Urine biomarker data analysis guide

Overview of the results files

ProjectID_Nightingale_Health_Urine_Biomarker_Results_Date.xlsx

This spreadsheet contains the biomarker results in ratio to creatinine as well as in absolute concentrations (mmol/L). It also includes tags that inform on the quantification accuracy and summary information on rates of successful quantification and tags. The success rates and tag rates commonly observed in other cohorts are provided to put the numbers into context.

ProjectID_Nightingale_Health_Urine_Biomarker_Results_Date.csv

This spreadsheet contains the same biomarker results as the xlsx file, but stored in machine readable format for loading into data analysis software. Results scaled relative to creatinine, absolute concentrations as well as tags are concatenated.

ProjectID Nightingale Health Urine Biomarker Quality Report.pdf

This report provides an overview of the biomarker data quality observed in your project in comparison to levels commonly observed in other cohorts. The biomarker quality is summarized in terms of quantification success rates, quality control tag rates, and biomarker distributions.

Nightingale Health - Venous blood analysis - Urine biomarker data analysis guide.pdf

This document.

Explanation of the Urine Biomarker Quality Report

This report illustrates the biomarker distributions, quantification success and tag rates in your project as compared to reference cohorts. This comparison is made based on a handful of general population cohorts with diverse patient characteristics and geographical origins.

For the biomarker distribution plots, the reference distributions are provided separately for general population cohorts and cohorts of patients living with diabetes. In both cases, information from around five cohorts was combined to determine the distributions of the urine biomarkers. This information may help to compare the consistency of the results from your study with what is commonly observed. However, it is important to note that considerable deviation from these reference distributions can still occur without indicating poor sample quality or measurement issues.

The distributions are shown for a range which excludes outlier samples with values over 5 IQRs from the median. To improve visualization, a stricter filtering was applied for urea, creatinine, and 3–aminoisobutyrate, and glucose.

Biomarker annotation and tags

The annotation of the biomarker levels is the following:

Zero (0)	Biomarker is absent or has very low concentration.
Crea_missing	Creatinine is missing, and the ratio is therefore not computed.
NaN	Quantification of the given biomarker was not successful. No information exists whether the true concentration is low or high.

The following two tags are used to inform on the accuracy of the measurements

NOISY	Indicates that the measurement error in the absolute concentration of the given biomarker may be larger than usual. Measurements tagged as NOISY are often useful in epidemiological analyses, but if you prefer to maximize accuracy at the cost of having more missing values, you may try excluding measurements tagged as NOISY as a sensitivity analysis.
AMBIGUOUS	Indicates that the signal could not be identified unambiguously; either the signal is not visible because the concentration is low, or there are multiple overlapping signals due to interfering substances. The concentration provided is measured from the most plausible candidate and can be interpreted as an upper bound on the true concentration. The concentration is likely not larger than this, but it may be smaller. Ambiguous measurements are useful for most epidemiological analyses, and we do not recommend excluding them entirely because that would mean excluding many samples with low concentrations.

Note that a biomarker can have both NOISY and AMBIGUOUS tags.

Overview of the biochemical role of each biomarker

An overview of each biomarker is provided in the document “Urine analysis - CoreMetabolomics - Urine biomarker description - Nightingale Health.pdf”. The description covers the biochemical role of each biomarker as well as information in relation to diet and health outcomes when known.

Human metabolome database (HMDB) identifiers and CAS numbers are provided for each biomarker in the Results spreadsheet (xlsx). This file also contains the measurement coefficient of variation (CV) for each biomarker.

Biomarker coverage and groupings

Nightingale Health's urine biomarker service is designed for high-throughput rather than the widest possible metabolite coverage. The metabolites included in the panel are therefore selected based on feasibility for automated quantification in high throughput NMR spectroscopy. This approach emphasizes metabolites at high abundance in urine, and those which generate modest signal overlap. As such, the biomarker panel is not based on prior biological relevance of the metabolites or the emphasis of certain metabolic pathways.

The biomarkers have been grouped into different categories in the xlsx spreadsheet based on their biochemical relations and origins in urine. Please note that this grouping is provided as a suggestive overview to facilitate plotting and initial biological insights on the biomarker relations. It is good to keep in mind that many of the biomarkers have roles in multiple metabolic pathways.

Biomarker correlations and multiple testing

An example of the correlation structure of the urine biomarkers can be found on page 5. The correlation structure is generally consistent across cohorts. Urine biomarkers are generally less correlated than blood biomarkers since there is less homeostatic regulation in urine.

For multiple testing correction we recommend to use the same as the number of biomarkers tested since the biomarkers are only weakly correlated with each other. It takes ~40 principal components to explain 95% of the variance in the biomarker data, and 48 principal components to explain 99% of the variance.

Consistency to creatinine measured by clinical chemistry and alternative normalization

Studies have reported good correlation for urinary creatinine measured by clinical chemistry and Nightingale Health NMR from the same samples. The Multi-Ethnic Cohort study conducted by the National University of Singapore reports a correlation of R=0.903 as illustrated on page 6. The scatter plot indicates good match in the absolute units for creatinine measured by clinical chemistry from fresh samples to levels measured by Nightingale Health NMR from frozen samples years later.

The FinnDiane study reported a correlation R=0.88 for ~2,600 patients with type 1 diabetes. This consistency was observed despite a gap of up to 18 years between the clinical chemistry and NMR measurements (Mutter et al, *Diabetologia* 2022;65:140).

A comprehensive study on the influence of different normalization schemes in urine biomarker profiling by NMR was done by Li et al. Characteristics of Normalization Methods in Quantitative Urinary Metabolomics-Implications for Epidemiological Applications and Interpretations. *Biomolecules*. 2022 12(7):903. The results of this study indicate consistent epidemiological results when the biomarkers are scaled by creatinine or when using alternative scaling methods.

Software package for plotting biomarker associations

Displaying results visually is helpful when analyzing many biomarkers. We provide a tool for plotting results of association testing in the ggforestplot package for R, available here https://nightingalehealth.github.io/ggforestplot/index.html.

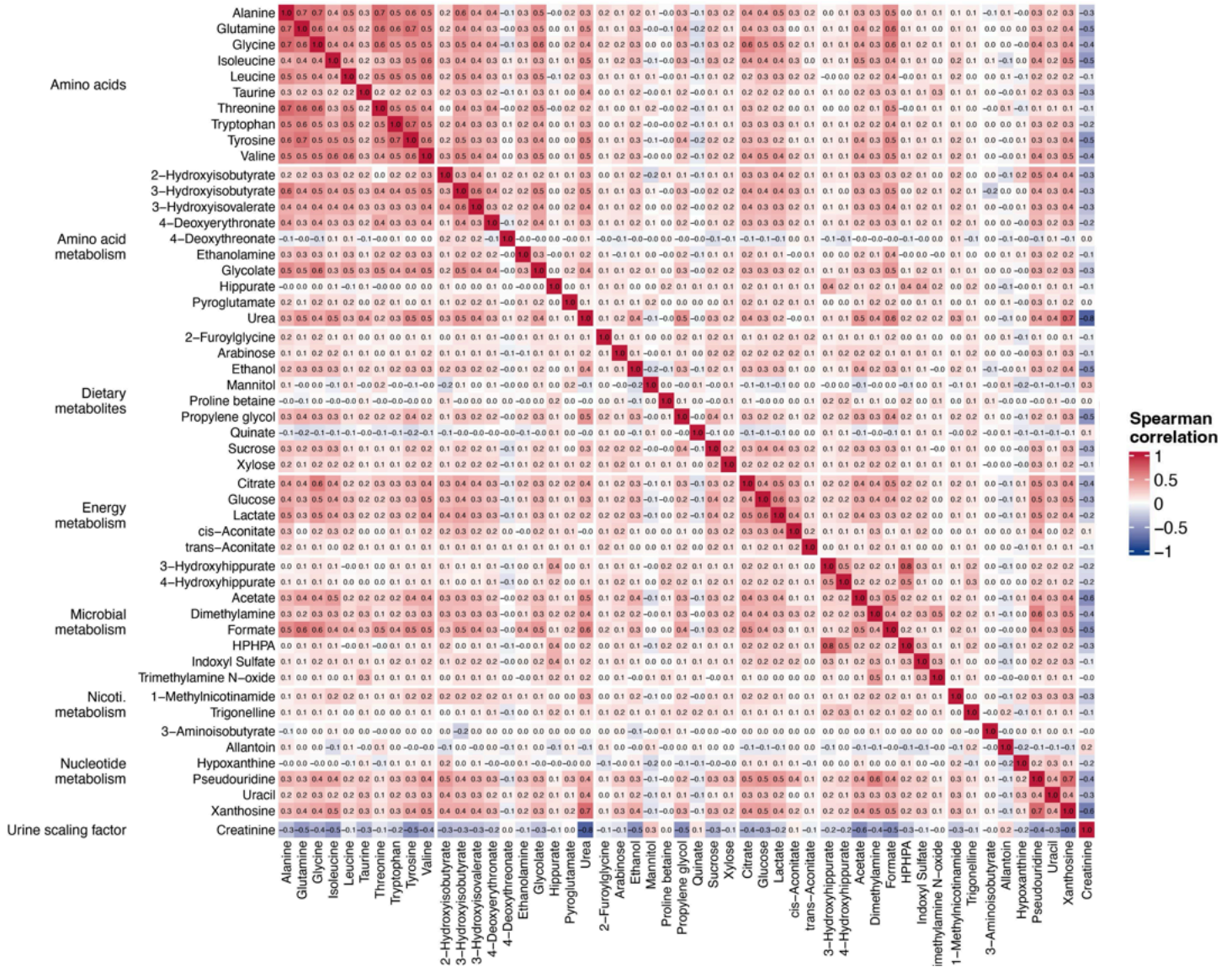
Although the examples in the ggforestplot R package are given for blood biomarkers, the plotting tool can be easily modified to display the urine biomarkers in a comprehensive overview. This can be helpful when comparing results with different covariate adjustments or for different subsets of the study population.

Example paragraph for citing in scientific publications

Metabolite biomarkers were measured from urine samples of XXXX individuals using proton nuclear magnetic resonance (NMR) (CoreMetabolomics by Nightingale Health Plc, Finland; quantification library 2024). In total, 50 metabolites and creatinine were quantified in absolute concentration (mmol/L) using an established NMR metabolomics protocol with automated total-line-shape fitting (Mutter et al, *Diabetologia* 2022;65:140). This approach simultaneously quantifies biomarkers from multiple metabolic pathways (e.g. amino acid and energy metabolism) and from various origins (e.g. microbial metabolism and diet). The NMR spectroscopy and bioinformatics setup was designed for quantification of identified metabolites at high-throughput rather than optimizing for widest possible metabolite coverage. Each metabolite was scaled relative to creatinine prior to analyses to normalize for variation in urine volume (Li et al. *Biomolecules*. 2022;12:903). All urine samples were [morning void/spot/24h] collection.

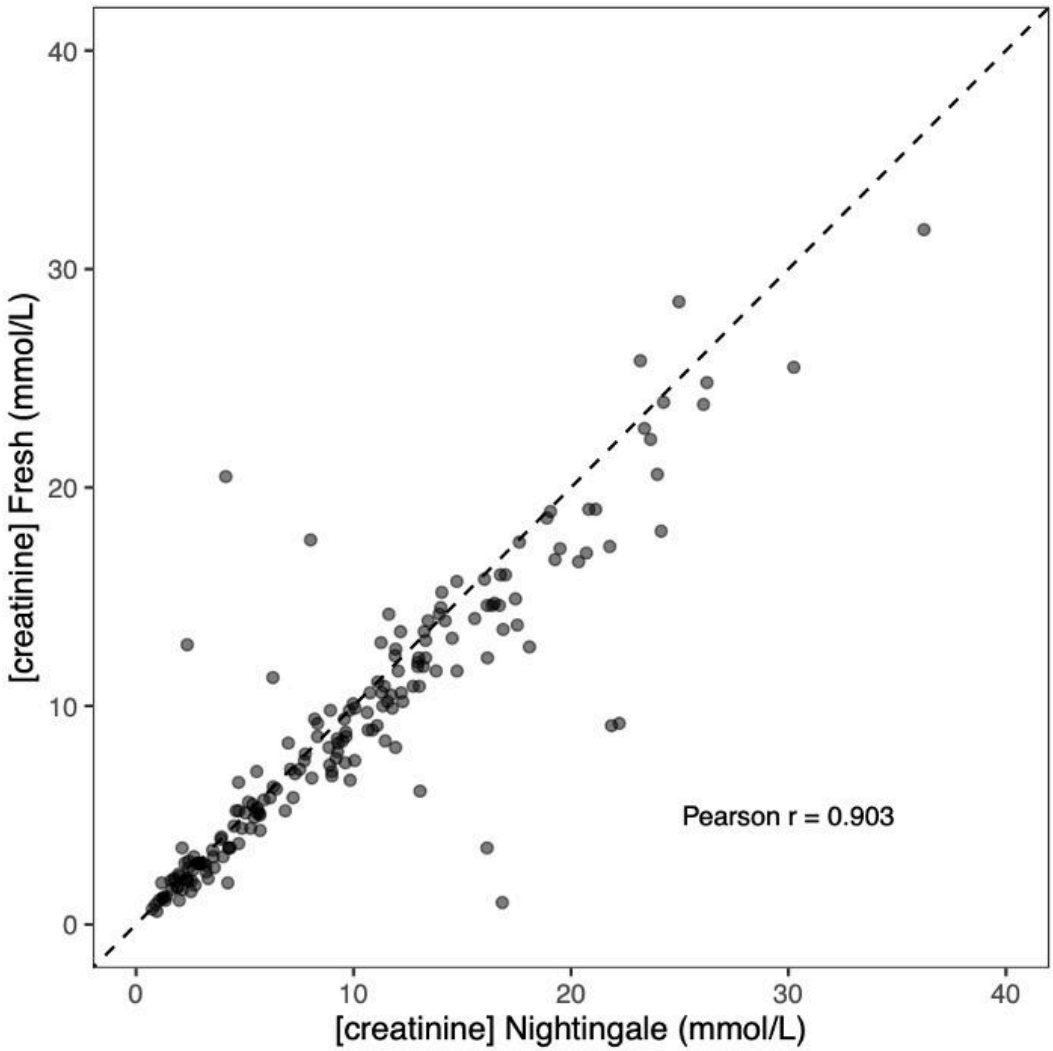
If further experimental details are desired consider the following text:
The experimental setup is as follows: urine samples were thawed overnight and subsequently centrifuged (5 min, 3000g). Robotic liquid handlers (JANUS 8-tip workstation; PerkinElmer Inc, USA) were used to mix 70 µL of phosphate buffer (1.5 M K2HPO4 and 1.5 M NaH2PO4 in D2O, pH 7.0; including 5.8 mM sodium 3-(trimethylsilyl)propionate-2,2,3,3-d4 and 30.8 mM sodium azide;) and 450 µL urine into 5mm NMR tubes placed on 96-tube racks. A 600 MHz Bruker AVANCE IIIHD NMR spectrometer with automated SampleJet sample changer and cryoprobe (CryoProbe Prodigy TCI) was used to acquire the spectral data at 298K. The spectroscopy settings were standard water-suppressed measurements (Bruker noesypresat pulse sequence with mixing time of 10 ms and irradiation field of 25 Hz) using 32 scans per sample with 5.1 s recycle time. Representative proton NMR spectra and metabolite assignment can be found in e.g. Tynkkynen et al *Int J Epidemiol*. 2019;48:978. A metabolite quantification protocol based on automated implementation of total-line-shape fitting was applied after piece-wise spectral alignment. The scientific literature on chemical shifts and J-couplings for high-abundance metabolites in urine was used to assign metabolite peak shapes. The metabolite identifications have further been confirmed with spiking and concentrations calibrated using standard addition (Mutter et al, *Diabetologia* 2022).

Correlation heatmap for the biomarkers



The correlations are ordered according to biomarker groups. Numbers indicate Spearman's correlations. All biomarkers are scaled to creatinine.

Urinary creatinine measured by Nightingale Health NMR and clinical chemistry from the same samples



Results from Multi-Ethnic Cohort, courtesy of National University of Singapore

Key publications related to urine metabolomics

Mutter et al. Urinary metabolite profiling and risk of progression of diabetic nephropathy in 2670 individuals with type 1 diabetes. *Diabetologia*. 2022;65:140.

Valo et al. Genome-wide characterization of 54 urinary metabolites reveals molecular impact of kidney function. *Nat Comms.* 2024, accepted.

Li et al. Clinical and biochemical associations of urinary metabolites: quantitative epidemiological approach on renal-cardiometabolic biomarkers. *Int J Epidemiol*. 2024;53(1):dyad162.

Li et al. Characteristics of normalization methods in quantitative urinary metabolomics-implications for epidemiological applications and interpretations. *Biomolecules*. 2022;12(7):903.

Tynkkynen et al. Proof of concept for quantitative urine NMR metabolomics pipeline for large-scale epidemiology and genetics. *Int J Epidemiol*. 2019;48(3):978.

Steinbrenner et al. Associations of urine and plasma metabolites with kidney failure and death in a chronic kidney disease cohort. *Am J Kidney Dis*. 2024:S0272-6386(24)00787-X.

Schlosser et al. Genetic studies of urinary metabolites illuminate mechanisms of detoxification and excretion in humans. *Nat Genet*. 2020;52(2):167-176.

Webinar on urine metabolomics by Nightingale Health

To learn more about epidemiological applications and recently published science on the Nightingale Health urine metabolomics service, please see the recording (https://vimeo.com/943654536) from the webinar of May 7, 2024, by Dr Peter Würtz.