

## RESEARCH ARTICLE

## Alterations in the molecular composition of COVID-19 patient urine, detected using Raman spectroscopic/computational analysis

John L. Robertson<sup>1,2,3†\*</sup>, Ryan S. Senger<sup>3,4,5‡</sup>, Janine Talty<sup>6‡</sup>, Pang Du<sup>3,7</sup>, Amr Sayed-Issa<sup>3</sup>, Maggie L. Avellar<sup>3,4</sup>, Lacey T. Ngo<sup>3</sup>, Mariana Gomez De La Espriella<sup>8</sup>, Tasaduq N. Fazili<sup>8</sup>, Jasmine Y. Jackson-Akers<sup>8</sup>, Georgi Guruli<sup>9</sup>, Giuseppe Orlando<sup>10</sup>

**1** Department of Biomedical Engineering and Mechanics, College of Engineering, Virginia Tech, Blacksburg, Virginia, United States of America, **2** Section of Nephrology, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, **3** DialySensors Incorporated, Blacksburg, Virginia, United States of America, **4** Department of Biological Systems Engineering, College of Life Sciences and Agriculture, Virginia Tech, Blacksburg, Virginia, United States of America, **5** Department of Chemical Engineering, College of Engineering, Virginia Tech, Blacksburg, Virginia, United States of America, **6** Clinical Biomechanics and Orthopedic Medicine, Roanoke, Virginia, United States of America, **7** Department of Statistics, College of Science, Virginia Tech, Blacksburg, Virginia, United States of America, **8** Internal Medicine/Infectious Disease, Carilion Clinic, Roanoke, Virginia, United States of America, **9** Division of Surgical Urology/Urologic Oncology, Department of Surgery, Virginia Commonwealth University, Richmond, Virginia, United States of America, **10** Department of Surgery, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America

☉ These authors contributed equally to this work.

‡ JLR, RSS and JT are joint senior authors on this work.

\* [drbob@vt.edu](mailto:drbob@vt.edu)



## OPEN ACCESS

**Citation:** Robertson JL, Senger RS, Talty J, Du P, Sayed-Issa A, Avellar ML, et al. (2022) Alterations in the molecular composition of COVID-19 patient urine, detected using Raman spectroscopic/computational analysis. PLoS ONE 17(7): e0270914. <https://doi.org/10.1371/journal.pone.0270914>

**Editor:** Benedetta Bussolati, Center for Molecular Biotechnology, ITALY

**Received:** December 21, 2021

**Accepted:** June 17, 2022

**Published:** July 18, 2022

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0270914>

**Copyright:** © 2022 Robertson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data relevant to this study are available from Github at <https://github.com/SengerLab/Raman-Scans/tree/COVID-19>.

## Abstract

We developed and tested a method to detect COVID-19 disease, using urine specimens. The technology is based on Raman spectroscopy and computational analysis. It does not detect SARS-CoV-2 virus or viral components, but rather a urine ‘molecular fingerprint’, representing systemic metabolic, inflammatory, and immunologic reactions to infection. We analyzed voided urine specimens from 46 symptomatic COVID-19 patients with positive real time-polymerase chain reaction (RT-PCR) tests for infection or household contact with test-positive patients. We compared their urine Raman spectra with urine Raman spectra from healthy individuals (n = 185), peritoneal dialysis patients (n = 20), and patients with active bladder cancer (n = 17), collected between 2016–2018 (i.e., pre-COVID-19). We also compared all urine Raman spectra with urine specimens collected from healthy, fully vaccinated volunteers (n = 19) from July to September 2021. Disease severity (primarily respiratory) ranged among mild (n = 25), moderate (n = 14), and severe (n = 7). Seventy percent of patients sought evaluation within 14 days of onset. One severely affected patient was hospitalized, the remainder being managed with home/ambulatory care. Twenty patients had clinical pathology profiling. Seven of 20 patients had mildly elevated serum creatinine values (>0.9 mg/dl; range 0.9–1.34 mg/dl) and 6/7 of these patients also had estimated glomerular filtration rates (eGFR) <90 mL/min/1.73m<sup>2</sup> (range 59–84 mL/min/1.73m<sup>2</sup>). We could not determine if any of these patients had antecedent clinical pathology abnormalities. Our

**Funding:** The authors received no specific funding for this work.

**Competing interests:** I have read the journal's policy and authors of this manuscript (JR, RS, PD) have the following competing interest. These authors have declared IP interests in the technology and Virginia Tech is seeking provisional patents for the technology that VT may intend to commercialize.

technology (Raman Chemometric Urinalysis—Rametrix<sup>®</sup>) had an overall prediction accuracy of 97.6% for detecting complex, multimolecular fingerprints in urine associated with COVID-19 disease. The sensitivity of this model for detecting COVID-19 was 90.9%. The specificity was 98.8%, the positive predictive value was 93.0%, and the negative predictive value was 98.4%. In assessing severity, the method showed to be accurate in identifying symptoms as mild, moderate, or severe (random chance = 33%) based on the urine multimolecular fingerprint. Finally, a fingerprint of 'Long COVID-19' symptoms (defined as lasting longer than 30 days) was located in urine. Our methods were able to locate the presence of this fingerprint with 70.0% sensitivity and 98.7% specificity in leave-one-out cross-validation analysis. Further validation testing will include sampling more patients, examining correlations of disease severity and/or duration, and employing metabolomic analysis (Gas Chromatography–Mass Spectrometry [GC-MS], High Performance Liquid Chromatography [HPLC]) to identify individual components contributing to COVID-19 molecular fingerprints.

## Introduction

Infection with SARS-CoV-2 and development of COVID-19 disease is associated with a deleterious effect in renal function and structure. This would be expected to potentially alter the molecular composition of urine. Since COVID-19 evolved in 2019, there have been numerous reports of acute kidney injury (AKI) associated with this infection [1–8]. The incidence of AKI in COVID-19 patients has been estimated to range from about 27–50+% [9, 10]. Early in the pandemic, several groups noted a correlation of disease severity, hospitalization, and intensive care admissions with increased risk for developing AKI [11, 12]. This was not surprising. The contribution of cardiopulmonary dysfunction, renal hypoperfusion, and/or multidrug therapy leading to the development AKI has been well-known for decades prior to COVID-19 [13]. However, the role of renal viral infection in the development of AKI was uncertain.

A recent review by Hassler and coworkers [14] considered evidence both for and against direct SARS-CoV-2 infection of the kidney. Rightly so, Hassler, et. al., and authors they cited, felt that viral infection might help explain the disproportionately high incidence of AKI and collapsing glomerulopathy seen in patients with COVID-19. Combining data from several studies, Hassler, et. al., presented at least putative evidence of viral infection in 102/235 kidneys (43%) from autopsied patients. A variety of techniques for viral/viral RNA detection were used in the cited studies, including immunohistochemistry, real-time polymerase chain reaction (RT-PCR), *in situ* hybridization, immunofluorescent microscopy, and electron microscopy [14]. No study they referenced used two or more methods for cross-checking and validating renal viral infection, a deficit in study design considered to affect interpretation in the results.

Hassler, et. al., [14] posited that renal biopsies (not autopsy-derived samples) and development of urine-based screening tests would be keys to understanding the effects of COVID-19 on renal function and structure and these would be needed to improve detection and management of disease.

Over the past four years, we have been developing and validating a novel approach to molecular urinalysis, using a combination of Raman spectroscopic, computational, and physicochemical analytical methods. This new method of urinalysis is termed Raman chemometric urinalysis—abbreviated as Rametrix<sup>®</sup>. We have successfully applied this method of urinalysis to determining the molecular characteristics and physical properties of urine specimens from

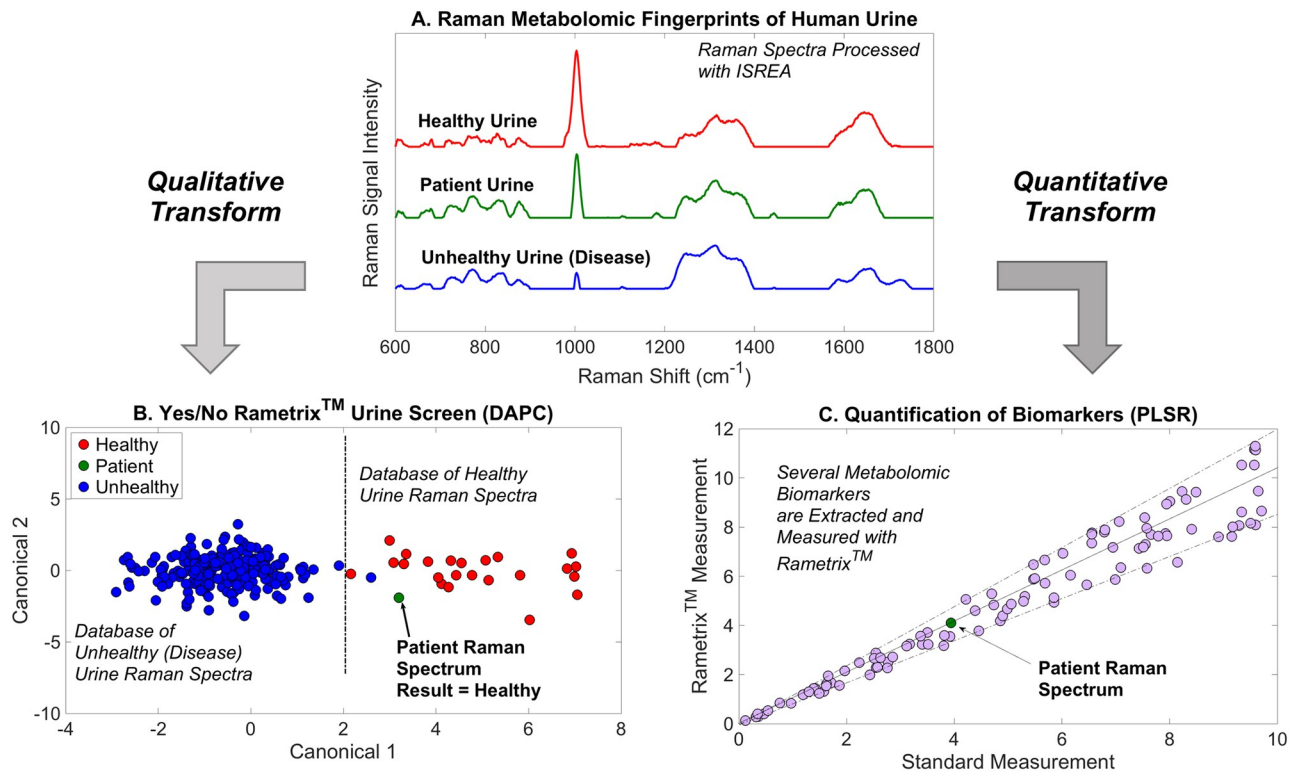
healthy human volunteers (n = 235 urine samples) [15], patients with chronic kidney disease (CKD) (n = 362 urine samples) [16], bladder cancer (BCa) and non-neoplastic genitourinary pathologies (such as cystitis and benign prostatic hypertrophy/urinary retention, among others) (n = 56 urine samples) [17, 18], and post-acute patients following exposure to Lyme pathogen (*Borrelia* sp.) (n = 30 urine samples) [19]. Data cited in these references was derived from Rametrix<sup>®</sup> analysis of urine specimens collected prior to December 2018.

Normal human urine contains over 2,000 separate chemical entities, reflective of systemic physiology/metabolism and the processes of renal function [20]. In the past decade, mass spectrometry, liquid/gas chromatography, nuclear magnetic resonance, and kinetic nephelometry methods have been used to detect analytes (i.e., biomarkers) associated with normal metabolism or disease [21]. Metabolomics has been used to identify kidney disorders (including renal cell carcinoma), coronary artery disease, diabetes, Alzheimer's disease and cognitive impairment, neurodegenerative disease, and colorectal cancer [22–27]. While metabolomics is often used to search for circulating/plasma disease biomarkers, it is now used to study how the presence of disease alters the urine metabolite profile (“fingerprint”).

Mass spectrometry-based urine biomarker and “-omics” technologies are used rarely by caregivers in patient care settings. This is due to expense, the daunting requirement for advanced technology, expertise required for interpretation of results, and the lack of assay validation requiring large datasets of normal and abnormal specimens. In fact, the complexity of both acute and chronic genitourinary tract pathologies makes large dataset sampling and validation with technology-intensive methods (like mass spectrometry and high-performance liquid chromatography) unlikely and cost-prohibitive.

As an alternative approach to mass spectroscopy-based urine metabolomics, we invented and extensively validated Rametrix<sup>®</sup> to analyze urine [15, 28, 29]. Raman spectroscopy is a mature, well-studied, and powerful technology that has been applied to analysis of the chemical composition of a wide variety of solids and liquids, including biological specimens [30–33]. Irradiation of molecular mixtures (like urine), with wavelength-specific laser energy, produces weak vibrational energy (Raman scatter radiation) from deformation/relaxation of the many chemical bonds in hundreds of distinct molecules in specimens. Different molecular constituents are represented by Raman ‘bands’ (i.e., signal intensity peaks) and these bands/peaks are indicative of chemical bond vibrations [34]. These vibrations may be present in several molecules with similar chemical bonds in a sample, meaning it can be difficult to assign individual Raman bands to specific molecules, unless they are present in abundance. This is the case for urea in urine, for example, where the C-N bond stretch at  $1,002\text{ cm}^{-1}$  is dominant and can be associated with urea concentration. We have also identified the bands/peaks of creatinine, heme, amino acids, albumin, collagen, and phospholipids in urine. A few of these and other broad molecular assignments are shown in Raman spectra of urine from healthy volunteers [15], CKD 4–5 patients [16], and Surine<sup>™</sup> urinalysis analytical control solution (Dyna-Tek Industries, Lenexa, KS) in Fig 1.

Because it is difficult to relate individual Raman bands to specific molecules, a chemometric approach is required to analyze Raman spectra of highly complex heterogeneous samples [28, 35–39]. The chemometric approach is unlike chromatographic and mass spectrometry approaches that resolve single molecules. The chemometric approach treats an entire Raman spectrum as a ‘fingerprint’ and then associates it with a condition (i.e., ‘healthy’, ‘chronic kidney disease,’ ‘COVID-19 infection’, etc.) using statistical models and artificial intelligence. Building an accurate model to predict the condition of an unknown sample requires a large dataset of pre-analyzed Raman spectra. This can be seen, for example, in Fig 1. Here, representative urine spectra are shown for healthy volunteers and patients with diagnosed disease. Chemometric models determine whether an “unknown” patient sample more closely resembles



**Fig 1. Schematic of qualitative and quantitative chemometric analyses of urine spectra using the Rametrix® Toolbox.**

<https://doi.org/10.1371/journal.pone.0270914.g001>

the healthy urine spectrum or one of a diseased state, without knowing the identities of all molecules in each sample.

Rametrix® computations are performed using the Rametrix® Toolbox for MATLAB, which is available to academic researchers through GitHub [28, 29]. The Rametrix® Toolbox offers two approaches to data analysis (also shown in Fig 1), (i) qualitative classification, and (ii) quantitative analysis. For classification (e.g., “yes/no” to the presence of disease) the Rametrix® Toolbox offers principal component analysis (PCA) followed by discriminant analysis of principal components (DAPC) [28, 29, 32]. Other deep learning and artificial neural network options are available, and other classifier models used with Raman spectra were surveyed recently [32]. Partial least-squares regression (PLSR) for quantitative analyses, and methods to identify quantifiable biomarkers from Raman spectra have been produced recently [17]. The Rametrix® Toolbox also includes ISREA [40, 41], which enables us to “exclude” non-diagnostic features in samples/spectra, such as the presence of blood/breakdown products, if such exclusion is warranted and logical [17, 32]. Finally, the Rametrix® Toolbox offers a graphical interface, spectral viewing pane, and predictive model cross-validation through leave-one-out analysis [29].

Analysis of urine by Rametrix® is inexpensive (uses off-the-shelf Raman spectrometers and costs a few dollars a sample for consumables and analysis), rapid (typically less than 30 minutes to process and interpret a sample), requires no urine sample preparation or chemical manipulation, and is non-invasive (voided samples are analyzed).

These characteristics, and ease-of-use, suggested to us that this Raman spectroscopy-based technology might be useful in determining if a COVID-19 ‘molecular fingerprint’ was present

in the urine of diseased patients. If so, this fingerprint could be useful in disease detection and patient management.

## Methods

### Patients and controls

**Informed consent.** The collection and analysis of all de-identified urine specimens reported in this study and Raman spectra from previously reported reference studies were performed in accordance with the principles of the Declaration of Helsinki. Informed consent for the collection of urine specimens was obtained from healthy individuals (no evidence of COVID-19 disease) and RT-PCR-positive COVID-19 patients. Following the description of the proposed study by study personnel to prospective participants, patients who provided verbal consent for the study were provided with specimen cups, urine collected, and the specimen was encoded by study personnel. Informed consent for participation was recorded in patient medical records securely retained by study personnel (JT). This study, and previously reported studies cited in this document, were approved by the Virginia Tech Institutional Review Board (VT IRB #s 15–703, 20–924, 21–569) and Virginia Commonwealth University Institutional Review Board (reference bladder cancer study) (VCU IRB # HM20006879).

**Study subjects: COVID-19 patients.** Forty-six (46) patients, with clinical signs of COVID-19 disease, RT-PCR confirmation of nasopharyngeal infection and/or household/congregate and temporal exposure to RT-PCR confirmed patients, were seen by a primary care physician for disease/symptom management. All patients were symptomatic, but ambulatory, at the time of evaluation. The patient population consisted of 32 female and 14 male patients. The age range of female patients was 18–68 years (average age 47.84 years/old) and of male patients was 18–62 years (average age 47.85 years/old). As would be expected, the clinical presentation of patients was highly variable, as was the duration and severity of clinical symptoms. Thirty out of 46 patients (30/46) were seen for evaluation (and specimen collection) within the first 14 days of clinical disease. Twelve of 46 patients (12/46) had clinical disease present for 30–300 days, and 10 of 46 patients (10/46) had clinical disease present for 60–300 days. Based on physician evaluation and self-assessment, 25/46 patients presented with ‘mild, symptomatic’ disease, while 14/46 patients presented with ‘moderate, symptomatic’ disease. Seven of 46 patients (7/46) were classified ‘severe, symptomatic’ at the time of presentation and specimen collection.

Several patients had pursued multiple avenues of diagnosis and variable courses and types of therapies (including antibiotics) prior to evaluation.

Clinical pathology evaluation (serum chemistry) was performed on a subset of 20/33 patients. Seven of 20 patients had mildly elevated serum creatinine values ( $>0.9$  mg/dl; range 0.9–1.34 mg/dl) and 6/7 of these patients also had eGFR  $<90$  mL/min/1.73m<sup>2</sup> (range 59–84 mL/min/1.73m<sup>2</sup>).

**Controls: Healthy volunteers (pre-COVID-19).** A full analysis of the healthy human volunteer urine dataset has been published [15]. This dataset contains 235 urine specimens collected from 39 females and 9 males, and all were collected prior to December of 2018. All volunteers were healthy (free of infectious or degenerative disease) at the time of collection and had no history or evidence of renal disease. The population ranged in age from 18 to 70 years, and 87.5% were between 19–22 years of age (median of 21 years). A total of 185 urine spectra were selected randomly from this dataset and used in this study.

**Controls: Healthy volunteers (fully vaccinated against COVID-19).** For this study, 19 additional urine specimens were collected from healthy volunteers who had been fully



vaccinated against COVID-19 and who had no history or evidence of either renal disease or COVID-19 disease.

**Controls: End-stage renal disease (ESRD) patients (pre-COVID-19).** The ESRD patient urine dataset has also been published [16]. It contains 362 urine specimens from 96 patients receiving treatment for ESRD with peritoneal dialysis therapy. The age range was 24 to 90 years, with a mean of 60 and median of 63.5 years. Twenty (20) spectra were selected randomly from this dataset and used in this study.

**Controls: Bladder cancer (BCa) patients (pre-COVID-19).** We have also published urine spectra from patients with active or remissive BCa [18]. The dataset contains 56 urine specimens (one per patient) from patients between 31–91 years old (mean and median of 62 years) were collected. The patients ranged in age from 31–91 years old. The mean and median age of 62 years. From this dataset, we selected 17 specimens from patients with active BCa at the time of collection for this study.

**Specimen collection and storage.** Voided, mid-stream urine specimens were collected, frozen immediately at  $-15^{\circ}\text{C}$ , and stored at  $-35^{\circ}\text{C}$  until analysis. We have demonstrated the suitability of this procedure for preserving samples [18].

**Raman methodology and measurements.** Urine specimens were analyzed at room temperature in bulk liquid form using 2 mL screw thread flat bottom borosilicate glass vials (Fisher Scientific). A Wasatch Photonics 785 nm dispersive Raman spectrometer (Wasatch Photonics, Morrisville, NC) was used with a Rametrix<sup>®</sup> AutoScanner (DialySensors, Inc., Blacksburg, VA) to automate sample scanning. The following settings were used:  $25^{\circ}\text{C}$ , 785 nm laser, 30 s excitation time, 30 mW laser power, 0.2 mm laser spot size, 200–2000  $\text{cm}^{-1}$  range, and spectral resolution of 8  $\text{cm}^{-1}$  (manufacturer default). Ten scans were obtained per vial. ENLIGHTEN<sup>™</sup> software (Wasatch Photonics) was used for spectrometer operation, and molecular contributions investigated with a published database [42]. In all cases, Raman intensity and wavenumber calibrations were performed during each operation of the Raman spectrometer using Surine<sup>™</sup> urine analytical control (see below) and published chemometric protocols.

**Analytical standards.** Surine<sup>™</sup> Urine Negative Control (Dyna-Tek Industries, Lenexa, KS) was used as a control in this study.

**Computational methodology.** Previously published computational methods were used [19, 28, 29, 42] with the Rametrix<sup>®</sup> Toolbox (LITE v1.1 and PRO v1.0) with added capabilities for ISREA baselining [40, 41]. Calculations were performed in MATLAB R2018A (Mathworks; Natick, MA). Raman spectra were truncated to 600–1800  $\text{cm}^{-1}$ , baseline corrected with ISREA, averaged over the 10 scans for each urine specimen, and vector normalized. ISREA was applied using nodes (or ‘knots’) at wavenumbers of 400, 950, 1100, 1500, and 1800  $\text{cm}^{-1}$ . In specified cases, the placement of nodes was also allowed to vary, as described previously [19], to exclude specific Raman shift regions of spectra selectively. Spectra were analyzed by principal component analysis (PCA) and discriminant analysis of principal components (DAPC) in the Rametrix<sup>®</sup> LITE Toolbox, and models were cross-validated with leave-one-out analysis with Rametrix<sup>®</sup> PRO. These methods have been implemented and described previously [19, 28, 32, 35–39]. This procedure allowed calculation of overall prediction accuracy, sensitivity, specificity, positive-predictive value (PPV), and negative-predictive value (NPV) for detecting COVID-19 in human urine. These metrics have also been defined in previous publications, and here, the RT-PCR nasopharyngeal swab test and/or proximate/congregate/temporal exposure to COVID-19 positive patients is treated as the “Gold-Standard” test when comparing to the Rametrix<sup>®</sup> urine screen.

**Statistical comparisons.** Of spectra were performed through the calculation of total spectral distance (TSD), as has been demonstrated [15, 16, 18, 19, 29, 32]. In calculation of TSD, the difference between each urine spectrum and that of Surine<sup>™</sup> was calculated at each

wavenumber and summed. One-Way Analysis-of-Variance (ANOVA) and pairwise comparisons using Tukey's honestly significant difference (HSD) procedure were used to determine if TSD values of COVID-19 patient urine were different from those of healthy volunteers and those with other diseases.

## Results

### Raman dataset

Urine Raman spectra from the patient dataset shown in Table 1 were used in this study. Spectra obtained from patients with active COVID-19 are a new contribution in this study, along with spectra obtained from healthy volunteers vaccinated for COVID-19 but with no known history or exposure to COVID-19. Other Raman scans for healthy human volunteers (collected and analyzed pre-2019; i.e., before COVID-19).

### Spectral processing

All spectra were truncated between 600–1,800  $\text{cm}^{-1}$ , baselined using ISREA [40, 41], vector normalized, and averaged for each urine specimen. For the ISREA implementation, nodes were applied at 400, 950, 1100, 1500, and 1800  $\text{cm}^{-1}$ . The concept of ISREA node placement and optimization has also been introduced recently [32]. Averaged spectra from all classes listed in Table 1 are shown in Fig 2. The most notable observable difference from the spectra of the COVID-19 and Healthy (pre-2019) classes was the height of the urea representative band (1,002  $\text{cm}^{-1}$ ). Inspection also revealed other minor differences (e.g., 970  $\text{cm}^{-1}$ ; 1100–1200  $\text{cm}^{-1}$ ), prompting additional differences to be investigated by chemometric methods. To date, patients with ESRD have the most visually different urine Raman spectra from the Healthy class (Fig 2). However, it is noted that the ESRD class is also identified by a reduced urea Raman band intensity.

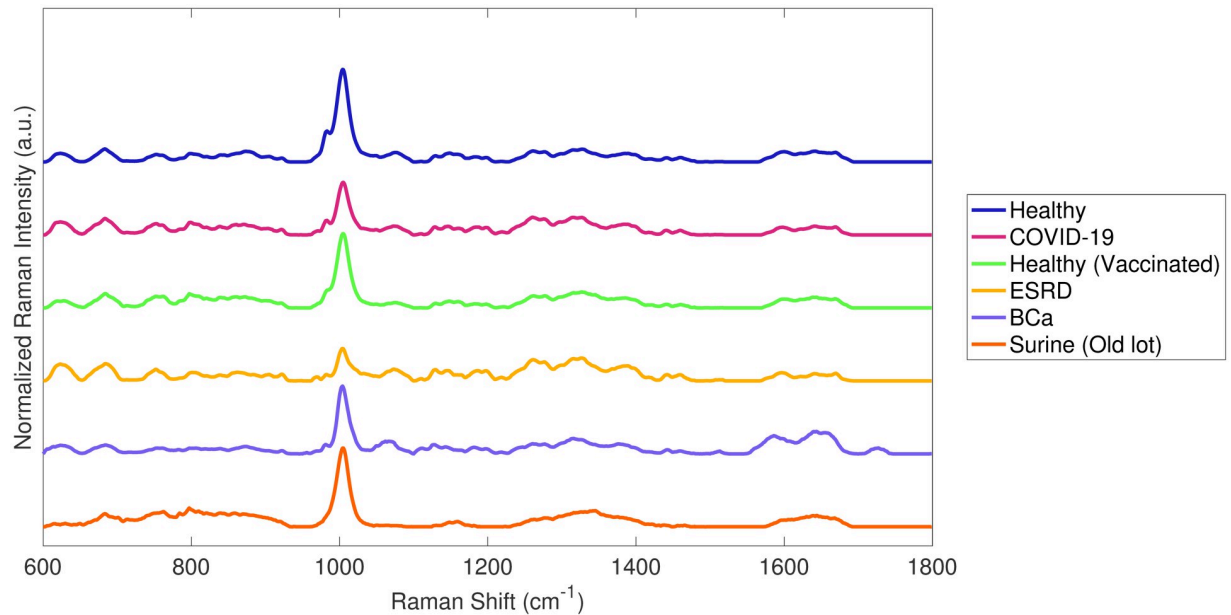
### Statistical significance

Total Principal Component Distance (TPD) [18, 28, 29] calculations were performed to determine if urine Raman spectra of the “COVID-19” class were different from all other classes (combined to form a “non-COVID-19” class). In short, TPD uses ISREA baselined and

**Table 1. Patient dataset analyzed in this study.**

Number of Urine Specimens	Description	Classification	Reference
185	Healthy human volunteers (pre-2019)	“Healthy”	Senger et al. 2019
20	Peritoneal dialysis patients with CKD 4–5	“ESRD”	Senger et al. 2020
17	Patients with active bladder cancer	“BCa”	Huttanus et al. 2020
6	Surine™ (lot from 2016)	“Surine”	Huttanus et al. 2020 and This study
5	Surine™ (lot from 2021)	“Surine”	This study
19	Healthy human COVID-19 vaccinated volunteers (2021)	“Healthy”	This study
46	Patients with active COVID-19	“COVID-19”	This study
25	Patients with ‘mild’ severity COVID-19 symptoms	“COVID-19 (Mild)”	This study
14	Patients with ‘moderate’ severity COVID-19 symptoms	“COVID-19 (Moderate)”	This study
7	Patients with ‘severe’ COVID-19 symptoms	“COVID-19 (Severe)”	This study
12	Patients with COVID-19 clinical disease lasting longer than 30 days	“COVID-19 (Long COVID 19)”	This study

<https://doi.org/10.1371/journal.pone.0270914.t001>



**Fig 2.** Averages of ISREA baselined and vector normalized spectra for classes specified in Table 1.

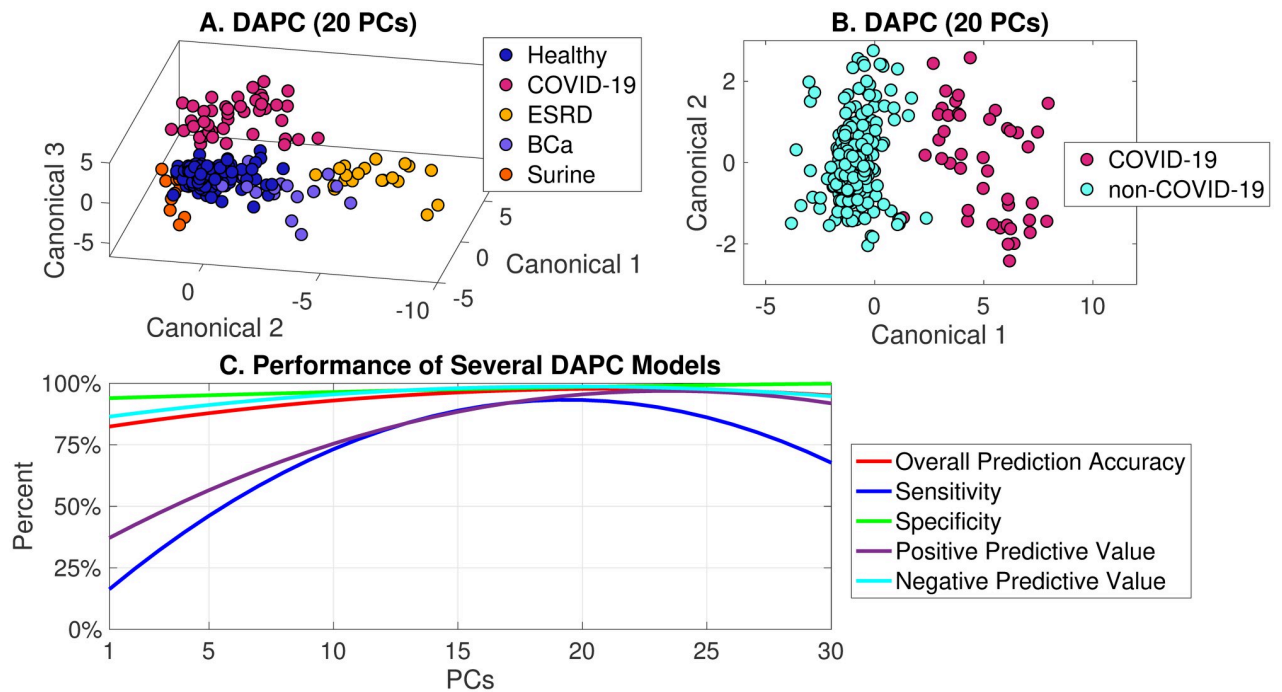
<https://doi.org/10.1371/journal.pone.0270914.g002>

processed spectra. The first five principal components (PCs) of PCA are used. For each sample, the distance (across all five PCs) is calculated between that sample and Surine™ (using a simple distance formula). This provides a distance calculation for every sample in the dataset. Then, ANOVA and pairwise comparisons are used to determine if statistically significant distances exist between classes of spectra. TPD was applied to the COVID-19 class and all other classes grouped as non-COVID-19. Through this method, the COVID-19 class and non-COVID-19 grouped class were found statistically significant ( $p < 0.001$ ). Pairwise comparisons were applied, and the COVID-19 class was found statistically different ( $p < 0.001$ ) against all other groups in Table 1. From experience, this indicates that an effective predictive model may be able to be constructed from PCA followed by DAPC. Of other pairwise comparisons, it was found that the Healthy and Surine™ groups were not statistically different from one another ( $p = 0.79$ ) according to TPD calculations.

### Predictive urine screen for COVID-19

With the COVID-19 class showing statistical significance from all other classes, predictive models were built using PCA and DAPC. The model inputs were truncated, baselined, and normalized urine spectra, and the ISREA nodes of 400, 950, 1100, 1500, and 1800  $\text{cm}^{-1}$  were used in this initial model-building. Spectra were processed further by PCA to produce principal components (PCs). A specified number of PCs were then fed into DAPC to return a “yes/no” for the presence of COVID-19. Predictive models differed by the number of PCs fed into DAPC, and these were evaluated for performance by leave-one-out cross-validation. Results are shown in Fig 3A for a model designed to separate all classes in Table 1 using 99% of the dataset variance (available in the top 20 PCs). This plot was effective in showing cluster separation, particularly that the COVID-19 cluster separated from the Healthy cluster more effectively than the BCa group did. The separation of COVID-19 and the non-COVID-19 groups for this model is shown in Fig 3B. When cross-validated with leave-one-out, an overall prediction accuracy of 97.6% for our dataset. The sensitivity of this model for detecting COVID-19





**Fig 3. DAPC models demonstrating cluster separation of COVID-19 Raman urine spectra from those of other groups (A, B), and DAPC predictive model performance (C).**

<https://doi.org/10.1371/journal.pone.0270914.g003>

was 90.9%. The specificity was 98.8%, the positive predictive value (PPV) was 93.0%, and the negative predictive value (NPV) was 95.8%. Finally, the performance of several DAPC models is given in Fig 3C and illustrate the influence of model architecture on performance. It is noted that at least 6 PCs were required to obtain overall prediction accuracy, sensitivity, specificity, positive predictive value, and negative predictive value above 50%.

Next, we varied the location of the ISREA nodes with the objective of reducing the number of PCs required in DAPC. Several iterations were tried, and node positions of 400, 439, 446, 605, 1045, 1163, 1247, 1443, 1739, 1768, and 1775  $\text{cm}^{-1}$  yielded similar results to those reported above; however, only 4 PCs of the dataset were required (as opposed to 20). A comparison of the two node sets for detecting the presence of COVID-19 in urine with Rametrix<sup>®</sup> analysis is shown in Table 2.

### Molecular contributions

The separation of COVID-19 and non-COVID-19 cluster separation in Fig 3B was investigated further through PC and canonical (DAPC) loadings. Significant Raman shifts (defined as above 0.2% total contribution) are given in Table 3. The molecular assignments were obtained by a published database [43]. It was observed the occurrences of lipids/cholesterol,

**Table 2. Detection of COVID-19 in urine by Rametrix<sup>®</sup> given two different ISREA node sets.**

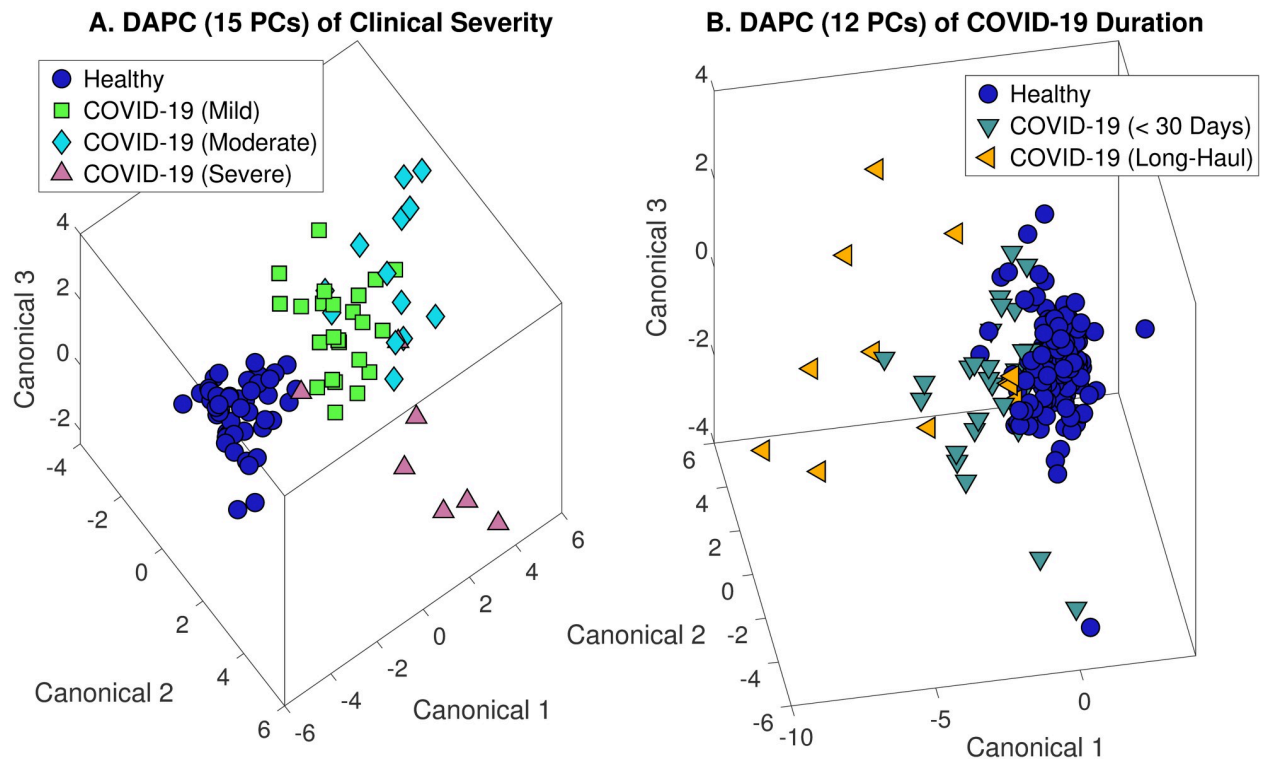
ISREA Nodes	PCs	Overall Accuracy	Sensitivity	Specificity	PPV	NPV
400, 950, 1100, 1500, and 1800 $\text{cm}^{-1}$	20	97.6%	90.9%	98.8%	93.0%	95.8%
400, 439, 446, 605, 1045, 1163, 1247, 1443, 1739, 1768, and 1775 $\text{cm}^{-1}$	4	97.6%	93.2%	98.4%	91.1%	98.8%

<https://doi.org/10.1371/journal.pone.0270914.t002>

**Table 3. Molecular assignments for Raman shifts leading to cluster separations in Fig 3.**

Raman Shift (cm <sup>-1</sup> )	Present in PCA, DAPC, or Both	Molecular Assignment [43]
425	Both	N/A
445	Both	N-C-S stretch
485	Both	Glycogen
518	DAPC	Phosphatidylinositol
614	DAPC	Cholesterol ester
621	Both	C-C twisting of phenylalanine
627	DAPC	N/A
682	DAPC	N/A
688	PCA	N/A
702	DAPC	Cholesterol ester
719	PCA	Lipids
776	PCA	Phosphatidyl inositol
782	PCA	DNA
810	PCA	Phosphodiester
817	PCA	Collagen
830	PCA	Phosphate stretching, Tyrosine
847	PCA	Monosaccharides
860	DAPC	Phosphate group
880	Both	Tryptophan
893	PCA	C-C backbone
900	DAPC	N/A
906	DAPC	Tyrosine
913	DAPC	Glucose
955	PCA	Carotenoids
980	Both	Beta-sheet proteins
992	DAPC	Red blood cell, phenylalanine, NADH
1002	Both	Urea
1006	Both	Carotenoids (absent in normal tissue)
1008	DAPC	Phenylalanine
1013	DAPC	N/A
1030	DAPC	Phenylalanine of collagen
1049	DAPC	Glycogen
1058	PCA	Lipids
1073	PCA	Fatty acids
1077	DAPC	Lipids, phospholipids, phosphate
1080	DAPC	Phospholipids, phosphate, collagen, tryptophan
1104	PCA	Phenylalanine
1107	PCA	N/A
1126	Both	Protein, disaccharides, lipids
1185	PCA	Phosphate
1240	DAPC	RNA, phosphate, collagen
1327	Both	Nucleic acids
1396	Both	Beta-carotene
1491	PCA	Amino radical cations
1607	Both	Tyrosine and phenylalanine
1630	PCA	N/A
1641	DAPC	N/A

<https://doi.org/10.1371/journal.pone.0270914.t003>



**Fig 4. DAPC models demonstrating cluster separation of COVID-19 Raman urine spectra by clinical severity (A) and duration of symptoms (Long-Haul) (B).**

<https://doi.org/10.1371/journal.pone.0270914.g004>

collagen, and phosphates occurred more regularly than our similar analyses of CKD and BCa [16, 18].

### Severity of clinical symptoms

The separation of COVID-19 urine Raman scans by severity of clinical symptoms from those of healthy volunteers is shown in Fig 4A. When clustered by DAPC, the COVID-19 data largely separated by severity, with mild symptoms clustering closer to healthy scans, on average. However, we note one urine sample from a COVID-19 patient with severe symptoms clustered with the healthy group. This is unexplained by our analysis. With separation by severity, the presence of COVID-19 was detected with 93.5% overall accuracy (87.5% sensitivity, 100% specificity, 100% PPV, and 88.0% NPV). When inspecting among COVID-19 samples to determine severity, 60–66% overall accuracy was achieved for the three levels (mild, moderate, and severe). Full results are given in Table 4. We note that given the three levels of severity, the random chance of correct prediction is 33%.

**Table 4. Prediction of COVID-19 clinical severity by Rametrix<sup>®</sup> analysis of urine.**

Clinical Severity*	Accuracy	Sensitivity	Specificity	PPV	NPV
Mild	59.1%	79.2%	35%	59.4%	58.3%
Moderate	65.9%	61.5%	67.7%	44.4%	80.8%
Severe	65.9%	57.1%	67.6%	25%	89.3%

\* Random chance of correct prediction is 33%.

<https://doi.org/10.1371/journal.pone.0270914.t004>

**Table 5. Detection of long COVID-19 in urine by Rametrix<sup>®</sup> given two different ISREA node sets.**

ISREA Nodes	PCs	Overall Accuracy*	Sensitivity	Specificity	PPV	NPV
400, 950, 1100,1500, and 1800 cm <sup>-1</sup>	12	95.1%	25.0%	98.7%	50.0%	96.2%
600, 1074, 1153, 1230, 1313, 1416, 1507, 1800 cm <sup>-1</sup>	4	97.6%	70.0%	98.7%	70.0%	98.7%

\* Random chance of correct prediction is 50%.

<https://doi.org/10.1371/journal.pone.0270914.t005>

## Identification of long COVID19

Of the patients treated for COVID-19, 26% (12/46) showed symptoms longer than 30 days. Of this group, over 83% showed symptoms for longer than 60 days (up to 300 days). Thus, this group of patients was defined as ‘COVID-19 (Long COVID19)’. A molecular fingerprint was sought to determine if urine metabolomic differences existed between the Long COVID19 group and those whose symptoms resolved in less than 30 days (regardless of clinical severity). Using the initial ISREA node set (400, 950, 1100,1500, and 1800 cm<sup>-1</sup>), no significance was found, where prediction accuracy, sensitivity, specificity, PPV, and NPV all exceeded 50% (the random chance of correct prediction). To find a signal, we shortened the truncation range to 600–1800 cm<sup>-1</sup> and searched for additional ISREA nodes. Ultimately, a node set was located that led to 70% sensitivity, with better than 98% specificity, for detecting Long-Haul COVID-19. The results are shown in Fig 4B, and a comparison of the ISREA node sets with prediction metrics is given in Table 5. We note that the plot in Fig 4B was constructed with 12 PCs (instead of 4) to better show the separation of Long-Haul samples.

## Discussion

We believe this is the first study demonstrating that SARS-CoV-2 infection changes the chemical composition of urine. These changes—complex, multimolecular ‘fingerprints’—can be detected using Raman spectroscopic examination and computational analysis. Sample analysis is low-cost (dollars per sample) and rapid (results <30 minutes). This analytical method does not detect virus or viral components. This method also does not identify a single “biomarker” of COVID-19 disease, but rather a “biomarker pattern” composed of molecular clusters associated with disease. These biomarker patterns reflect systemic inflammatory, immunologic, and metabolic reactions to infection. We hypothesize that viral infection of the kidney (if substantiated) may affect renal form/function and urine composition. Our results support information in many of the studies reviewed/critiqued by Hassler, et. al. [14].

This study would not have been possible without access to a large database of urine Raman spectra (n = 235) from healthy volunteers collected prior to SARS-CoV-2/COVID-19 and other large databases of urine Raman spectra (pre-COVID-19 CKD and bladder cancer patients) These provided the critical context for interpretation of spectra from COVID-19 patients. Without access to large, pre-COVID-19 spectral databases, we could not have been sure that what was observed in our study was related to SARS-CoV-2 infection. Other investigators may find that a lack of pre-COVID-19 clinical samples (urine, serum, from healthy and diseased individuals, for example) is a challenge as they search for COVID-19-related biomarkers.

Our understanding of the myriad effects of COVID-19 disease is in its infancy. The tropisms of the pathogen, clinical course of infection, the ongoing evolution of variants under immunologic pressure, and individual responses to infection have only been intensively studied for slightly more than 23 months. Largely unknown are long-term effects of infection. We do not know if acute infection evolves to chronic disease and organ dysfunction. This may be

especially important for COVID-19 patients who had antecedent renal disease. By analogy, Yu and Bonventre [44] noted the complex interplay between diabetic kidney disease (DKD), diabetes mellitus (DM), and AKI. They noted that DKD/DM patients were more likely to develop AKI and that the interactions of these disease process affected development of CKD and ESRD. One might expect that patients with COVID-19-associated AKI might be at increased risk for progression of chronic renal disease.

We acknowledge limitations of this study. The small number of patients studied ( $n = 46$ ) limits computational comparisons and correlations of disease severity and duration. These comparisons will require much more sampling and data analysis. None of the patients studied had AKI, although standard laboratory metrics (serum creatinine, eGFR) for patients indicated a mild degree of renal dysfunction. We did not follow patients with serial samples and therefore it is not possible to determine if there is a progression of renal dysfunction or a return to pre-COVID renal metrics. These limitations will be addressed in ongoing and planned studies.

What are the next steps? First, we need to conduct a larger study of COVID-19 patients to determine 1) are the results of this preliminary study validated with further sampling, 2) how does the molecular fingerprint vary among individuals, 3) does the molecular fingerprint differ in individuals with different disease severities and durations, 4) how long does the molecular fingerprint persist following acute infection, 5) can the molecular fingerprint specifically indicate renal infection, and 6) can the molecular fingerprint predict development of AKI? Second, the important molecular clusters contributing to the molecular fingerprint will need to be studied with confirmatory metabolomic analysis. With this additional information, we may have a validated, non-invasive, inexpensive method to monitor systemic manifestations of disease. This could be used to detect infections potentially missed with current PCR/antigen-based technologies and to monitor the efficacy of therapy and/or detect possible disease progression. We expect it will be a very useful tool for monitoring direct/indirect renal effects of COVID-19 disease. This technology, once more fully validated, could easily be used for non-invasive, repetitive monitoring of individuals who choose not to be vaccinated, to detect 'break-through' infections in vaccinated individuals, and to differentiate COVID-19 disease from seasonal respiratory infections (influenza), and may be especially useful in the detection and management of Long COVID19.

## Conclusions

Our preliminary data shows that SARS-CoV-2 infection and COVID-19 disease alters the molecular composition of urine, as determined by Raman spectroscopy and computational analysis. We believe our findings could be applied to disease detection, early screening for serious renal complications such as AKI, and overall management of COVID-19.

## Author Contributions

**Conceptualization:** John L. Robertson, Ryan S. Senger, Mariana Gomez De La Espriella, Georgi Guruli.

**Data curation:** John L. Robertson, Ryan S. Senger, Janine Talty, Pang Du, Amr Sayed-Issa, Lacey T. Ngo, Mariana Gomez De La Espriella, Jasmine Y. Jackson-Akers, Georgi Guruli.

**Formal analysis:** John L. Robertson, Ryan S. Senger, Pang Du, Amr Sayed-Issa, Maggie L. Avellar, Lacey T. Ngo.

**Investigation:** John L. Robertson, Ryan S. Senger, Amr Sayed-Issa, Mariana Gomez De La Espriella, Tasaduq N. Fazili, Jasmine Y. Jackson-Akers, Georgi Guruli, Giuseppe Orlando.



**Methodology:** John L. Robertson, Ryan S. Senger, Pang Du, Amr Sayed-Issa, Maggie L. Avellar, Lacey T. Ngo, Mariana Gomez De La Espriella, Georgi Guruli, Giuseppe Orlando.

**Project administration:** John L. Robertson.

**Resources:** John L. Robertson.

**Software:** Ryan S. Senger, Pang Du.

**Supervision:** John L. Robertson.

**Validation:** John L. Robertson, Ryan S. Senger.

**Visualization:** John L. Robertson, Ryan S. Senger.

**Writing – original draft:** John L. Robertson, Ryan S. Senger, Janine Talty, Pang Du, Maggie L. Avellar, Mariana Gomez De La Espriella, Tasaduq N. Fazili, Jasmine Y. Jackson-Akers, Georgi Guruli, Giuseppe Orlando.

**Writing – review & editing:** John L. Robertson, Ryan S. Senger, Mariana Gomez De La Espriella, Tasaduq N. Fazili, Jasmine Y. Jackson-Akers, Georgi Guruli, Giuseppe Orlando.

## References

1. Battle D, Soler MJ, Sparks MA, Hiremath S, South AM, Welling PA, et al. COVID-19 and ACE2 in Cardiovascular, Lung, and Kidney Working Group: Acute kidney injury in COVID-19: Emerging evidence of a distinct pathophysiology. *J Am Soc Nephrol* 2020; 31: 1380–1383 <https://doi.org/10.1681/ASN.2020040419> PMID: 32366514
2. Chen YT, Shao SC, Lai EC, Hung MJ, Chen YC. Mortality rate of acute kidney injury in SARS, MERS, and COVID-19 infection: A systematic review and meta-analysis. *Crit Care* 2020; 24: 439 <https://doi.org/10.1186/s13054-020-03134-8> PMID: 32677972
3. Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int* 2020; 97: 829–838 <https://doi.org/10.1016/j.kint.2020.03.005> PMID: 32247631
4. Ng JH, Bijol V, Sparks MA, Sise ME, Izzedine H, Jhaveri KD: Pathophysiology and pathology of acute kidney injury in patients with COVID-19. *Adv Chronic Kidney Dis* 2020; 27: 365–376 <https://doi.org/10.1053/j.ackd.2020.09.003> PMID: 33308501
5. Pei G, Zhang Z, Peng J, Liu L, Zhang C, Yu C, et al. Renal involvement and early prognosis in patients with COVID-19 pneumonia. *J Am Soc Nephrol* 2020; 31: 1157–1165, 2020 <https://doi.org/10.1681/ASN.2020030276> PMID: 32345702
6. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA* 2020; 323: 2052–2059 <https://doi.org/10.1001/jama.2020.6775> PMID: 32320003
7. Sharma P, Uppal NN, Wanchoo R, Shah HH, Yang Y, Parikh R, et al. COVID-19–associated kidney injury: A case series of kidney biopsy findings. *J Am Soc Nephrol* 2020; 31: 1948–1958 <https://doi.org/10.1681/ASN.2020050699> PMID: 32660970
8. Su H, Yang M, Wan C, Yi L-X, Tang F, Zhu H-Y, et al. Renal histopathological analysis of 26 postmortem findings of patients with COVID-19 in China. *Kidney Int* 2020; 98: 219–227, 2020 <https://doi.org/10.1016/j.kint.2020.04.003> PMID: 32327202
9. Hirsch JS, Ng JH, Ross DW, Sharma P, Shah HH, Barnett RL, et al. Acute kidney injury in patients hospitalized with COVID-19. *Kidney Int* 2020; 98: 209–218, 2020 <https://doi.org/10.1016/j.kint.2020.05.006> PMID: 32416116
10. Mohamed MM, Lukitsch I, Torres-Ortiz AE, Walker JB, Varghese V, Hernandez-Arroyo CF. Acute kidney injury associated with coronavirus disease 2019 in urban New Orleans. *Kidney360* 2020; 1: 614–622 <https://doi.org/10.34067/KID.0002652020> PMID: 35372932
11. Fisher M, Neugarten J, Bellin E, Yunes M, Stahl L, Johns TS, et al. AKI in hospitalized patients with and without COVID-19: A comparison study. *J Am Soc Nephrol* 2020; 31: 2145–2157 <https://doi.org/10.1681/ASN.2020040509> PMID: 32669322
12. Moledina DG, Simonov M, Yamamoto Y, Alausa J, Arora T, Biswas A, et al. The association of COVID-19 with acute kidney injury independent of severity of illness: A multicenter cohort study. *Am J Kidney Dis* 2021; 77: 490–499.e1 <https://doi.org/10.1053/j.ajkd.2020.12.007> PMID: 33422598

13. Murugan R, Kellum JA: Acute kidney injury: what's the prognosis? *Nat Rev Nephrol.* 2011; 7(4):209–217 <https://doi.org/10.1038/nrneph.2011.13> PMID: 21343898
14. Hassler L, Reyes F, Sparks MA, Welling P, Battle D. Evidence for and against direct kidney Infection by SARS-CoV-2 in patients with COVID-19. *CJASN* 2021; 16: <https://doi.org/10.2215/CJN.04560421> PMID: 34127485
15. Senger R, Kavuru V, Sullivan M, Gouldin A, Lundgren S, Merrifield K, et al. Spectral characteristics of urine specimens from healthy human volunteers analyzed using Raman Chemometric Urinalysis (Rametrix). *PLoS One.* 2019; 14(9): e0222115. Published online 2019 Sep 27. <https://doi.org/10.1371/journal.pone.0222115> PMID: 31560690
16. Senger R, Sullivan M, Gouldin A, Lundgren S, Merrifield K, Steen C, et al. Spectral characteristics of urine from patients with end-stage kidney disease, analyzed using Raman Chemometric Urinalysis (Rametrix). *PLoS One.* 2020; 15(1): e0227281. Published online 2020 Jan 10. <https://doi.org/10.1371/journal.pone.0227281> PMID: 31923235
17. Carswell W, Guruli G, Tracy A, Xu Y, Du P, Senger R, et al. Raman Spectroscopy as a non-cytological detection and quantification urinalysis method for microhematuria in human urine. *Appl Spectro* 2022. 76(3): 1–11 <https://doi.org/10.1177/00037028211060853> PMID: 35102755
18. Huttanus H, Vu T, Guruli G, Tracey A, Carswell W, Said N, et al. Raman Chemometric Urinalysis (Rametrix) as a screen for bladder cancer. *PLoS One.* 2020; 15(8): e0237070. Published online 2020 Aug 21. <https://doi.org/10.1371/journal.pone.0237070> PMID: 32822394
19. Senger R, Sayed Issa A, Agnor B, Talty J, Hollis A, Robertson JL. Disease-associated multimolecular signature in the urine of patients with Lyme disease, detected using Raman spectroscopy and chemometrics (Rametrix). *Appl Spectro* 2022 76; 284–299 <https://doi.org/10.1177/00037028211061769> PMID: 35102746
20. Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, et al. The human urine metabolome. *PLoS One.* 2013; Sep 4; 8(9):e73076 <https://doi.org/10.1371/journal.pone.0073076> PMID: 24023812
21. Emwas AH, Luchinat C, Turano P, Tenori L, Roy R, Salek RM, et al. Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. *Metabolomics* 2015; 11(4):872–894 <https://doi.org/10.1007/s11306-014-0746-7> PMID: 26109927
22. Castelli FA, Rosati G, Moguet C, Fuentes C, Marrugo-Ramírez J, Lefebvre T, et al. Metabolomics for personalized medicine: the input of analytical chemistry from biomarker discovery to point-of-care tests. *Anal Bioanal Chem.* 2021; 25:1–31 <https://doi.org/10.1007/s00216-021-03586-z> PMID: 34432105
23. Chen Z, Kim J. Urinary proteomics and metabolomics studies to monitor bladder health and urological diseases. *BMC Urol* 2016; 16 (Mar 22):11 <https://doi.org/10.1186/s12894-016-0129-7> PMID: 27000794
24. Cheng S, Shah SH, Corwin EJ, Fiehn O, Fitzgerald RL, Gerszten RE, et al. Potential impact and study considerations of metabolomics in cardiovascular health and disease: A scientific statement from the American Heart Association. *Circ Cardiovasc Genet.* 2017; 10(2):e000032 <https://doi.org/10.1161/HCG.000000000000032> PMID: 28360086
25. Dator R, Villalta PW, Thomson N, Jensen J, Hatsukami DK, Stepanov I, et al. Metabolomics profiles of smokers from two ethnic groups with differing lung cancer risk. *Chem Res Toxicol.* 2020; 33(8):2087–2098 <https://doi.org/10.1021/acs.chemrestox.0c00064> PMID: 32293874
26. Gowda GA, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D. Metabolomics-based methods for early disease diagnostics. *Expert Rev Mol Diagn* 2008; 8(5):617–33 <https://doi.org/10.1586/14737159.8.5.617> PMID: 18785810
27. Turi KN, Romick-Rosendale L, Ryckman KK, Hartert TV. A review of metabolomics approaches and their application in identifying causal pathways of childhood asthma. *J Allergy Clin Immunol.* 2018; 141(4):1191–1201 <https://doi.org/10.1016/j.jaci.2017.04.021> PMID: 28479327
28. Fisher AK, Carswell WF, Athamneh AIM, Sullivan MC, Robertson JL, Bevan DR, et al. The Rametrix LITE Toolbox v1.0 for MATLAB®. *J Raman Spectrosc* 2018; 49: 885–896
29. Senger R, Robertson J. The Rametrix PRO Toolbox V1.0 for MATLAB®. *Peer J.* 2020; 8: e8179. Published online 2020 Jan 6. <https://doi.org/10.7717/peerj.8179> PMID: 31934499
30. Celia-May D, Schmitt M, Popp J. Theoretical principles of Raman spectroscopy. *Phys Sci Rev.* 2019; 19:4
31. Movasaghi Z, Rehman S, Rehman DIU. Raman spectroscopy of biological tissues. *Appl Spectrosc Rev* 2007; 42: 493–541
32. Senger RS, Scherr D. Resolving complex phenotypes with Raman spectroscopy and chemometrics. *Curr Opin Biotechnol* 2020; 66:277–282 <https://doi.org/10.1016/j.copbio.2020.09.007> PMID: 33142112

33. Shen Y, Hu F, Min W. Raman imaging of small biomolecules. *Annu Rev Biophys* 2019; 48:347–369 <https://doi.org/10.1146/annurev-biophys-052118-115500> PMID: 30892920
34. Gardiner D. *Practical Raman Spectroscopy*. NY, Springer-Verlag; 1989
35. Athamneh AIM, Senger RS. Peptide-guided Surface-Enhanced Raman Scattering probes for localized cell composition analysis. *Appl Env Microbiol* 2012; 78: 7805–7808 <https://doi.org/10.1128/AEM.02000-12> PMID: 22923413
36. Athamneh AIM, Alajlouni RA, Wallace RS, Seleem MN, Senger RS. Phenotypic profiling of antibiotic response signatures in *Escherichia coli* using Raman spectroscopy. *Antimicrob Agents Chemother* 2014; 58: 1302–1314 <https://doi.org/10.1128/AAC.02098-13> PMID: 24295982
37. Lussier F, Thibault V, Charron B, Wallace GQ, Masson J-F. Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *Trends Anal Chem* 2020; 124:115796
38. Zu TNK, Athamneh AIM, Wallace RS, Collakova E, Senger RS. Near-real-time analysis of the phenotypic responses of *Escherichia coli* to 1-butanol exposure using Raman spectroscopy. *J Bacteriol*. 2014; 196: 3983–3991 <https://doi.org/10.1128/JB.01590-14> PMID: 25157078
39. Zu TNK, Athamneh AIM, Collakova E, Robertson J, Hawken T, Aardema C, et al. Assessment of *ex vivo* perfused liver health by Raman spectroscopy. *J Raman Spectrosc*. 2015; 46: 551–558
40. Xu Y, Du P, Senger R, Robertson J, Pirkle J. ISREA: An Efficient Peak-Preserving Baseline Correction Algorithm for Raman Spectra. *Appl Spectro*. 2021; 75(1):34–45. <https://doi.org/10.1177/0003702820955245> PMID: 33030999
41. Xu, Y, Du, P, Senger, R, Robertson, J. Sparse logistic regression on functional data. *Stat Interface*, accepted, 2021
42. Liu J, Sun J, Huang X, Li G, Liu B. Goldindec: A novel algorithm for Raman spectrum baseline correction. *Appl Spectrosc* 2015; 69: 834–42 <https://doi.org/10.1366/14-07798> PMID: 26037638
43. Talari A, Movasaghi Z, Rehman S, Rehman I. Raman spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* 2015; 50(1): 46–111
44. Yu S, Bonventre J. Acute kidney injury and progression of diabetic kidney disease. *Adv Chronic Kidney Dis* 2018; 25: 166–180 <https://doi.org/10.1053/j.ackd.2017.12.005> PMID: 29580581