, C. Garcia-Viguera, P. Zafrilla, J. Abellan, O. Jauregui, F. A. Tomas-Barberan, A. Gil-Izquierdo and C. Andres-Lacueva, *Electrophoresis*, 2014, 35, 1599-1606.
61. G. X. Xie, A. H. Zhao, L. J. Zhao, T. L. Chen, H. Y. Chen, X. Qi, X. J. Zheng, Y. Ni, Y. Cheng, K. Lan, C. Yao, M. F. Qiu and W. Jia, *J Proteome Res*, 2012, 11, 3449-3457.
62. C. L. Tseng and L. Li, *Sci China Chem*, 2014, 57, 678-685.
63. L. Ridder, J. J. J. van der Hooft, S. Verhoeven, R. C. H. de Vos, J. Vervoort and R. J. Bino, *Anal Chem*, 2014, 86, 4767-4774.
64. C. H. Johnson, O. Slanar, K. W. Krausz, D. W. Kang, A. D. Patterson, J. H. Kim, H. Luecke, F. J. Gonzalez and J. R. Idle, *Am J Clin Nutr*, 2012, 96, 818-830.
65. M. B. S. Andersen, A. Rinnan, C. Manach, S. K. Poulsen, E. Pujos-Guillot, T. M. Larsen, A. Astrup and L. O. Dragsted, *J Proteome Res*, 2014, 13, 1405-1418.
66. T. Zhang, D. J. Creek, M. P. Barrett, G. Blackburn and D. G. Watson, *Anal Chem*, 2012, 84, 1994-2001.
67. A. Roux, Y. Xu, J. F. Heilier, M. F. Olivier, E. Ezan, J. C. Tabet and C. Junot, *Anal Chem*, 2012, 84, 6429-6437.
68. D. P. Kloos, H. Lingeman, W. M. A. Niessen, A. M. Deelder, M. Giera and O. A. Mayboroda, *J Chromatogr B*, 2013, 927, 90-96.
69. D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. F. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. G. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert, *Nucleic Acids Res*, 2013, 41, D801-D807.
70. H. P. Benton, E. Want, H. C. Keun, A. Amberg, R. S. Plumb, F. Goldfain-Blanc, B. Walther, M. D. Reily, J. C. Lindon, E. Holmes, J. K. Nicholson and T. M. D. Ebbels, *Anal Chem*, 2012, 84, 2424-2432.
71. C. G. Juo, C. L. Chen, S. T. Lin, S. H. Fu, Y. T. Chen, Y. S. Chang and J. S. Yu, *Rapid Commun Mass Sp*, 2014, 28, 1813-1820.
72. G. Glauser, N. Veyrat, B. Rochat, J.-L. Wolfender and T. C. Turlings, *J Chromatogr A*, 2013, 1292, 151-159.
73. W. Koch, S. Forcisi, R. Lehmann and P. Schmitt-Kopplin, *J Chromatogr A*, 2014, 1361, 209-216.
74. X. M. Zhang, M. R. Clausen, X. L. Zhao, H. Zheng and H. C. Bertram, *Anal Chem*, 2012, 84, 7785-7792.

Analyst Accepted Manuscript

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

75. P. Liu, Y. Q. Huang, W. J. Cai, B. F. Yuan and Y. Q. Feng, *Anal Chem*, 2014, 86, 9765-9773.

76. J. Peng and L. Li, *Anal Chim Acta*, 2013, 803, 97-105.

77. A. Stanislaus, K. Guo and L. Li, *Anal Chim Acta*, 2012, 750, 161-172.

78. R. K. Zhou, K. Guo and L. Li, *Anal Chem*, 2013, 85, 11532-11539.

79. R. Zhou, C. L. Tseng, T. Huan and L. Li, *Anal Chem*, 2014, 86, 4675-4679.

80. C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan and G. Siuzdak, *Anal Chem*, 2006, 78, 779-787.

81. M. Katajamaa, J. Miettinen and M. Orešič, *Bioinformatics*, 2006, 22, 634-636.

82. Y. H. Chen, J. Xu, R. P. Zhang, G. Q. Shen, Y. M. Song, J. H. Sun, J. M. He, Q. M. Zhan and Z. Abliz, *Analyst*, 2013, 138, 2669-2677.

83. H. Zheng, M. R. Clausen, T. K. Dalsgaard, G. Mortensen and H. C. Bertram, *Anal Chem*, 2013, 85, 7109-7116.

84. R. Zhang, D. G. Watson, L. Wang, G. D. Westrop, G. H. Coombs and T. Zhang, *J. Chromatogr. A*, 2014, 1362, 168-179.

85. E. Mattarucchi and C. Guillou, *Biomed Chromatogr*, 2012, 26, 512-517.

86. C. A. Crutchfield, M. T. Olson, E. Gourgari, M. Nesterova, C. A. Stratakis and A. L. Yergey, *J Am Soc Mass Spectr*, 2013, 24, 230-237.

87. Y. V. Karpievitch, S. B. Nikolic, R. Wilson, J. E. Sharman and L. M. Edwards, *PloS one*, 2014, 9, e116221.

88. B. M. Warrack, S. Hnatyshyn, K.-H. Ott, M. D. Reily, M. Sanders, H. Zhang and D. M. Drexler, *J Chromatogr B*, 2009, 877, 547-552.

89. J. Zieliński, T. Rychlewski, K. Kusy, K. Domaszewska and M. Laurentowska, *European J. App Physiol*, 2009, 106, 867-876.

90. R. Lehmann, X. Zhao, C. Weigert, P. Simon, E. Fehrenbach, J. Fritsche, J. Machann, F. Schick, J. Wang and M. Hoene, *PloS One*, 2010, 5, e11519.

91. G. D. Lewis, L. Farrell, M. J. Wood, M. Martinovic, Z. Arany, G. C. Rowe, A. Souza, S. Cheng, E. L. McCabe and E. Yang, *Sci Trans Med*, 2010, 2, 33ra37-33ra37.

92. Wu Y and Li L *Anal. Chem.*, 2012, 84, 10723-10731.

Analyst Accepted Manuscript

93. W. M. B. Edmands, P. Ferrari and A. Scalbert, *Anal Chem*, 2014, 86, 10925-10931.

94. E. Saccenti, H. C. Hoefsloot, A. K. Smilde, J. A. Westerhuis and M. M. Hendriks, *Metabolomics*, 2014, 10, 361-374.

95. J. Boccard and D. N. Rutledge, *Anal Chim Acta*, 2013, 769, 30-39.

96. J. Xia, D. I. Broadhurst, M. Wilson and D. S. Wishart, *Metabolomics*, 2013, 9, 280-299.

97. M. M. Hendriks, F. A. van Eeuwijk, R. H. Jellema, J. A. Westerhuis, T. H. Reijmers, H. C. Hoefsloot and A. K. Smilde, *TrAC*, 2011, 30, 1685-1698.

Analyst Accepted Manuscript