Urine is a non-invasive biofluid that is rich in polar metabolites and well-suited for metabolomic epidemiology. However, due to individual variability in health and hydration status, the physiological concentration of urine can differ >15-fold, which can pose major challenges in untargeted LC-MS metabolomics. Although numerous urine normalization methods have been implemented (e.g., creatinine, specific gravity – SG), most are manual and therefore not practical for population-based studies. To address this issue, we developed a method to measure SG in 96-well-plates using a refractive index detector (RID), which exhibited accuracy within 85-115% and <3.4% precision. Bland-Altman statistics showed a mean deviation of -0.0001 SG units (limits of agreement: -0.0014-0.0011) relative to a hand held refractometer. Using this RID-based SG normalization, we developed an automated LC MS workflow for untargeted urinary metabolomics in 96-well-plate format. The workflow uses positive and negative ionization HILIC chromatography and acquires mass spectra in data independent acquisition (DIA) mode at 3 collision energies. Five technical internal standards (tISs) were used to monitor data quality in each method, all of which demonstrated raw coefficients of variation (CVs) <10% in the quality controls (QCs) and <20% in the samples for a small cohort (n=87 samples, n=22 QCs). Application in a large cohort (n=842 urine samples, n=248 QCs), demonstrated CVQC<5% and CVsamples<16% for 4/5 tISs after signal drift correction by cubic spline regression. The workflow identified >540 urinary metabolites including endogenous and exogenous compounds. This platform is suitable for performing urinary untargeted metabolomic epidemiology and will be useful for applications in population-based molecular phenotyping.
High-precision automated workflow for urinary untargeted metabolomic epidemiology

Isabel Meister1,2, Pei Zhang1,2,†, Anirban Sinha3,4, C. Magnus Sköld5, Åsa M. Wheelock5, Takashi Izumi1,6, Romanas Chaleckis1,2,*,‡, Craig E. Wheelock1,2,*,‡

1 Gunma University Initiative for Advanced Research (GIAR), Gunma University, Maebashi, Gunma, Japan
2 Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden
3 Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
4 Department of Experimental Immunology, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands
5 Division of Respiratory Medicine and Allergy, Department of Medicine, Karolinska Institutet, Stockholm, Sweden
6 Department of Biochemistry, Gunma University Graduate School of Medicine, Maebashi, Gunma, Japan

†Current address: Key Laboratory of Drug Quality Control and Pharmacovigilance (Ministry of Education), State Key Laboratory of Natural Medicine, China Pharmaceutical University, Nanjing 210009, China

Corresponding Authors:
Romanas Chaleckis
Gunma University Initiative for Advanced Research (GIAR), Gunma University, Gunma, Maebashi, Japan.
Email: romcha@gunma-u.ac.jp.

Craig E Wheelock
Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden.
Email: craig.wheelock@ki.se
ABSTRACT

Urine is a non-invasive biofluid that is rich in polar metabolites and well-suited for metabolomic epidemiology. However, due to individual variability in health and hydration status, the physiological concentration of urine can differ >15-fold, which can pose major challenges in untargeted LC-MS metabolomics. Although numerous urine normalization methods have been implemented (e.g., creatinine, specific gravity – SG), most are manual and therefore not practical for population-based studies. To address this issue, we developed a method to measure SG in 96-well-plates using a refractive index detector (RID), which exhibited accuracy within 85-115% and <3.4% precision. Bland-Altman statistics showed a mean deviation of -0.0001 SG units (limits of agreement: -0.0014-0.0011) relative to a hand-held refractometer. Using this RID-based SG normalization, we developed an automated LC-MS workflow for untargeted urinary metabolomics in 96-well-plate format. The workflow uses positive and negative ionization HILIC chromatography and acquires mass spectra in data independent acquisition (DIA) mode at 3 collision energies. Five technical internal standards (tISs) were used to monitor data quality in each method, all of which demonstrated raw coefficients of variation (CVs) <10% in the quality controls (QCs) and <20% in the samples for a small cohort (n=87 samples, n=22 QCs). Application in a large cohort (n=842 urine samples, n=248 QCs), demonstrated CV_{QC}<5% and CV_{samples}<16% for 4/5 tISs after signal drift correction by cubic spline regression. The workflow identified >540 urinary metabolites including endogenous and exogenous compounds. This platform is suitable for performing urinary untargeted metabolomic epidemiology and will be useful for applications in population-based molecular phenotyping.
INTRODUCTION

The use of metabolomics is increasing in clinical research and metabolomics data have become an essential component of molecular phenotyping.\(^1\) In conjunction with these developments, the discipline of metabolomic epidemiology has been created, which involves the systematic use of epidemiological methods and principles to study population-based variation in the human metabolome as it associates with health-related outcomes and exposures.\(^2\) These efforts regularly involve the need to analyze large-scale studies of thousands of individuals, which can be a significant obstacle for many untargeted analytical methods. In order to meet the demands of this field in terms of experimental throughput and data quality, there is a need to develop untargeted metabolomics methods that have high precision and can be fully automated.

Due to its ease of collection, urine is a particularly well-suited biofluid for metabolomic epidemiology and population-based molecular phenotyping. The urinary chemical composition represents an integrated snapshot of the entire organism, providing insight into systemic metabolism. However, urine poses a number of challenges for untargeted LC-MS-based metabolomics, including the large amount of salts (inorganic salts, urea and creatinine collectively represent ~84% of the total urine solutes\(^3\)), wide variation according to the hydration status of the individual (urinary specific gravity, [SG], varies >15-fold\(^4\)), and large dynamic range of metabolites (>10 orders of magnitude\(^5\)), which can collectively result in retention time shifts and matrix effects.\(^6\)\(^-\)\(^8\) The dilution and normalization of urine samples to a uniform concentration prior to untargeted LC-MS measurements is crucial to achieve high data quality because post-acquisition normalization methods are unable to fully compensate for concentration-related matrix effects.\(^6\)\(^,\)\(^9\)\(^,\)\(^10\) It is therefore common practice to normalize the urinary concentration using a variety of methods including SG,\(^11\) creatinine,\(^12\) or osmolality.\(^13\) SG is the preferred urinary normalization method employed by the World Anti-Doping Agency (WADA),\(^14\) due to its ease of measurement and reduced confounding effects. SG measurements are currently performed by refractometry, while true SG measurements by gravimetry are comparatively complicated and have become obsolete.\(^15\)\(^,\)\(^16\) The refractometry SG is an indirect measurement of SG, where the refractive index (RI) observed in urine is converted to an SG value using experimental conversion tables.\(^16\) In practice, SG is generally measured using a hand-held refractometer, which is time-consuming and not amenable to automation.

To address these needs and challenges, we report the development of an automated untargeted metabolomics workflow for urine. In particular, a differential RI method was
developed to automate the SG measurement and all sample preparation steps are performed
using a liquid handling system and 96-well plates. Samples are then analyzed using a
combination of positive and negative ionization HILIC chromatography to provide wide
metabolic coverage of urinary metabolites.\(^{17}\) Taken together, this combined workflow
presents a fully automated method for urinary untargeted metabolomics that offers the
throughput and precision necessary for applications in molecular phenotyping and
metabolic epidemiology.

**EXPERIMENTAL SECTION**

**Chemicals, solvents and urine samples.** The complete list of solvents and chemicals,
including standards used to create the spectral libraries, is provided in Table S1. LC-MS
grade solvents were used throughout this work. Spectral libraries were acquired in data-
dependent (DDA) and data-independent (DIA) acquisition modes.\(^{18}\) Five technical internal
standards (tIS; Table S2) were selected for monitoring the system performance.

The method was characterized using fasting urine samples (n=87) from subjects from
the LUNg obstruction in Adulthood of PREmaturely born (LUNAPRE) study
(ClinicalTrials.gov NCT02923648).\(^{19}\) Stability assessments were performed using pooled
urine from all individuals (hereafter referred to as pooled study quality control sample, SQC)
as well as internal laboratory reference urine from healthy volunteers (pooled laboratory
reference quality control, RQC). Proof-of-concept studies for large-scale application were
performed using urine samples from a rhinovirus challenge study (n=842; Netherlands Trial
Register NTR5426/NL5317).\(^{20}\)

**Sample aliquoting.** Urine samples were thawed overnight at 4°C, vortexed, and 1.5 mL was
transferred to each well of a 2 mL 96-deepwell plate placed on a metal block on ice
(instrumentation and materials are listed in Table S3). An additional aliquot of 1.5 mL of
each sample was pooled and used to create the SQC sample, which was then transferred to
the dedicated wells of the deepwell-plate (Fig. S1). To avoid multiple freeze-thaw cycles, the
deepwell plate content was directly aliquoted in 15 aliquots of 120 µL into 0.2 mL 96-well
plates using a Bravo automated liquid handling platform (Agilent Technologies, Inc.)
equipped with a cooling unit set at 4°C using 250 µL tips. Both deepwell and aliquot plates
were sealed with peel seals on a thermal microplate sealer for 2 s at 170°C and stored at -
80°C until use.
For all subsequent sample processing steps, one aliquot plate was placed on a metal insert for 1 h at 4°C for uniform thawing. The plate was then shaken for 10 min at 1600 rpm and 4°C in a thermomixer, centrifuged at 4390 g in a plate centrifuge for 40 min at 4°C, and 105 µL of supernatant was pipetted using the liquid handler to a new 0.2 mL plate.

**Specific gravity measurement.** Manual measurements of SG were performed using an UG-D hand-held refractometer (Atago, Inc.). Samples were equilibrated to room temperature prior to SG measurements. The refractometer was calibrated with high-purity water, wiped with a lint-free tissue after each measurement and re-calibrated after every 20 samples.

Automated SG measurement were performed using an Agilent 1260 differential refractive index detector (RID) coupled to an Agilent 1290 multisampler and an Agilent 1260 quaternary pump with high-purity water as the single mobile phase. In contrast to the static RI cell of the hand-held refractometer, the differential RI is a single flow-through cell divided into two compartments: the reference compartment (filled with mobile phase and is static over the time of measurement) and the sample compartment (receives urine samples delivered in the mobile phase). As the urine passes into the sample cell, the light beam is diffracted, producing a differential with the reference cell, which is translated into nano RI units (nRIU). The mobile phase flow rate was set at 2 mL/min, the urine injection volume was 0.6 µL, and the optical unit temperature was 35°C. The total run time was 0.09 min per sample and the acquisition rate was 37 Hz. OpenLab ver 2.4 software was used to control the system and integrate the RID signal nRIU peak areas. For the conversion of nRIU peak area values into SG, a 10-point calibration curve was prepared in NaCl by serial dilution of a 2 M stock solution in water with corresponding SGs of 1.002-1.057 measured with the UG-D refractometer. A commercial 1 M NaCl solution was used to prepare QC samples for the calibration, with 1 M as the highest QC, and further diluted 2-, 10- and 20-fold with water for the middle, low and LLOQ QCs. These correspond to SG values of 1.030, 1.015, 1.003 and 1.002 on the UG-D refractometer, respectively. Sample preparation and measurement parameters are detailed in Table S4. A linear calibration curve with 1/x² weighting was applied. Intra- and inter-day accuracy and precision were assessed following the FDA guidelines. All estimates were calculated with SG values-1 to avoid underestimation of accuracy and precision values due to the narrow range of values compared to the blank value (the SG of water is 1.000). Manual vs. automated SG measurements were compared using Bland-Altman plots in R.22
**Sample preparation.** Urine normalization was performed on the liquid handler cooling unit set at 4°C and equipped with a metal insert. High-purity water was used to dilute urine to a common SG value (1.002) in 0.45 mL 96-well plates. Normalization was performed using the equation adapted from Levine and Fahys\(^{23}\)  
\[
   \text{volume}_{\text{urine}} = \frac{\text{total volume} \cdot (\text{SG}_{\text{target}} - 1)}{(\text{SG}_{\text{sample}} - 1)},
\]
where the total volume is 340 µL and the target SG value is 1.002 (lowest normal physiological value\(^{4}\)). To avoid evaporation during the 2.7 h of the normalization process, aliquot plates and normalized urine plates were covered with slit seals.

Depending on the LC-MS method, normalized urine was diluted to 100 µL with acetonitrile containing tIS in 0.2 mL 96-well plates using the liquid handler (1:9 v/v for positive ionization or 3:7 v/v for negative ionization, see Table S2 for tIS concentrations), sealed with peel seals and incubated at 4°C for 2 h. Plates were centrifuged at 4390 g and 4°C for 40 min, followed by transfer of 50 µL of supernatant to new 0.2 mL plates using the liquid handler, and then sealed with pierce seals for direct use in the LC-MS multisampler.

**LC-MS measurements.** LC-MS settings for the measurement of urine samples are detailed in Table S5. Briefly, samples were measured on an Agilent 1290 Infinity II ultrahigh-performance liquid chromatography (UHPLC) system coupled to a 6550 iFunnel quadrupole-time-of-flight (Q-TOF) mass spectrometer equipped with a dual AJS electrospray ionization source tuned for the 50-750 m/z range. The positive mode chromatography was adapted from Naz et al.\(^{24}\) and Chaleckis et al.\(^{25}\) using a SeQuant ZIC-HILIC column and a gradient between (A) water containing 0.1% formic acid (pH=2.6) and (B) acetonitrile containing 0.1% formic acid. The separation gradient included an isocratic step at 95% B for 1.5 min followed by a gradient to 40% B in 10.5 min. The negative mode chromatography was run on a SeQuant ZIC-pHILIC column with (A) ammonium acetate 5 mM in water (pH=9.3) and (B) pure acetonitrile as mobile phases. The gradient was set at 88-60% B from 0-8.5 min and the column oven was heated at 35°C. The acquisition was performed in DIA mode using a mass range of 40-1200 m/z with three different collision energies (0, 10 and 30 eV). The ZIC-pHILIC method was developed using a previous column version equipped with PEEK frits. However, the manufacturer changed the frit material from PEEK to titanium in 2019, resulting in exceeding the column pressure limit during the wash step. The wash and reequilibration steps have been subsequently modified to remain within the column pressure limits for the current column version as described in Table S5.
**Data availability.** The LUNAPRE study datasets have been deposited in the EMBL-EBI MetaboLights repository\(^\text{26}\) with the identifier MTBLS2295. Chemical standard retention times were submitted to PredRet\(^\text{27}\) and MS spectra to MoNA (MassBank of North America).

**Data quality check and pre-processing.** The data quality check and pre-processing procedure are described in the supplemental methods and Tables S6-S10. Briefly, raw files were converted to mzML using ProteoWizard\(^\text{28}\) and an initial quality check was performed in MZmine 2.53\(^\text{29}\) (Table S6) using a pre-defined list of metabolites (Table S7). For pre-processing of the annotated dataset, mzML files were converted to ‘Analysis Base File’ (ABF) format using Reifycs Abf Converter and loaded into MS-DIAL 4.20\(^\text{30}\) (Table S8-S9).\(^\text{31}\) Data was deconvoluted using MS2Dec\(^\text{31}\) and CorrDec\(^\text{32}\). Identifications were based on in-house compound libraries containing 404 and 295 chemical standards for the ZHP and ZHN platforms, respectively (Table S1 and previous studies).\(^\text{18, 24}\) Metabolites annotated by MS-DIAL were further curated using the following criteria: a retention time (RT) shift <0.5 min, a mass shift <10 mDa, and, for spectral match, both a dot product score without weighting above 700 and at least 3 matching MS\(^2\) peaks with the reference spectra to avoid spurious high scores from too low number of peaks.\(^\text{33}\) Any annotations that did not fulfill one of the RT or m/z criteria, but were still determined to be accurate received an explanatory comment in the identification tables. CVs and mean peak intensities were computed for the SQC samples. Missingness (percentage of samples below 5x the blank signal for a given metabolite), skewness, interquartile range (IQR) and D-ratios (percentage ratio of the SQC to sample standard deviations\(^\text{34}\)) were calculated in the samples for all annotated metabolites using R scripts.\(^\text{22}\) For pre-processing the all-feature dataset, peak detection at the MS1 level was performed in MZmine (Table S10).\(^\text{35}\) The features were then filtered in R using the following criteria: CV in the SQCs <35%, peak heights within the dynamic range of the instrument (for the system reported here: 1.5\(\times\)10\(^3\)-3.5\(\times\)10\(^6\)), missingness <90%, D-ratio <55%, IQR >80 and skewness >-2. After exporting the peak lists from MS-DIAL (annotated list) or MZmine (all-feature list), signals were corrected for measurement drift using a Matlab algorithm based upon the SQC signals.\(^\text{34}\)

**Stability assessment of urine samples.** The stability during the normalization process on the liquid handler was tested by leaving normalized urine samples on the cooling deck for 3 h (the time required to normalize a full 96-well plate) compared to 3 h on wet ice and 3 h at
The stability in the multisampler was evaluated by placing the processed sample plate in the multisampler set at 4°C for 50 or 96 h (required time to measure 1 or 2 plates) compared to urine processed and measured immediately. We also evaluated whether short-term storage of the extracted urine at -80°C (implying a freeze-thaw cycle) was more advantageous than leaving the urine in the multisampler queue for days. The stability to 1 vs. 3 freeze-thaw cycles was assessed for normalized urine samples by allowing urine samples stored at -80°C to thaw at 4°C for 4 h, which corresponds to 2 h of thawing and 2 h of processing time. Prior to the experiment, both the SQC and RQC samples experienced at least 2 additional freeze-thaw cycles. The storage time stability at -80°C during 2 and 10 months was compared between unprocessed urine, normalized urine and processed urine (extracted with acetonitrile).

RESULTS AND DISCUSSION

Workflow description. We present here an untargeted LC-MS-based metabolomics workflow for the automated processing of urine samples. The use of the 96-well plate format enables the sample preparation to be primarily performed by an automated liquid handing platform (Fig. 1). The workflow incorporates batch structures that can be used in both small and large studies, and contains multiple QC samples to monitor instrument performance across the analytical runs as well as for eventual signal correction (Fig. S1). Due to the large number of performance parameters and associated metabolites to evaluate, this workflow was developed and characterized using a small study of 87 urine samples. As proof-of-concept for application in larger cohorts, the workflow was then tested with a cohort of 842 urine samples.

A major advantage of working with urine is that it is available in large quantities and relatively easy to obtain. As part of the workflow proposed herein, we recommend a starting volume of 600 µL of urine (Fig. 1), of which 540 µL are transferred to 2 mL 96-well plates and 40 µL are used to create the SQC sample, which is aliquoted into 20 wells per plate (Fig. S1). To reduce freeze-thaw cycles, the 2 mL urine plates are then aliquoted into 4 plates (120 µL/well), of which 1 is used for SG measurements, 2 are dedicated to metabolomics measurements, and 1 serves as an eventual back-up. The urine normalization protocol results in a normalized urine volume of 340 µL per well (if the SG is 1.004-1.006, then normalized urine end volumes have to be reduced to 180 µL; if the SG is <1.004, then the final diluted well volume is 100 µL). Because the extraction protocol only uses a maximum of 30 µL of normalized urine per platform, the remaining urine can be stored as needed.
The method was developed with an analytical batch based upon a 96-well plate containing 70 samples, 20 SQCs, 2 RQCs and 4 blanks (Fig. S1). Given that the small cohort for method development was just over the maximal allowance for one plate, samples were equally divided between 2 plates and measured as a single batch. The injection sequence followed recommendations from Broadhurst et al.\textsuperscript{34} with samples interspaced with SQCs after each 5\textsuperscript{th} sample (Fig. S1, Table S11).

The gains in automating the sample preparation workflow can be described in terms of operator time (Table S12). Overall, the automation resulted in an almost 7-fold gain in person-time per plate (from 6.9 to 1 h), and decreased operator fatigue and related errors. For application in large-scale cohorts, for example processing 1050 samples in 15 batches, the total gain in time for the automated protocol is 88.6 h, equating to \textasciitilde11 days of full-time work.

\textbf{Urine normalization.} Normalization of the urinary solute concentration is a crucial step due to physiological fluctuations in the matrix composition.\textsuperscript{6} Multiple normalization strategies have been proposed including urinary SG,\textsuperscript{11} creatinine,\textsuperscript{12} and osmolality.\textsuperscript{13} Creatinine is widely used, but is susceptible to inter-individual characteristics including age, sex, diet and muscle cachexia.\textsuperscript{38, 39} The WADA has adopted SG due to its general applicability.\textsuperscript{14} In addition, given the time-consuming nature of normalizing the urine samples to a common dilution factor prior to LC-MS analysis, some efforts have applied either normalization of the MS signals post-acquisition\textsuperscript{13} or injection of variable urine amounts.\textsuperscript{9} The main issue of post-acquisition normalization is that there are significant matrix effects that stem from the high variability in the concentration and composition of urine. This can lead to ion suppression or enhancement as well as solid phase binding competition, which are all analyte-specific.\textsuperscript{6, 8} In contrast, targeted methods that employ internal standards generally provide good results with post-acquisition normalization.\textsuperscript{40} However, it is not possible to correct for these analyte-specific matrix effects in untargeted metabolomics. To demonstrate these effects, we measured a small cohort (n=87 samples) with no-normalization as well as pre- and post-acquisition urinary SG normalization (Fig S2).

The WADA protocol for measurement of urinary SG uses a hand-held refractometer, which converts the refractive index of urine to the corresponding SG.\textsuperscript{41, 42} We tested the reproducibility of refractometry SG readings between laboratories (in Japan and Sweden), as well as their stability after 5 freeze-thaw cycles (Fig. S3). The reproducibility of SG readings
between laboratories is good, with a mean bias of 0.0001 (limit of agreement, LOA: -0.0009–0.0012). The effects of 5 freeze-thaw cycles on urine SG readings translated to a small bias of 0.0008 (LOA: -0.0002–0.0019).

The measurement of SG using a hand-held refractometer is time-consuming for large sample numbers and is a significant bottleneck for automating metabolomics. We therefore developed an automated 96-well plate format method to measure RI using a multisampler connected to a RID. The RID measurements are converted to SG using a calibration curve consisting of NaCl, with SG measured with a refractometer. This method was validated with the UG-D refractometer model (available only in Japan) and the UG-alpha model (corresponding international version) that are based on a slightly different SG conversion.41, 42 Intra- and inter-day accuracy and precision were within the recommended ±15% thresholds (Table S13), except for the UG-D intra-day NaCl QC low (117%), which can be attributed to the 3-digit SG reading precision of this model. Comparison of SG measurements obtained from RID vs. the refractometers showed a mean bias of 0.0003 for the UG-D (LOA=−0.0007–0.0012; Fig. 2A) and a mean bias of -0.00003 for the UG-alpha (LOA=−0.0009–0.0008; Fig. S4). The RID measurements show a good agreement to both hand-held refractometer values, demonstrating the utility of the RID method. In addition, the automation of the SG measurements can increase biosafety due to decreased transfer of the urine by hand.

**Analytical methods and data pre-processing.** Given that urine is the main route for the elimination of water-soluble waste products, we propose here a combination of two analytical methods using HILIC chromatography. The HILIC method at pH=2.6 in positive ionization (ZHP) was adapted from previously published work24 and a new method was developed for negative ionization at pH=9.3 using a ZIC-pHILIC column (ZHN).

Features were annotated in MS-DIAL using accurate mass (AM) and retention time (RT) match, as well as spectral match (MS/MS) to our in-house libraries, and MS-FINDER 3.4243 for the identification of unknown compounds. For the ZHP dataset, of the raw 10,363 features, 406 features were annotated (including fragments and adducts), their integration checked and their CVs across SQCs and D-ratios calculated, as well as intensity plots across the injection sequence. After filtering for single-species and CV values, the final ZHP annotated dataset was comprised of 295 metabolites (Table S14), of which 126 were AM, RT and MS/MS matched, 73 were AM and RT matched, and 96 were AM matched with MS-FINDER. For the ZHN dataset, 465 out of 9332 features were annotated, resulting after filtering in 358 single metabolites (47 AMRT-MS/MS, 87 AMRT and 224 AM; Table S15).
Median absolute AM and RT differences were 0.3 mDa and 0.2 min for the ZHP method and 0.3 mDa and 0.03 min for the ZHN method.

Because the annotated dataset does not include unknown peaks, we processed an additional dataset at the feature level (a feature is defined as a pair of m/z and its corresponding RT). The all-feature dataset as exported from MZmine underwent several additional filtering steps in R, including: CVQC <35% (filtering out signals displaying low precision), D-ratio <55% (removing signals with low biological to technical variability ratio), and interquartile range (IQR) >0 (to eliminate features with low variability). In the ZHP untargeted dataset of 10,797 features, 632 (6%), 1320 (12%) and 33 (0.3%) entries were identified as falling outside the respective CV, D-ratio and IQR thresholds, while, for the 8,966 ZHN features, 503 (6%), 925 (10%) and 24 (0.2%) were outside CV, D-ratio and IQR thresholds.

**Metabolomic coverage of urine components.** To offer the broadest coverage of urine components by untargeted LC-MS methods, we chose two complementary HILIC methods in positive and negative ionization. The methods overlap for 109 compounds, while providing unique coverage of 186 and 249 compounds for ZHP and ZHN, respectively. Taken together, 544 unique metabolites were detected covering major chemical classes, including amino acids, nucleobases, lipids, vitamins and exogenous compounds (e.g., diet-derived, medicines) (Fig. 3). Both methods perform equally well for the majority of amino acids and their derivatives. However, the combination of two chromatographies and polarities enables the coverage of specific compound classes. The ZHP method offers an extensive coverage of acylcarnitines and betaines, while the ZHN method outperforms for the detection of lipids (SCFA, MCFA, steroids) and sugars as well as microbiota- and diet-derived compounds. The performance in the latter category is mostly due to phase II conjugated compounds, such as sulfated, glucuronides and other glycosylated species that account for >25% of the total annotations. These results stress the importance of phase II metabolic products in urine for characterizing dietary patterns. There are few available standards for this class of compounds, and instead enzymatic treatment combined with MS methods followed by custom synthesis have been proposed as strategy to rapidly identify these metabolites. Accordingly, the untargeted ZHN platform developed for the current study is useful for providing a snapshot of these metabolic processes as well as promoting structure characterization and long-term identification efforts.
We evaluated our coverage of urinary metabolites by comparing our annotations to the current Urine HMDB repository (4364 entries as of 2020-09-10).\(^5\) We observed an overlap of 286 compounds, while our methods include 258 compounds not yet reported in Urine HMDB. As this repository provides concentration data, we also compared our annotation list with a subset of 1199 compounds with available urine concentrations in healthy adults. Of the 286 compounds overlapping with our annotations, 90% had mean concentrations >0.1 µmol/mmol creatinine, which could be taken as an overall estimation of our platform sensitivity.

We also assessed our coverage compared to two other recent metabolomics platforms: one offering a coverage of 142 urinary compounds, of which half can be quantified,\(^4\)\(^5\) and another targeting exposome research with a wide coverage of endogenous and exogenous compounds quantifying 690 compounds for a total of 1022 annotations.\(^4\)\(^6\) Of the first study, only 60 compounds overlap with our annotations, due to the strong focus on urinary lipids that are for most only present in low abundance and for which targeted MS/MS is necessary. When comparing our coverage with the exposome platform, we observed an overlap of 222 compounds, with 318 uniquely reported by our methods. Here again, the use of a targeted MS/MS approach, which is well-suited to detect low abundant compounds, can explain the relatively small annotation overlap. However, by definition, targeted approaches can only focus on already known compounds, while our untargeted methods can perform de novo annotations at the MS\(^2\) level. Given that our general sensitivity threshold is ~0.1 µmol/mmol creatinine, it is likely that the 258 compounds detected by our method, but not yet reported in urine are in fact present in relatively high abundance. These findings support the use of untargeted metabolomics to map unknown components in urine.

**Precision of the workflow.** In contrast to targeted methods, untargeted metabolomics aims to measure hundreds to thousands of potentially unknown metabolites. It is not possible to assess the data quality using the standard targeted metrics of accuracy and precision for each feature. To address this issue, we implemented technical internal standards (tIS) to monitor analytical performance. These selected standards are exogenous compounds that are not generally found in biofluids, are readily available, and are affordable (Table S1, Fig. S5).\(^4\)\(^7\) In automated liquid handling systems, the extraction solvent containing the tIS is dispensed from a reservoir containing excess volume, justifying the need for affordable standards. Moreover, isotopically labelled metabolites might interfere with deconvolution of untargeted
data, complicating spectral information. The use of tIS enables evaluation of the LC-MS system prior to initiating an analytical run and provides a means to monitor the precision of the measurements within and between analytical batches across the entire LC-MS workflow. For example, shifts in retention time, injection volume inconsistencies or mass spectrometer ionization issues can rapidly be identified and the potential impact on the measurements assessed. Based upon our experience with both platforms, we set acceptance criteria for the analytical run to be CVs in the SQCs <10% and CVs across samples <20%. When measuring the small cohort, we observed good compliance with the acceptance criteria with tISs displaying CVs <5% and <8% in the SQCs and samples, respectively for the ZHP method, while for the ZHN method, the CVs were <11% and <20% in the SQCs and samples, respectively (Table 1, Fig. S5). For the 290 annotated single species metabolites measured in the ZHP platform prior to drift correction, 231 have CV ≤10%, with only 12 metabolites above the 20% threshold set by Klavus et al. (Fig. 4). Similarly, out of 354 annotated metabolites in the ZHN platform, 296 display CV ≤10% and only 2 have CV >20% In the raw all-feature ZHP and ZHN datasets also, >80% of the features display CVQC ≤20%.

**Metabolite stability within the workflow.** The stability of the metabolites was assessed for the entire sample preparation protocol and LC-MS measurement. In the transfer from the manual to automated workflow, the samples could no longer be processed on ice. We accordingly complemented the liquid handling system with a cooling unit and metal inserts for the 96-well plates. This provided cooling down to 6°C in plates on metal blocks and <10°C for the reservoirs. All manipulations by the liquid handler can be performed within 10 min, except for the urine normalization, which requires a maximum of 3 h for a full plate. We therefore assessed the stability of urine samples on the cooling deck of the liquid handler during 3 h compared to incubation on wet ice or at room temperature for raw and normalized SQC and RQC urine samples. Of the >200 metabolites that could be reliably observed in these two urines, <6% were affected by any of the conditions and displayed an increase >20% (Table S16, Fig. S6).

Another potential issue affecting metabolite stability was the time spent in the multisampler (set at 4°C). In our setup, the temperature in the bottom compartment is <5°C, while the upper compartment generally does not achieve temperatures <8°C. Less than 1% of the metabolites monitored showed a change in intensity after 50 or 96 h at 5°C. However, short-term storage of extracted urine at -80°C is not advised; this additional freeze-thaw led
to a decrease of >20% in intensity for 40% of the metabolites in the RQC, while the SQC urine had only 12% of its metabolites impacted. During the sample processing, there is the option to freeze normalized urine and/or extracted urine for later use. We therefore investigated the stability of metabolites stored in normalized or extracted urine and observed minimal differences (<11% of the metabolites affected) between unprocessed, normalized and extracted urine when stored for 2 weeks, and up to 10 months. Finally, we also demonstrated stability of normalized urine after several freeze-thaw cycles (>85% are stable), which indicates that normalized urine can be extracted following storage if necessary (Table S16, Fig. S6).

**Performance in a large cohort.** We tested the applicability of our workflow for a large cohort (n=842 samples). We first evaluated SG measurements performed with the developed RID method compared to measurements with the UG-D refractometer. The method showed excellent agreement between the refractometer and the RID, with a mean deviation of -0.0001 and LOA=-0.0014–0.0011 (Fig. 2B). These findings demonstrate the suitability of the RID SG measurement method for large sets of urine samples. In addition, this RID method might be of use for normalizing other less common biofluids. Initial trials with saliva were encouraging (data not shown).

In the case of large study sizes, on the scale of >1000 samples, analytical measurements can last several weeks and the probability of encountering technical issues significantly increases. As proof-of-concept for the application of the present workflow for large cohorts, we provide in Table 1 CVs for 842 urine samples measured in 24 batches (1090 urine samples in total, including 248 SQCs).

For the ZHP platform, 4 tISs (pyrantel, PIPES, CHES and HEPES) met the acceptance criteria of CV_{QCs} <10% and CV_{samples} <20% across all 24 batches. Fluorocytosine did not meet the criteria in 8 plates for the CV_{QCs} and the CV_{samples}. This tIS elutes in a busy RT window and is injected at >5 times lower concentration than the other tIS due to a strong ionization. For the ZHN platform, 4 tISs (fluorocinnamic acid, CHES, PIPES and HEPES) also displayed CVs generally below set thresholds (fluorocinnamic acid failed in one plate for CV_{QCs}). The 5th tIS (tricarballylic acid) evidenced a higher variability, especially in the CV_{samples}. This tIS elutes at the far end of the gradient, where the percentage of water in the mobile phase is high and desolvation more difficult. Also, it elutes closely to two abundant metabolites (citric acid and ascorbic acid sulfate) that interfere with the signal. While fluorocytosine and tricarballylic acid demonstrated high variability, they are still of utility.
because they demonstrate potential performance issues at those RT in the chromatograms and can be used as anchors for RT correction algorithms. The remaining tISs display high precision at the large cohort scale. In the case when a tIS exceeds threshold values, a more thorough investigation is performed. For example, in one plate, CHES showed a slightly higher CV. By plotting the intensities across injections, we observed two samples with markedly lower CHES abundance (Fig. S7). The 3D plots revealed a highly abundant compound that eluted close to CHES, which was identified as the antibiotic trimethoprim and was suppressing the CHES signal. Therefore, tIS performance should be carefully assessed depending on the context of the urinary cohort, especially when the urine sample strongly deviates from the usual composition. The performance of these tISs also demonstrates the analytical challenge of untargeted metabolomics.

After drift correction, we report CV\textsubscript{QCs} <5\% and CV\textsubscript{samples} <16\% for 4/5 tIS for the entire dataset of 1090 chromatograms in both ZHP and ZHN. In a recent targeted metabolomics study with 690 reference standards,\textsuperscript{46} the authors report a mean intra-day precision of 6.9\% (range: 0.1-16.1\%) as calculated in 5 pooled QCs, which are close to our CV\textsubscript{QCs} values at the single batch level, and mean inter-day of 10.9\% (1.5-19.6\%), markedly higher than our mean post-correction tIS CV\textsubscript{QCs}. In another targeted platform used on 1800 samples and measuring 142 urinary metabolites,\textsuperscript{45} intra-day CVs ranged from 0.23\% to 8.34\% and inter-day CVs from 0.77\% to 12.8\% in triplicate pooled QCs, which are similar to our reported tIS CVs. Also, considering that >75\% of our raw peak intensities show CV\textsubscript{QCs} \leq 10\% in both annotated datasets and >82\% of the raw features have CV\textsubscript{QCs} \leq 20\%, our untargeted workflow provides a precision level similar to recent targeted metabolomics platforms.

**CONCLUSIONS**

The workflow presented here offers an automated platform for urinary profiling in metabolomic epidemiology. The adaptation to a 96-well plate format enabled automation while increasing the potential for high-throughput. The implementation of a novel RID method for high-throughput measurement of urinary SG offered significant savings in operator time, while also developing a method that can be of utility for measuring SG in non-metabolomics studies. The untargeted methods cover a wide range of known urinary metabolites, while enabling the discovery of novel compounds. We also provide metrics for the quality assessment of analytical runs. There are however limitations in the method including the variability in metabolite RT inherent to HILIC columns and time-consuming
data processing. In addition, in spite of the excellent performance of 4/5 tISs in a large cohort, there is still variability associated with the performance of some metabolites in both ZHP and ZHN, highlighting that data from untargeted metabolomics require confirmation with targeted methods using the appropriate internal standards. Taken together, the developed platform offers the potential for automated urinary metabolomic epidemiology analyses with high precision. This will be useful for performing large-scale studies and molecular phenotyping efforts as part of a personalized medicine strategy.

Supporting Information
a) Supplemental Experimental; b) Supplemental Results and Discussion; c) Tables S−1 to S−16; d) Figures S−1 to S−8.

Author Contributions
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

ACKNOWLEDGMENT
We thank Cristina Gómez and Yusuke Jikumaru for discussions and suggestions. We acknowledge support from the Gunma University Initiative for Advanced Research (GIAR), the Japan Society for the Promotion of Science (JSPS) (19K17662), the Swedish Heart Lung Foundation (HLF 20170734, HLF 20170603), the Swedish Research Council (2016-02798), the Swedish Respiratory Society, the Swedish Asthma and Allergy Foundation and through the Regional Agreement on Medical Training and Clinical Research (ALF) between Stockholm County Council and Karolinska Institutet. This work was supported in part by The Environment Research and Technology Development Fund (ERTDF) (Grant No 5-1752) and by Japan-Sweden Research Cooperative Program between JSPS and STINT (Grant No JPJSBP-1201854). IM was supported by JSPS postdoctoral fellowship (P17774).
REFERENCES

(1) Karczewski, K. J.; Snyder, M. P. Nature Reviews in Genetics 2018, 19 (5), 299-310.
(2) Lasky-Su, J.; Kelly, R. S.; Wheelock, C. E.; Broadhurst, D.; Zanetti, K. 2021, submitted.
(3) Putnam, D., Composition and Concentrative Properties of Human Urine. 1971.
(4) Mayo Clinic Laboratories - Test Catalogue. https://www.mayocliniclabs.com/test-catalog/clinical-and-interpretive/9318.
(5) Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorndahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P., et al. PLoS One 2013, 8 (9), e73076.
(6) Edmands, W. M.; Ferrari, P.; Scalbert, A. Anal Chem 2014, 86 (21), 10925-31.
(7) Elmsjö, A.; Haglöf, J.; Engskog, M. K. R.; Erngren, I.; Nestor, M.; Arvidsson, T.; Pettersson, C. J. Chromatogr. A 2018, 1568, 49-56.
(8) Chamberlain, C. A.; Rubio, V. Y.; Garrett, T. J. Metabolomics 2019, 15 (10), 1-9.
(9) Chen, Y.; Shen, G.; Zhang, R.; He, J.; Zhang, Y.; Xu, J.; Yang, W.; Chen, X.; Song, Y.; Abliz, Z. Anal Chem 2013, 85 (16), 7659-65.
(10) Chetwynd, A. J.; Abdul-Sada, A.; Holt, S. G.; Hill, E. M. J. Chromatogr. A 2016, 1431, 103-110.
(11) Haddow, J. E.; Knight, G. J.; Palomaki, G. E.; Neveux, L. M.; Chilmonczyk, B. A. Clin Chem 1994, 40 (4), 562-4.
(12) Boeniger, M. F.; Lowry, L. K.; Rosenberg, J. Am Industrial Hyg Assoc J 1993, 54 (10), 615-627.
(13) Warrack, B. M.; Hnatyshyn, S.; Ott, K.-h.; Reily, M. D.; Sanders, M.; Zhang, H.; Drexler, D. M. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 2009, 877, 547-552.
(14) World Anti-Doping Agency (WADA) ISTI: Urine Sample Collection Guidelines. https://www.wada-ama.org/sites/default/files/resources/files/wada_guidelines_urene_sample_collection_2014_v1_0_en.pdf.
(15) Chadha, V.; Garg, U.; Alon, U. S. Pediatr Nephrol 2001, 16 (4), 374-82.
(16) George, J. W. Vet Clin Pathol 2001, 30 (4), 201-210.
(17) Zhang, T.; Creek, D. J.; Barrett, M. P.; Blackburn, G.; Watson, D. G. Anal Chem 2012, 84 (4), 1994-2001.
(18) Tada, I.; Tsugawa, H.; Meister, I.; Zhang, P.; Shu, R.; Katsumi, R.; Wheelock, C. E.; Arita, M.; Chaleckis, R. Metabolites 2019, 9 (11).
(19) Um-Bergstrom, P.; Hallberg, J.; Pourbazargan, M.; Berggren-Brostrom, E.; Ferrara, G.; Eriksson, M. J.; Nyren, S.; Gao, J.; Lilja, G.; Linden, A., et al. Respir Res 2019, 20 (1), 102.
(20) Sinha, A.; Lutter, R.; Xu, B.; Dekker, T.; Dierdorp, B.; Sterk, P. J.; Frey, U.; Delgado-Eckert, E. eLife 2019, 8, e47969.
(21) US Food and Drug Administration, Bioanalytical Method Validation: Guidance for Industry. https://www.fda.gov/files/drugs/published/Bioanalytical-Method-Validation-Guidance-for-Industry.pdf (accessed - May).
(22) R Development Core Team R: A language and environment for statistical computing, Viena, Austria, 2010.
(23) Levine, L.; Fahy, J. P. J. Ind Hyg Toxicol 1946, 28 (3), 98-99.
(24) Naz, S.; Gallart-Ayala, H.; Reinke, S. N.; Mathon, C.; Blankley, R.; Chaleckis, R.; Wheelock, C. E. Anal Chem 2017, 89 (15), 7933-7942.
(25) Chaleckis, R.; Naz, S.; Meister, I.; Wheelock, C. E., LC-MS-Based Metabolomics of Biofluids Using All-Ion Fragmentation (AIF) Acquisition. In Methods in Molecular Biology, Giera, M., Ed. Springer Nature: 2018; Vol. 1730, pp 45-58.
(26) Haug, K.; Cochrane, K.; Nainala, V. C.; Williams, M.; Chang, J.; Jayaseelan, K. V.; O'Donovan, C. Nucleic Acids Res 2020, 48 (D1), D440-D444.
(27) Stanstrup, J.; Neumann, S.; Vrhovsek, U. Anal Chem 2015, 87 (18), 9421-8.
(28) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J., et al. Nat Biotechnol 2012, 30 (10), 918-20.
(29) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. Bmc Bioinformatics 2010, 11.
(30) Tsugawa, H.; Ikeda, K.; Takahashi, M.; Satoh, A.; Mori, Y.; Uchino, H.; Okahashi, N.; Yamada, Y.; Tada, I.; Bonini, P., et al. Nat Biotechnol 2020, 38 (10), 1159-1163.
(31) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; Vandergheynst, J.; Fiehn, O.; Arita, M. Nat Metho 2015, 12 (6), 523-526.
(32) Tada, I.; Chaleckis, R.; Tsugawa, H.; Meister, I.; Zhang, P.; Lazarinis, N.; Dahlén, B.; Wheelock, C. E.; Arita, M. Anal Chem 2020, 92 (16), 11310-11317.
(33) Cajka, T.; Fiehn, O. Methods Mol Biol 2017, 1609, 149-170.
(34) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. Metabolomics 2018, 14 (6), 72.
(35) Myers, O. D.; Sumner, S. J.; Li, S.; Barnes, S.; Du, X. Anal Chem 2017, 89 (17), 8696-8703.
(36) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N., et al. Nat Protocols 2011, 6 (7), 1060-1083.
(37) Saigusa, D.; Okamura, Y.; Motoike, I. N.; Katoh, Y.; Kurosawa, Y.; Sajiyo, R.; Koshiba, S.; Yasuda, J.; Motohashi, H.; Sugawara, J., et al. PLoS One 2016, 11 (8), e0160555.
(38) Cook, J. D.; Caplan, Y. H.; LoDico, C. P.; Bush, D. M. J. Anal Toxicol 2000, 24 (7), 579-588.
(39) Miller, R. C.; Brindle, E.; Holman, D. J.; Shofer, J.; Klein, N. A.; Soules, M. R.; O'Connor, K. A. Clin Chem 2004, 50 (5), 924-32.
(40) Gomez, C.; Gonzalez-Riano, C.; Barbas, C.; Kolmert, J.; Hyung Ryu, M.; Carlsten, C.; Dahlen, S. E.; Wheelock, C. E. J Lipid Res 2019, 60 (6), 1164-1173.
(41) Wolf, A. V., Aqueous Solutions and Body Fluids: Their Concentrative Properties and Conversion Tables. Harper and Row: New York, 1966; p 182.
(42) Standards, U. S. G. S. o. t. J. C. f. C. P. Japanese J. Clin. Pathol. 1979, 27 (11), 1026-1032.
(43) Tsugawa, H.; Nakabayashi, R.; Mori, T.; Yamada, Y.; Takahashi, M.; Rai, A.; Sugiyama, R.; Yamamoto, H.; Nakaya, T.; Yamazaki, M., et al. Nat Methods 2019, 16 (4), 295-298.
(44) Ballet, C.; Correia, M. S. P.; Conway, L. P.; Locher, T. L.; Lehmann, L. C.; Garg, N.; Vujasinovic, M.; Deindl, S.; Lohr, J. M.; Globisch, D. Chem Sci 2018, 9 (29), 6233-6239.
(45) Zheng, J.; Zhang, L.; Johnson, M.; Mandal, R.; Wishart, D. S. Anal Chem 2020, 92 (15), 10627-10634.
(46) González-Domínguez, R.; Jáuregui, O.; Isabel Queipo-Ortuño, M.; Andres-Lacueva, C. Anal Chem 2020, 92 (20), 13767-13775.
(47) Pluskal, T.; Hayashi, T.; Saitoh, S.; Fujisawa, A.; Yanagida, M. FEBS J 2011, 278 (8), 1299-315.
(48) Klávus, A.; Kokla, M.; Noerman, S.; Koistinen, V. M.; Tuomainen, M.; Zarei, I.; Meuronen, T.; Häkkinen, M. R.; Rummukainen, S.; Babu, A. F., et al. Metabolites 2020, 10 (4), 135.
Table 1 Performance of technical internal standards (tIS) at the large cohort scale*

| Standard       | Raw data | Raw data | Pre QC-correction | Post QC-correction |
|----------------|----------|----------|-------------------|--------------------|
|                |          | Small cohort | Large cohort |                 |                    |
| Pyrantel       | CV_{QC} / CV_{sample} | Mean CV_{QC} (min-max) | Mean CV_{sample} (min-max) | CV_{QC} / CV_{sample} | CV_{QC} / CV_{sample} |
| ZHP            | 3.0 / 2.5 | 2.7 (1.0 - 4.8) | 3.7 (1.5-6.8) | 23.0 / 23.0 | 2.3 / 3.7 |
| CHES           | 3.7 / 6.2 | 3.6 (1.4 - 6.9) | 6.8 (4.1-16.0) | 22.0 / 23.0 | 4.5 / 9.1 |
| Fluorocytosine | 3.8 / 7.0 | 8.5 (2.3 - 19.0) | 16.0 (5.5-31.0) | 40.0 / 40.0 | 19.0 / 19.0 |
| PIPES          | 4.0 / 3.5 | 4.4 (1.7 - 9.2) | 6.7 (3.1-13.0) | 20.0 / 19.0 | 4.3 / 8.8 |
| HEPES          | 4.4 / 3.1 | 4.5 (1.5 - 9.2) | 6.2 (1.9-15.0) | 20.0 / 21.0 | 3.8 / 7.0 |
| ZHN            | Raw data | Raw data | Pre QC-correction | Post QC-correction |
|                |          | Small cohort | Large cohort |                 |                    |
| Fluorocinnamic acid | CV_{QC} / CV_{sample} | Mean CV_{QC} (min-max) | Mean CV_{sample} (min-max) | CV_{QC} / CV_{sample} | CV_{QC} / CV_{sample} |
| CHES           | 10.0 / 20.0 | 5.9 (2.8 – 12.0) | 12.5 (5.9 - 18.0) | 15.0 / 23.0 | 2.5 / 16.0 |
| HEPES          | 4.1 / 6.3 | 4.4 (2.7 - 7.0) | 6.9 (4.9 - 11.0) | 12.0 / 16.0 | 2.8 / 8.9 |
| Pipes          | 7.5 / 8.2 | 5.3 (1.5 - 7.9) | 9.4 (6.7 - 12) | 19.0 / 20.0 | 3.6 / 11.0 |
| Tricarballylic acid | 5.7 / 10.0 | 4.7 (2.4 - 7.4) | 10.0 (6.2 - 16.0) | 25.0 / 27.0 | 3.2 / 12.0 |
|                | 5.6 / 18.0 | 12.4 (6.8 - 18.0) | 36.5 (14.0 – 66.0) | 32.0 / 54.0 | 10.0 / 44.0 |

*Coefficient of variation (CVs) of the peak area across 24 plates (n=842 samples, n=248 SQCs). Raw data are CV values per plate, while pre- and post-QC correction CVs are calculated at the whole cohort scale before and after applying the QC correction algorithm. SQCs=pooled study QCs; ZHP=ZIC-HILIC positive ionization mode; ZHN=ZIC-pHILIC negative ionization mode.
Figure 1. Urinary untargeted metabolomics workflow. Manual steps are framed in orange and automated steps are shown in blue. QC=quality control; RI=refractive index; SG=specific gravity; POS=positive ionization; NEG=negative ionization; DIA=data independent acquisition.
Figure 2. Bland-Altman plots of urine specific gravity (SG) measured with hand-held refractometer (model UG-D) vs. refractive index detector (RID). A) The small cohort (n=87). B) The large cohort (n=842). Each sample is represented by a blue diamond. Mean deviation is shown as a solid red line with the 95% confidence intervals (limits of agreement) as red dotted lines.
Figure 3. **Metabolic coverage and annotation confidence levels.** Annotated compounds are provided based upon chemical class. Shades of red indicate metabolites from the ZHP platform. Shades of blue indicate metabolites from the ZHN platform. The darkest shades indicate compounds that matched to the in-house library with accurate mass (AM), retention time (RT) and MS/MS spectrum; the mid-level shades indicate compounds that matched to the in-house library, but lack a spectral match; the lightest shades indicate compounds that are identified with AM from external databases (most of the time also using *in-silico* spectral match). ZHP=ZIC-HILIC positive ionization; ZHN=ZIC-pHILIC negative ionization.
Figure 4. Bar plot of metabolite or feature coefficients of variation (CV) in each of the platforms. A) Annotated metabolite datasets with ZHP platform represented by blue bars (n=295 metabolites) and ZHN platform by red bars (n=358); B) All-feature datasets (n=10,795 features in ZHP and n=8,961 in ZHN). Data are from the 22 pooled study quality control samples (SQC) from the small cohort. ZHP=ZIC-HILIC positive ionization; ZHN=ZIC-pHILIC negative ionization.
Supplementary Information

High-precision automated workflow for urinary untargeted metabolomic epidemiology

Isabel Meister\textsuperscript{1,2}, Pei Zhang\textsuperscript{1,2,•}, Anirban Sinha\textsuperscript{3,4}, C. Magnus Sköld\textsuperscript{5}, Åsa M. Wheelock\textsuperscript{5}, Takashi Izumi\textsuperscript{1,6}, Romanas Chaleckis\textsuperscript{1,2,•,‡}, Craig E. Wheelock\textsuperscript{1,2,•,‡}

\textsuperscript{1} Gunma University Initiative for Advanced Research (GIAR), Gunma University, Maebashi, Gunma, Japan

\textsuperscript{2} Division of Physiological Chemistry 2, Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden

\textsuperscript{3} Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

\textsuperscript{4} Department of Experimental Immunology, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

\textsuperscript{5} Division of Respiratory Medicine and Allergy, Department of Medicine, Karolinska Institutet, Stockholm, Sweden

\textsuperscript{6} Department of Biochemistry, Gunma University Graduate School of Medicine, Maebashi, Gunma, Japan

\textsuperscript{•} Current address: Key Laboratory of Drug Quality Control and Pharmacovigilance (Ministry of Education), State Key Laboratory of Natural Medicine, China Pharmaceutical University, Nanjing 210009, China

\textbf{Corresponding Authors:}

Romanas Chaleckis
Gunma University Initiative for Advanced Research (GIAR),
Gunma University,
Gunma, Maebashi, Japan.
Email: romcha@gunma-u.ac.jp

Craig E Wheelock
Division of Physiological Chemistry 2,
Department of Medical Biochemistry and Biophysics,
Karolinska Institute,
Stockholm, Sweden.
Email: craig.wheelock@ki.se
EXPERIMENTAL SECTION

Technical internal standards (tISs)
The tIS concentrations were optimized to achieve intensities 100-300-fold higher than the system noise level (1×10^3) and between 10-30-fold lower than the saturation level in the LC-MS system (3×10^6).

Data quality check
For the data quality check, raw data was first converted to mzML format using Proteo Wizard before loading into MZmine. Extracted ion chromatograms (EICs) of the lockmasses (Table S5) were obtained with 20 ppm tolerance. The EICs of the 5 tIS and 25 selected metabolites for the data quality check (Table S7) were extracted by targeted peak detection (Table S6). The quality of the dataset was assessed by examining the following parameters: binary and isopump pressures, overlaid total ion chromatograms (TICs) for all samples and for SQCs only, EICs of lockmasses for all samples and for SQCs only, 3D views (x=RT, y=intensity, z=m/z) of one SQC sample and the reference blank sample, EICs of each tIS and calculations of their CVs in the SQCs and in the samples. Recommended acceptance criteria for the analysis sequence are: binary pump pressure profile drift <10 bar across acquisition, presence of the lockmass signals above threshold in all spectra, baseline overlay of TICs for all the samples, excellent overlay of TICs in the SQC samples (baseline overlay without retention time shifts or outlier peaks) and tIS EICs in both samples and QCs, 3D views without signs of mobile phase contamination (unusual shark fins) and CVs of tIS ≤10% in the SQCs and ≤20% in the samples. Failure to meet these criteria does not necessarily lead to the rejection of the run, but rather flags the plate for detailed troubleshooting and careful evaluation as to whether the identified problem compromises the run quality.

Data pre-processing
Unidentified peaks, especially features among the 500 most intense peaks, were tentatively annotated using AM (<10 mDa tolerance) corresponding to MSI level 3. Supporting information for the annotation of the metabolites was at least 4 matching peaks with in silico fragmentation patterns in MS-FINDER, correlation with related compounds in the dataset, previous reporting of the metabolite in urine, RT relative to structurally similar known compounds, presence of specific fragments, etc.

All the annotated metabolite peak alignments and integrations were individually curated before exporting as a text file. Peak area and height intensities were plotted across the injection sequence for visual inspection using R ggplot2 scripts. As one metabolite can produce several signals (fragments or adducts), the annotated peak list was curated to have only one signal per metabolite. The molecular ion ([M+H]^+ or [M-H]^-) was generally preferred for choosing the representative molecular species; however, in some cases, a fragment or adduct displaying marked improved analytical performances was selected instead (e.g., creatinine Fig. S8).

RESULTS AND DISCUSSION

Sample preparation
The thawed urine samples are not centrifuged prior to aliquoting. This decision was made based upon the fact that there is no general harmonization on the selection of urine collection tubes. Accordingly, tubes are provided from clinical collaborators in various brands and shapes, and often in larger containers that do not fit available rotors of the centrifuges. To
avoid time-consuming manual transfers to centrifugeable tubes, we opted to move the urine clean-up centrifugation step to after aliquoting. In addition, because aliquot plates are frozen at -80°C, thawing will release debris that must be removed. Given that the original samples are not centrifuged prior to transfer to 96-well plates, the amount of particulates between the aliquots may vary; a centrifugation step after thawing is therefore necessary to ensure aliquot homogeneity.

**Analytical methods**

The ZHP method was optimized from the published version in terms of injection volume and ionization settings. The injection volume was decreased from 2 μL to 0.8 μL to reduce saturation of highly abundant compounds (e.g., creatinine). Saturation can result in a mass shift that may translate into a split chromatogram, and interfere with the alignment of untargeted data resulting in artifact peaks and low-quality data.

The ZHN stationary phase was selected after comparison between ZIC-pHILIC and the BEH amide (Waters), the former providing a slightly better performance for acidic amino acids, sugars and keto-acids, while the latter provided a better retention of steroids. We chose the ZIC-pHILIC with the objective of providing the widest chemical coverage. Switching the mobile phase A buffer from ammonium carbonate, as in several previous publication that use high pH with ZIC-pHILIC,11,12 to ammonium acetate and maintaining a low buffer concentration of 5mM enabled the partial rescue of steroid retention and increased performance for tricarboxylic acids. The method was then further optimized in terms of gradient and run time. The starting gradient was set to 88% B, which revealed the best compromise between low organic content that would impact peak shapes of carboxylic acids and high organic content that would affect steroid peak shape and retention. An isocratic start did not lead to improvements and increased run length time. Gradient steepness was also optimized to avoid a high level of co-elution while optimizing run time and peak shapes of later eluting compounds, which were further improved with the addition of a steeper sequence at the end of the gradient. Finally, a column wash with 25% B was necessary to avoid long-term column clogging, thus calling for a reequilibration time of 10.5 min to achieve a reproducible baseline.

The pH reproducibility of the ZHN method was evaluated by comparing the pH of mobile A solutions prepared on 4 different days. Values differed by a maximum of 0.08 pH units (9.36-9.42). As basic mobile phases are prone to absorb ambient CO₂, which can consequently affect the pH, we evaluated the pH stability of the ZHN mobile phase A stored at room temperature vs. at 4°C, and did not observe any variation after 7 days. However, care should be taken to have proper mobile phase cap fittings.

Another challenge in the automated workflow is the centrifugation force and duration. While common centrifuges with a fixed angle rotor for 1.5 mL tubes (e.g., Eppendorf 5430) can achieve centrifugation forces >20,000 g, plate centrifuges with swing rotors are limited to a maximum of 3000-4000 g. This limitation can be partially compensated for with longer centrifugation time. We observed differences in metabolite profiles for centrifugation times between 10 and 40 min, while 40 min and 2 h did not evidence major shifts. We therefore opted for 40 min centrifugation time. Although plate centrifuges are not able to offer the same pelleting efficiency as fixed angle rotor centrifuges, the supernatant is sufficiently clean to achieve reproducibility and avoid clogging in the LC system.
REFERENCES

(1) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J., et al. Nat Biotechnol 2012, 30 (10), 918-20.

(2) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. Bmc Bioinformatics 2010, 11.

(3) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; Vanderheynst, J.; Fiehn, O.; Arita, M. Nat Methods 2015, 12 (6), 523-526.

(4) Tada, I.; Chaleckis, R.; Tsugawa, H.; Meister, I.; Zhang, P.; Lazarinis, N.; Dahlén, B.; Wheelock, C. E.; Arita, M. Anal Chem 2020, 92 (16), 11310-11317.

(5) Tada, I.; Tsugawa, H.; Meister, I.; Zhang, P.; Shu, R.; Katsumi, R.; Wheelock, C. E.; Arita, M.; Chaleckis, R. Metabolites 2019, 9 (11).

(6) Fiehn, O.; Robertson, D.; Griffin, J.; van der Werf, M.; Nikolau, B.; Morrison, N.; Sumner, L. W.; Goodacre, R.; Hardy, N. W.; Taylor, C., et al. Metabolomics 2007, 3 (3), 175-178.

(7) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Environ Sci Technol 2014, 48 (4), 2097-8.

(8) R Development Core Team R: A language and environment for statistical computing, Viena, Austria, 2010.

(9) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. Metabolomics 2018, 14 (6), 72.

(10) Myers, O. D.; Sumner, S. J.; Li, S.; Barnes, S.; Du, X. Anal Chem 2017, 89 (17), 8696-8703.

(11) Gallart-Ayala, H.; Konz, I.; Mehl, F.; Teav, T.; Oikonomidi, A.; Peyratout, G.; van der Velpen, V.; Popp, J.; Ivanisevic, J. Anal Chim Acta 2018, 1037, 327-337.

(12) Zhang, T.; Creek, D. J.; Barrett, M. P.; Blackburn, G.; Watson, D. G. Anal Chem 2012, 84 (4), 1994-2001.

(13) Liu, K. H.; Nellis, M.; Uppal, K.; Ma, C.; Tran, V.; Liang, Y.; Walker, D. I.; Jones, D. P. Anal Chem 2020, 92 (13), 8836-8844.

(14) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N., et al. Nat Protocols 2011, 6 (7), 1060-1083.

(15) Sindelar, M.; Patti, G. J Am Chem Soc 2020, 142 (20), 9097-9105.

(16) Elmsjö, A.; Haglöf, J.; Engskog, M. K. R.; Ern gren, I.; Nestor, M.; Arvidsson, T.; Pettersson, C. J. Chromatogr. A 2018, 1568, 49-56.

(17) Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J. Nucleic Acids Res 2018, 46 (W1), W486-W494.

(18) Klávus, A.; Kokla, M.; Noerman, S.; Koistinen, V. M.; Tuomainen, M.; Zarei, I.; Meuronen, T.; Häkkinen, M. R.; Rummukainen, S.; Babu, A. F., et al. Metabolites 2020, 10 (4), 135.
Figure S1. Analytical batch layout and sample organization on the 96-well plate.

SQC = pooled study quality control sample (pooled mix of all study samples);
condQC = quality control samples for column conditioning (usually additional SQC samples);
RQC = reference quality control sample to be included in all studies (prepared from laboratory reference urine); tISmix = technical internal standard mix.
Figure S2. Effects of urine normalization upon the multivariate data structure. Principal components analysis (PCA) score plots of the 87-sample LUNAPRE cohort measured A) without any normalization prior to acquisition of LC-MS data, B) urinary normalization post-acquisition of LC-MS data, and C) urinary normalization pre-acquisition of LC-MS data. Normalization was performed using the measured urinary specific gravity (SG) of each sample. LC-MS data were acquired using the methods described in the main text. Metabolites were annotated at the all-feature level and data were log transformed and UV scaled for inclusion in the PCA model. PCA plots were generated using SIMCA v15 (Satorius AG, Göttingen, German). PCA model metrics are as follows: (A) $R^2_X=0.648$, $Q^2=0.464$; (B) $R^2_X=0.606$, $Q^2=0.303$; (C) $R^2_X=0.557$, $Q^2=0.282$. 
Figure S3. Bland-Altman plots of inter-laboratory refractometry specific gravity (SG) readings (A) and freeze-thaw cycle consistency (B). Each sample is represented by a blue lozenge (n=87), mean deviation as a solid red line, and 95% confidence intervals (limits of agreement, LOA) with red dotted lines. A) SG readings from the refractometer model UG-alpha between 2 different laboratories (Lab 1=Japan, Lab 2=Sweden). B) SG readings from the refractometer model UG-D of samples following 5 additional freeze-thaw cycles compared to controls.
Figure S4. Bland-Altman plots of refractometry specific gravity (SG) readings (model UG-alpha) with refractive index (RI) units (n=87). Each sample is represented by a blue lozenge, mean deviation as a solid red line, and 95% confidence intervals (limits of agreement, LOA) with red dotted lines. RI units were measured on the Agilent refractive index detector (RID) as described in the main text.
Figure S5. Overlaid chromatograms of the technical internal standards (tIS) across all samples and CVs of the corresponding peak areas. The CVs are shown from both the primary LUNAPRE study samples (n=87) as well as the pooled study quality control (SQC) samples (n=22). ZHP=ZIC-HILIC positive ionization; ZHN=ZIC-pHILIC negative ionization.
Figure S6. Stability of the ZHP platform metabolites across the reported workflow. A) The temperature-dependent stability of the observed metabolites in raw urine during the sample preparation. B) The temperature-dependent stability of the observed metabolites in normalized urine during the sample preparation. C) The effect of multiple freeze-thaw cycles on metabolite stability in normalized urine. D) Stability of metabolites in extracted urine for up to 96-hours. E) Stability of normalized, extracted, and raw urine for 2 weeks at -80°C. F) Stability of normalized, extracted, and raw urine for 10 months at -80°C. Metabolite data (in semi-transparent red, metabolites from the SQC urine; in blue from the RQC) are from the annotated dataset (Fig. 4A in the main text) and acquired by LC-MS as described in the main text. ZHP=ZIC-HILIC positive ionization; RT=room temperature; RQC=reference quality control; SQC=pooled study quality control; Bravo=automated liquid sample handler (Agilent Technologies).
Figure S7. Effect of the co-eluting antibiotic trimethoprim on the intensities of the technical internal standard (tIS) CHES. ZHP=ZIC-HILIC positive ionization; SQC=pooled study quality control.
Figure S8. Creatinine molecular ion (A) and reporter ion (B, dimer) intensities across the injection sequence and extracted chromatograms of the molecular ion (C) and reported ion (D, dimer). SQC=pooled study quality control sample; RQC=reference quality control sample; tIS mix=technical internal standard mix.
