A lipid atlas of the human kidney

Farrow, Melissa A.; Tideman, Léonore E.M.; Migas, Lukasz G.; Yang, Haichun; Rivera, Emilio S.; Romer, Carrie E.; Fogo, Agnes B.; Van de Plas, Raf; Spraggins, Jeffrey M.; More Authors

DOI
10.1126/sciadv.adu3730

Publication date
2025

Document Version
Final published version

Published in
Science Advances

Citation (APA)
Farrow, M. A., Tideman, L. E. M., Migas, L. G., Yang, H., Rivera, E. S., Romer, C. E., Fogo, A. B., Van de Plas, R., Spraggins, J. M., & More Authors (2025). A lipid atlas of the human kidney. *Science Advances*, *11*(24), Article eadu3730. https://doi.org/10.1126/sciadv.adu3730

BIOCHEMISTRY

A lipid atlas of the human kidney

Melissa A. Farrow[1,2]†, Léonore E. M. Tideman[3]†, Elizabeth K. Neumann[1,2]†‡, Lukasz G. Migas[3], Nathan Heath Patterson[1,2]§, Madeline E. Colley[1,2], Jamie L. Allen[2,4], Ellie L. Pingry[2,4], Martin Dufresne[1,2], Haichun Yang[5], Maya Brewer[5], Emilio S. Rivera[1,2]¶, Carrie E. Romer[1,2], Katerina Djambazova[2,4], Kavya Sharman[2,4], Angela R. S. Kruse[2,4], Danielle B. Gutierrez[1,2]#, Raymond C. Harris[5,6,7], Agnes B. Fogo[5,6,7], Mark P. de Caestecker[5], Richard M. Caprioli[1,2,7,8], Raf Van de Plas[1,2,3]*, Jeffrey M. Spraggins[1,2,4,6,9]*

Tissue atlases provide foundational knowledge on the cellular organization and molecular distributions across molecular classes and spatial scales. Here, we construct a comprehensive spatiomolecular lipid atlas of the human kidney from 29 donor tissues using integrated multimodal molecular imaging. Our approach leverages high-spatial-resolution matrix-assisted laser desorption/ionization imaging mass spectrometry for untargeted lipid mapping, stained microscopy for histopathological assessment, and tissue segmentation using autofluorescence microscopy. With a combination of unsupervised, supervised, and interpretable machine learning, the atlas provides multivariate lipid profiles of specific multicellular functional tissue units (FTUs) of the nephron, including the glomerulus, proximal tubules, thick ascending limb, distal tubules, and collecting ducts. In total, the atlas consists of tens of thousands of FTUs and millions of mass spectrometry measurements. Detailed patient, clinical, and histopathologic information allowed molecular data to be mined on the basis of these features. As examples, we highlight the discovery of how lipid profiles are altered with sex and differences in body mass index.

INTRODUCTION

The human kidney is a highly organized organ responsible for filtering waste products from the blood, maintaining ion and fluid balance, releasing hormones that regulate blood pressure, and controlling the production of red blood cells. Filtration and maintenance of ion and water balance are achieved in the nephron, which is vital for maintaining homeostasis. Each human kidney has ~1 million nephrons composed of distinct multicellular functional tissue units (FTUs) surrounded by a network of capillaries. FTUs of the nephron include the glomerulus (GL), where blood filtration and formation of ultrafiltrate are initiated, as well as a series of tubules that are composed of proximal convoluted tubules, the loop of Henle, distal convoluted tubules, and collecting ducts (CDs). Each renal FTU specializes in balancing ion and nutrient concentration and osmolality of the urine by mechanisms of secretion and absorption.

Fundamental to comprehending the architectural and functional complexity of an organ are determining the organization of cell types and FTUs and cataloging the localized molecular profiles of these tissue features. Model systems, such as cell cultures and organoids, have provided insight into aspects of tissue function but not necessarily into the spatial anatomical and cellular organization of tissue. Animal models are better suited to yield the latter but may not fully recapitulate human physiology. Therefore, recently, there has been a push to develop cellular atlases of human tissue instead. Nevertheless, such studies have primarily focused on transcriptional data with limited spatial context (*1*–*3*). To address these gaps, multiple large-scale research consortia have been established to develop and deploy spatial technologies for deep molecular profiling and mapping of the transcriptome, proteome, lipidome, and metabolome of human tissues. Comprehensive spatiomolecular atlases offer a means to generate new hypotheses and advance biomedical research by providing an unprecedented view into tissue at cellular resolution (*4*–*6*). Exploring relationships between cellular and molecular organization of tissues enables the discovery of underlying drivers of functional efficiency, transition to disease, and disease severity as a function of key patient factors (e.g., age, sex, race, and comorbidities). Recently, the National Institutes of Health and private organizations have funded various atlas efforts to address this grand challenge, such as the Human Biomolecular Atlas Program (HuBMAP) (*6*), Human Cell Atlas (*7*), BRAIN Initiative (*8*), Kidney Precision Medicine Project (*9*), and Human Tumor Atlas Network (*10*), targeting an array of normal and/or diseased organs.

Integration of multimodal molecular imaging technologies can provide a more rounded, systems biology view of organ function (and dysfunction) across a wide range of molecular classes (e.g., lipids, metabolites, proteins, and RNA) and spatial scales (e.g., whole organs to single cells). The HuBMAP is a consortium funded by the National Institutes of Health that is leading research efforts in this area with the goal of creating an open, comprehensive molecular atlas of the human body at cellular resolution (*6*). Here, as part of HuBMAP, we have developed an FTU-specific lipidomic atlas of the human kidney to characterize the molecular organization of the nephron in normal-appearing tissue. This atlas incorporates untargeted

[1]Department of Biochemistry, Vanderbilt University, Nashville, TN 37232, USA. [2]Mass Spectrometry Research Center, Vanderbilt University, Nashville, TN 37232, USA. [3]Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, Netherlands. [4]Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN 37232, USA. [5]Division of Nephrology and Hypertension, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA. [6]Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN 37232, USA. [7]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA. [8]Department of Pharmacology, Vanderbilt University, Nashville, TN 37232, USA. [9]Department of Chemistry, Vanderbilt University, Nashville, TN 37232, USA.
*Corresponding author. Email: jeff.spraggins@vanderbilt.edu (J.M.S.); raf.vandeplas@tudelft.nl (R.V.d.P.)
†These authors contributed equally to this work.
‡Present address: Department of Chemistry, University of California Davis, Davis, CA 95616, USA.
§Present address: Aspect Analytics, 3600 Genk, Belgium.
¶Present address: Los Alamos National Laboratory, Los Alamos, NM 87545, USA.
#Present address: Proteome Center, Discovery Life Sciences, Huntsville, AL 35806, USA.

imaging mass spectrometry (IMS) data (*11*, *12*) that are fully integrated with various forms of microscopy (*13*, *14*) to enable the discovery of spatially specific molecular marker candidates for key patient factors such as obesity and sex. Each component of the atlas is publicly available (https://portal.hubmapconsortium.org/). This atlas spans multiple scales of spatial granularity, ranging from large tissue areas including multiple anatomical regions to specific FTUs and cell types and from averaged populations to individual subjects, serving as a resource for exploring the biomolecular landscape that informs renal function and disease. To punctuate the potential of this resource, we demonstrate how our integration of IMS and microscopy can be used to reveal spatiomolecular relationships that are difficult to access through other means and how these unique insights can yield key lipidomic marker candidates for, e.g., obesity for various functional components of the nephron. Although this work is focused on the human kidney, the analytical and computational pipelines we have developed as part of our atlasing efforts have broad utility and are applicable to any organ system or disease type. Together, these developed pipelines provide critical insight into how molecules and cells are organized into multicellular functional components that work in concert to drive macroscopic tissue function.

RESULTS

We have collected nephrectomies from 29 donors (table S1) and analyzed these tissues with a customized workflow that integrates pathological assessment and image segmentation on the basis of autofluorescence (AF) microscopy with high-spatial-resolution matrix-assisted laser desorption/ionization (MALDI) imaging IMS and unsupervised and supervised machine learning (ML) approaches that exploit the spatial co-registration of the different data types (Fig. 1). The spatial match achieved by acquisition of microscopy images and IMS data from the same tissue section outperforms what would be possible with measurements from subsequent tissue sections. Whole slide microscopy images were segmented into GL, proximal tubule (PT), thick ascending limb (TAL), distal tubule (DT), and CD features, and spatially associated untargeted IMS data were mined to discover FTU-specific lipid profiles. This approach provides insight into the FTU-level organization of the kidney at the molecular level. Figure 2 (A.a to A.c) shows example whole slide microscopy images collected from a selected donor (56-year-old white female). AF microscopy (Fig. 2A.a) was used to drive image co-registration and FTU segmentation (Fig. 2A.b). Periodic acid–Schiff (PAS)–stained images (Fig. 2A.c and fig. S1) were collected after MALDI IMS acquisition for histological assessment of overall specimen composition (i.e., % cortex versus medulla), tubular atrophy, glomerular sclerosis, and measures of other key tissue features and pathologies (table S2). AF images (fig. S2), acquired before MALDI IMS, were used to automatically segment the FTUs of the nephron, including GLs, PTs, TAL, DTs, and CDs (fig. S3), as described previously (*15*). MALDI IMS measurement regions were selected to capture a mixture of FTUs and anatomical regions in both positive and negative ion modes (Fig. 2, A.d to A.f). It is noted that both polarities are collected from each tissue section. Examples of IMS-provided lipid distribution images are shown in Fig. 2 (B.a to B.e). All MALDI IMS data in the atlas were collected with a pixel size of 10 μm, which, together with multimodal acquisition from the same tissue section, provided sufficient spatial specificity and accuracy to associate mass spectrometry (MS) signals with specific FTUs in situ. For example, the sphingomyelin SM(34:1);2O [mass/charge ratio (*m*/*z*) 687.545] was found to localize to the GLs and other surrounding structures (Fig. 2B.a). The phosphatidylethanolamine PE(36:4) (*m*/*z* 738.508) localized primarily to the PTs (Fig. 2B.b), whereas PE(36:1) (*m*/*z* 744.555) was found with greater abundance in the CDs (Fig. 2B.c). The substantial variation in distribution between these latter two lipids is particularly notable given that they only differ by the number of double bonds in their fatty acyl tails. Last, the sulfatide SHexCer(42:1);3O (*m*/*z* 906.635) was found specifically in the TAL (Fig. 2B.d). The overlay of these four lipid species (Fig. 2B.e) highlights the localized diversity of these molecules and their ability to differentiate FTUs as compared to complimentary microscopy. Besides a spatial view into a selection of the hundreds of lipid species recorded, the atlas can also yield spectrum-wide molecular signatures for each FTU. One approach is demonstrated in Fig. 2B.f, where cohort-wide average mass spectra specific for each FTU type are provided. Using the automated AF microscopy–based FTU segmentations and the common spatial coordinate system as a guide, MALDI IMS mass spectra (i.e., IMS pixels) specific to each FTU type can be collected at a scale from across all tissues in the atlas. These FTU-specific mass spectra are then combined to provide donor cohort-wide average mass spectra for each FTU in both negative (figs. S4 and S5) and positive (figs. S6 and S7) ion modes. The negative ion mode average mass spectra for each FTU are also shown in Fig. 2B.f for a selected mass window (*m*/*z* 650 to 900). While subtle changes in ion intensity or in the presence of low abundant molecular species can be difficult to discern in average spectra, the manner in which average spectra diverge between FTUs is more readily apparent when these data are shown as difference spectra (figs. S8 to S27). For example, the difference spectrum comparing GLs to PTs in negative ion mode shows variations in lipid ion intensities between these two critical structures of the nephron more clearly (Fig. 2B.g).

Characteristics of the donor tissue cohort

The molecular atlas was constructed with normal portions of fresh-frozen renal cancer nephrectomy tissue (i.e., remnant tissue) collected from 29 human donors (table S1). All tissues were collected through the Cooperative Human Tissue Network at Vanderbilt University Medical Center. Warm and cold ischemia times were shorter than 5 and 30 min, respectively. Figure 3A highlights the distributions of sex, age, and body mass index (BMI) of the donor cohort. The cohort includes 14 female and 15 male donors with an age distribution ranging from 20 to 78 years and an average age of 60. The BMI values of the included donors were similar to averages for the population of the United States (*16*), ranging from 22.3 to 45.5 with an overall average of 30.6, and averages of 31.8 and 29.5 for female and male donors, respectively. It is noted that the cohort used for this study is limited on the basis of tissues available and currently only includes non-Hispanic white individuals. This limitation will be specifically addressed in future versions of the atlas.

Although it is difficult to obtain tissues from completely normal individuals and most adult tissues display some pathological features, histopathological analysis by expert clinical pathologists determined that all tissues included in the atlas were age-appropriate normal tissues (table S2). PAS-stained images show no clear indication of cancer or immune cell infiltration. Tissues were assessed for anatomical composition (i.e., percentage of cortex and medulla) to ensure that all key renal FTUs are represented. Tissue normalcy was

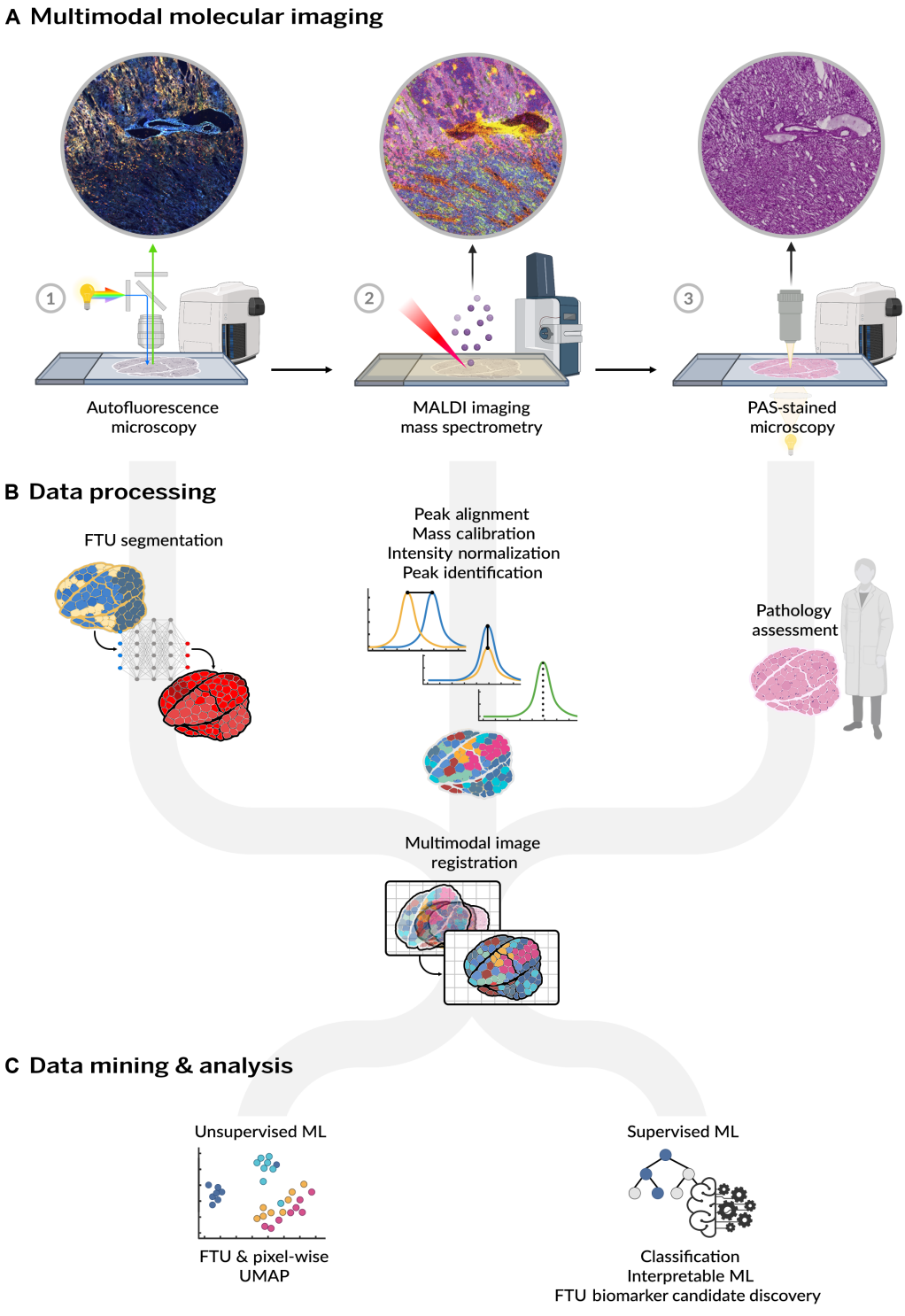

Fig. 1. Construction of the FTU lipid atlas of the human kidney. Multimodal molecular imaging data were collected from 29 donor kidney tissues. Each tissue section was subjected to AF microscopy, MALDI IMS, and PAS-stained microscopy, in sequential order (**A**). Each modality is processed individually to provide AF-driven FTU segmentations, to ensure that MALDI IMS measurements are comparable and to remove potential batch effects, and to provide histopathological assessment of each tissue (**B**). These datasets are then integrated by spatially co-registering them onto the same spatial coordinate system and by performing a combination of unsupervised and cross-modal supervised machine learning (ML) analyses (**C**). Interpretable machine learning is then used to uncover spatially distinct biomarker candidates for FTUs across the overall data cohort as well as scoped to specific donor metadata such as BMI or sex.

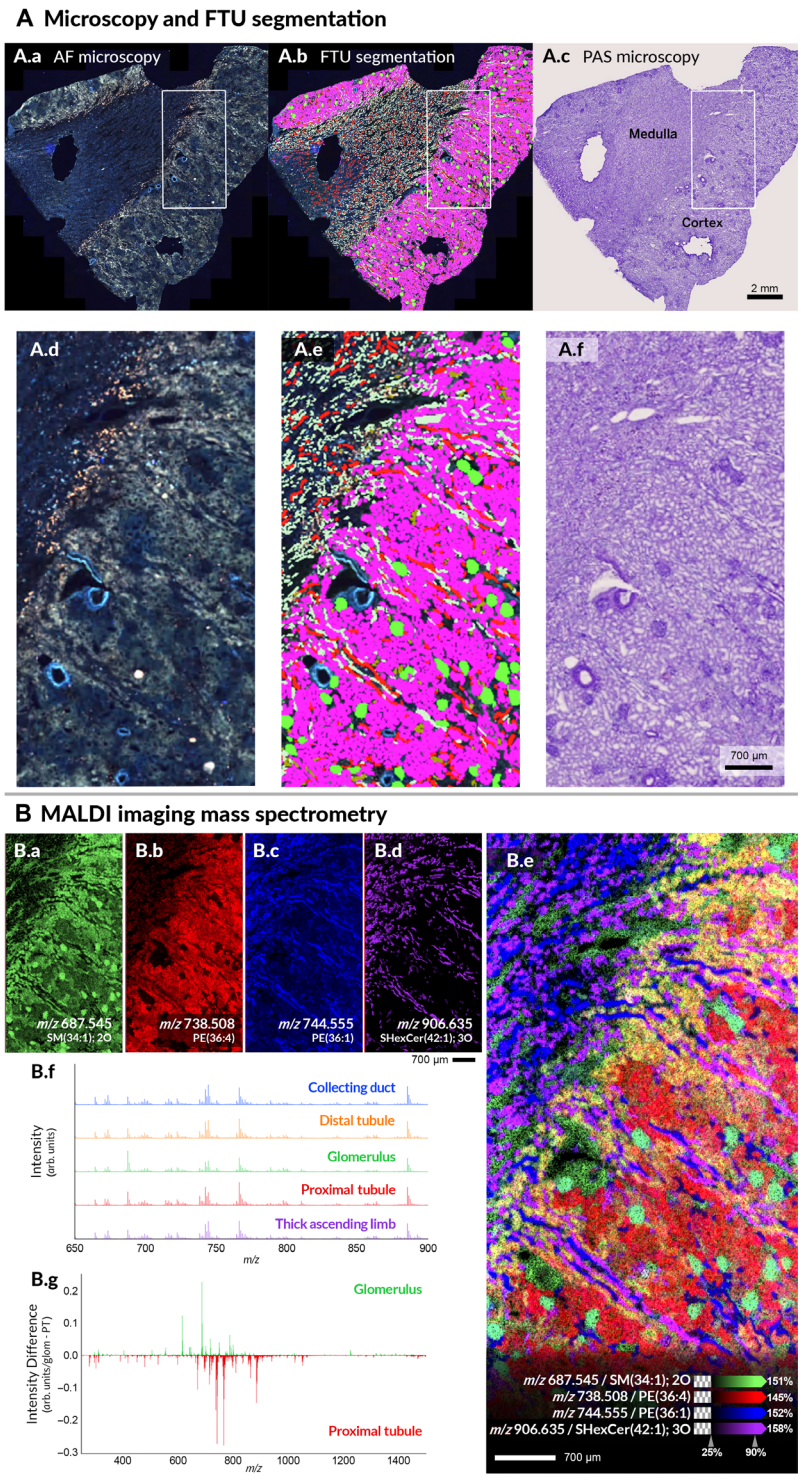

Fig. 2. Example of multimodal molecular characterization of human kidney tissue. Whole slide microscopy images from donor VAN0028 (56-year-old white female) were collected using AF (**A.a** and **A.b**) before high-spatial-resolution (10-µm pixel size) MALDI IMS measurement, and PAS-stained microscopy (**A.c**) data were acquired post-IMS. AF microscopy data were automatically segmented into renal FTUs (A.b), including the GLs (green), PTs (magenta), TAL (light green), DTs (brown), and CDs (red). MALDI IMS measurement regions (white boxes) were selected to include a mixture of tissue features. The microscopy data for the MALDI IMS measurement region are highlighted in (**A.d**) to (**A.f**). Selected individual molecular distribution images from the negative ion mode MALDI IMS measurement and an overlay image are provided in (**B.a**) to (**B.d**) and (**B.e**), respectively. The selected ions demonstrate unique localizations within the kidney without the need for prior labeling, and these are just four of the hundreds of lipids that make up this molecular atlas. The mean mass spectrum associated with each FTU, obtained by averaging FTU-specific IMS-pixels across all donors, displays subtle differences in the lipids detected and their intensities (**B.f**). It is noted that full size versions of the spectra can be found in the Supplementary Materials. Variations in intensity profiles of FTUs are more evident using difference spectra, as shown in the comparison between the normalized (between 0 and 1 to allow for direct comparison) average spectrum of PTs subtracted from that of the GLs (**B.g**).

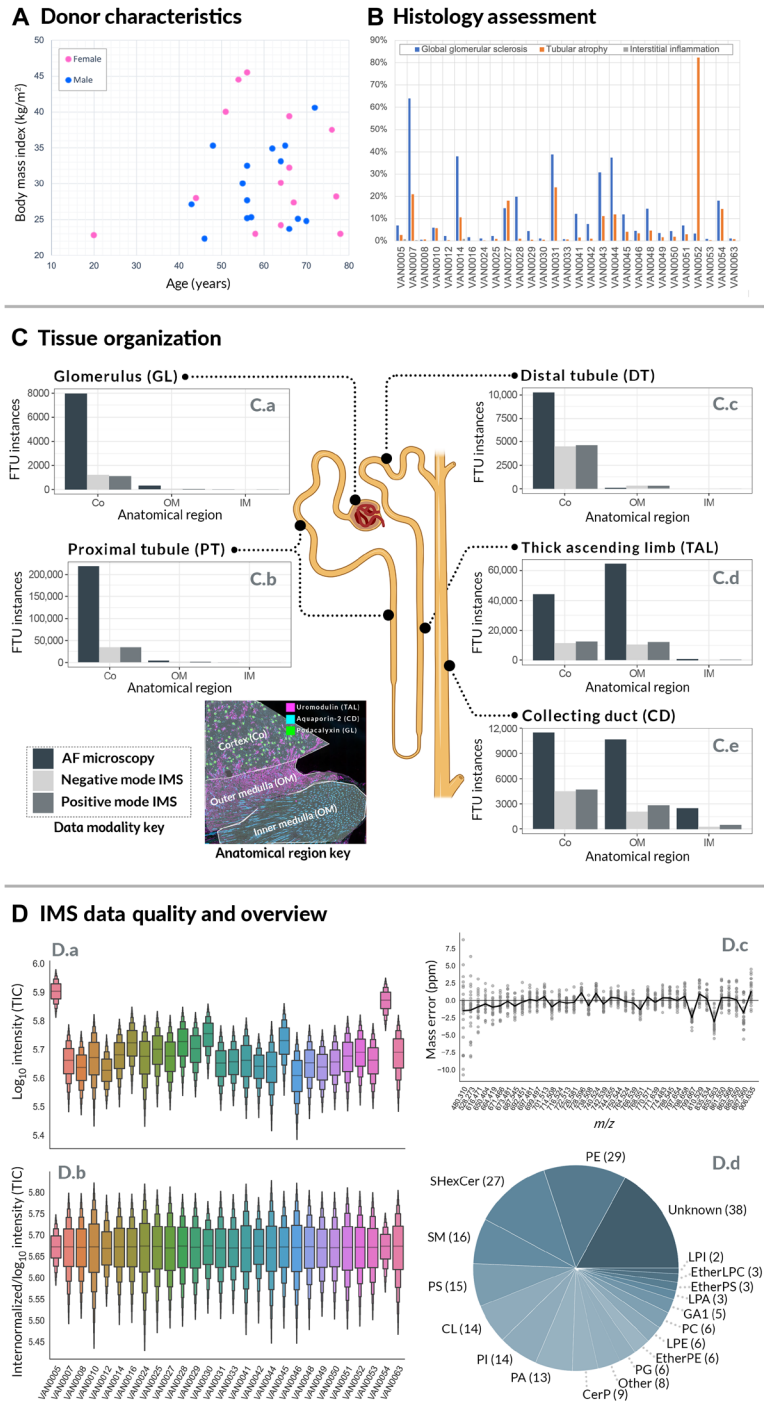

Fig. 3. Global characteristics of the atlas. Key donor characteristics such as age and BMI (**A**) and detailed histopathology (**B**) are available for 29 donors. Selected measures of tissue normalcy are highlighted in the bar plot, including % global glomerular sclerosis, % interstitial fibrosis tubular atrophy, and % interstitial inflammation. AF microscopy was used to comprehensively segment renal FTUs such as the GL (**C.a**), PT (**C.b**), DT (**C.c**), TAL (**C.d**), and CD (**C.e**). The total number of detected instances for each FTU was quantified within the cortex (Co), outer medulla (OM), and inner medulla (IM) of the kidney across all AF whole slide images (dark gray) and specific to the MALDI IMS measurement regions in both negative (light gray) and positive ion (medium gray) modes. The example immunofluorescence data show how markers used to train the AF-based segmentation algorithms were also able to differentiate the broader anatomical zones of the kidney. The integrated MALDI IMS data underwent data preprocessing to address nonbiological variability, including peak alignment, calibration, and intensity normalization. Boxen plots show the variability of the TIC (log_{10} intensity) for the negative ion mode data following intrasample normalization (**D.a**) and the consistency after intersample normalization (**D.b**). (**D.c**) Mass error in parts per million for selected negative ion mode lipids. The black line represents the mean mass error from all pixels collected from all samples, and the gray dots represent the spread of the data for each *m/z*. Following preprocessing, *m/z* features are annotated using mass accuracy to compare to LC-MS/MS–based identifications and on-tissue fragmentation. The provided pie chart summarizes the number of annotations for various lipid classes from the negative ion mode data (**D.d**). It is noted that 28 of the 29 samples were analyzed in negative ion mode.

determined using measures of interstitial fibrosis tubular atrophy, arteriosclerosis, global glomerular sclerosis, arteriolar hyalinosis, and interstitial inflammation. The frequency of key pathologies determined for each donor tissue is summarized in Fig. 3B. As expected, the levels of tissue pathologies vary greatly in human subjects, yet our analytical and computational pipelines were still able to uncover lipidomic profiles for specific FTUs that were consistent across the entire atlas donor cohort.

FTU frequency and distribution

Multimodal analysis of renal tissue provides a means of connecting rich molecular information from IMS with relevant tissue cell types and structures when integrated with complementary microscopy and the derived FTU segmentation masks (fig. S3). By applying the multimodal collection pipeline to 29 donors, FTU segmentation labels and molecular profiles (i.e., mass spectra) on the order of millions of data points are obtained and tied to specific spatial tissue areas with high fidelity. Figure 3C shows the number of instances (discrete FTUs) found across the AF microscopy images overall as well as specific to the MALDI IMS measurement regions. Each donor sample has similar numbers of FTU instances per class, but relative counts are influenced by the gross morphology of each section, which was calculated as the percent of cortex or medulla in a section during pathological assessment (*15*) (table S2). FTU instances were distributed as expected, with GLs and PTs almost exclusive to the cortex (Fig. 3, C.a and C.b, respectively), TAL segments found in the outer medulla and cortex (Fig. 3C.d), DTs in the cortex (Fig. 3C.c), and CDs found both in medullary and cortical regions (Fig. 3C.e). In terms of the robustness of the atlas and the FTU-specific molecular signatures provided by MALDI IMS, it is critical to note that the analyses highlighted herein are based on thousands of individual FTU instances for each targeted kidney structure, together with millions of MS measurements that are co-registered and integrated with them. For example, positive ion mode MALDI IMS measurements from all 29 donors amassed 6,779,166 mass spectra (individual IMS pixels). From this, ~68.8% (4,662,712 pixels) were spatially annotated into the five targeted FTUs: GLs, PTs, TAL, DTs, and CDs. Similarly, in negative ionization mode, 6,568,017 mass spectra were collected from 29 of the donor samples, with 70.8% assigned to FTUs. The fraction of MALDI IMS data that could be labeled as part of targeted FTUs is as expected. The unlabeled fraction of the data is partially due to how the segmentation models were constructed, namely with an emphasis on avoiding false positives. Given the number of instances available, for integration with MALDI IMS data, it is less important to detect every last FTU than it is to be sure that the FTU label assignments are of high confidence. Therefore, the resulting segmentation masks label rather conservatively and only high-confidence FTUs so that only high-quality data are included in subsequent IMS data mining and modeling approaches. Furthermore, the current version of the kidney lipid atlas does not include tissue features such as vasculature and extracellular matrix, which also account for a nontrivial tissue area. More granular delineation of the nephron and the addition of these other key tissue features will be the focus of future extensions of the atlas.

IMS data quality and annotations

Ensuring MALDI IMS data quality was essential when constructing the kidney FTU lipid atlas. Extensive quality control procedures and data preprocessing were performed to minimize nonbiological variability and to ensure comparability of the data. Peak alignment and calibration were critical for enabling high-confidence annotation of IMS data. IMS peak intensity normalization was performed both on a per-sample basis (i.e., intrasample normalization) and across the entire data cohort (i.e., intersample normalization). Following intrasample normalization, the sample-level intensity variance across the entire cohort was ±8% (Fig. 3D.a and fig. S28) and ±9% (fig. S31) for negative and positive ion modes, respectively. This was improved upon further with intersample normalization reducing the global intensity variance for negative ion mode to ±2% (Fig. 3D.b and fig. S28) and ±3% (fig. S31) for positive ion mode. A combination of liquid chromatography–tandem MS (LC-MS/MS) and on-tissue fragmentation (table S4) was used to identify lipids from serial tissue sections from each of the donor samples. Although thousands of MS features were detected, the current atlas is built around 212 (negative ion mode) and 211 (positive ion mode) peaks that were either annotated or that remain unidentified at the moment but that do exhibit a high signal-to-noise ratio and that are clearly distinguishable from the isotopic envelopes of other annotated molecular species. All IMS peak annotations were made on the basis of mass accuracy with a threshold of less than 5 parts per million (ppm) to link to the theoretical m/z of the lipids identified by MS/MS. Plots are provided to highlight the consistency of the data, showing the spread of the mass error for selected negative (Fig. 3D.c and fig. S29) and positive (fig. S30) ion mode lipids. The MALDI IMS analysis detected a wide range of lipid classes. Negative ion mode data included annotations and images from phosphatidylethanolamines (PEs), sulfatides [sulfatidyl hexose ceramides (SHexCers)], sphingomyelins (SMs), phosphatidylserines (PSs), cardiolipins (CLs), phosphatidylinositols (PIs), and others (Fig. 3D.d and fig. S32A). In positive ion mode, the atlas includes lipid classes such as phosphatidylcholines (PCs), SMs, ether-linked PCs (etherPC), and nonpolar lipids such as diglycerides (DGs) and monoglycerides (MGs), among others (fig. S32B). Full datasets are available as part of the atlas, and additional features can be added to future analyses if other ions are identified or found to be important.

Unsupervised analysis

IMS measurements were acquired in both negative and positive ionization modes, yielding two data cohorts. All mass spectra within a cohort (one mass spectrum per IMS pixel) have been preprocessed to remove nonbiological variation where possible and to enable ion intensity values to be compared across experiments. This preprocessing phase includes m/z alignment, m/z calibration, ion intensity normalization, and peak picking, and it is complemented with sample-specific variation removal by means of reComBat (*17*). Each cohort dataset consists of all IMS pixels (i.e., spectra) collected across donor tissue samples and uses the same set of peak-picked features throughout. Specifically, the negative ionization mode cohort dataset entails 6,568,017 pixels, each reporting a vector of 212 features. The features correspond to peak intensities of a selection of 212 lipid species (table S4) that were chosen for identifiability and broad coverage from all peaks reported by the full profile mass spectra acquired.

Initial insight into the variation captured by this dataset was obtained through an unsupervised machine learning approach, uniform manifold approximation and projection (UMAP) (*18*), with a focus on exploring the structure underlying the measurements. This resulted in the dimensionality-reduced representation of the spectral data in Fig. 4. The latent space representation of

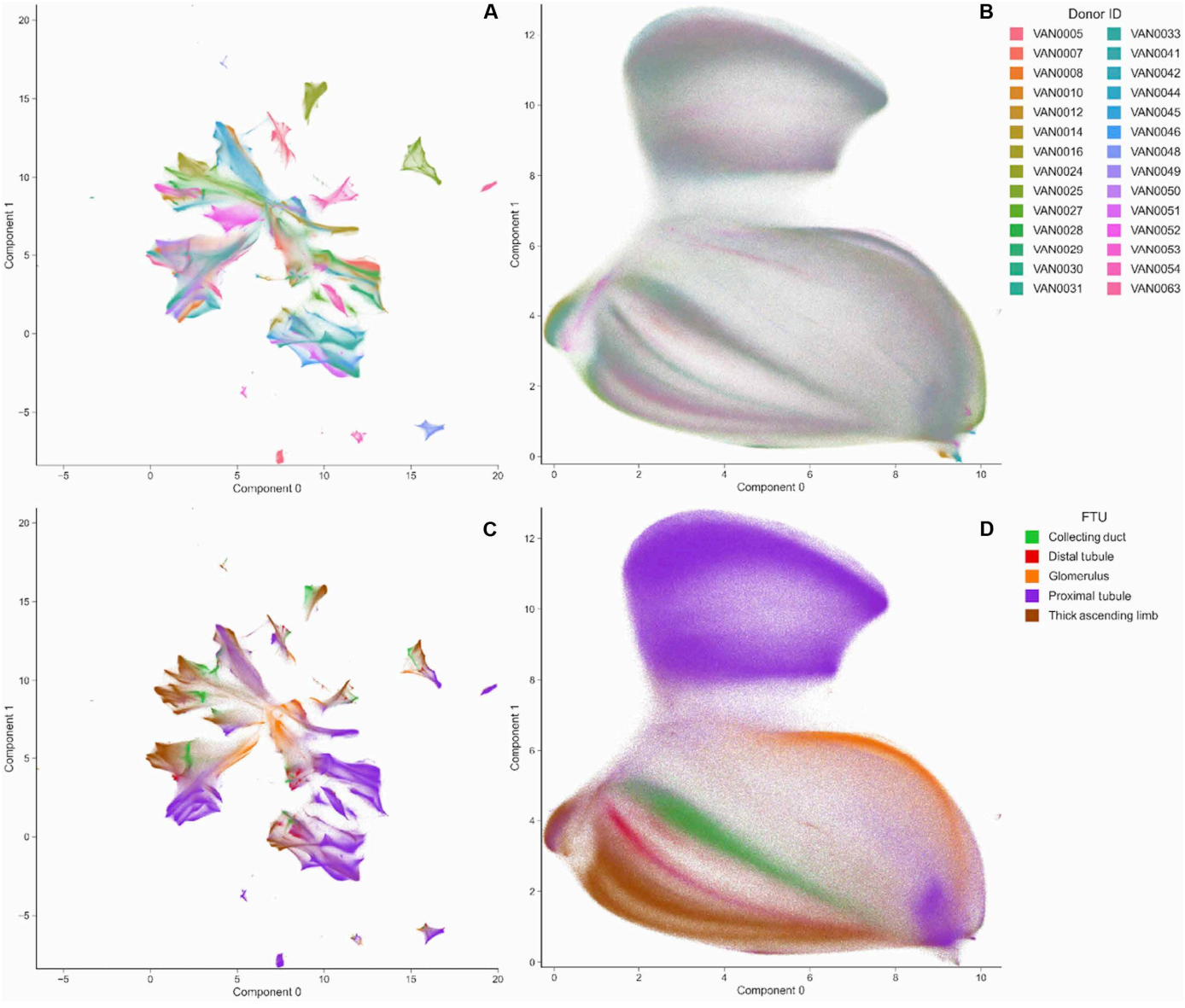

Fig. 4. Pixel-level chemical variation of MALDI IMS measurements across 28 of the donors analyzed by negative ion mode. Two-dimensional visualization of chemical variation in the negative ionization mode experiment cohort using UMAP to cast a matrix of 6,568,017 observations (i.e., IMS pixels across 28 donor tissues) by 212 features (i.e., lipid species) into a table of 6,568,017 observations by two latent variables while retaining neighborhood relationships between observations as captured by a cosine distance measure. (**A**) Latent space representation of chemical variation after preprocessing, with pixels color coded for donor origin. (**C**) Same latent space representation as in (A), with pixels color coded for FTU type (as automatically recognized from microscopy). (**B**) Latent space representation of chemical variation after preprocessing and reduction of donor variation by reComBat, with pixels color coded for donor origin. (**D**) Same latent space representation as in (B), with pixels color coded for FTU type (as automatically recognized from microscopy). Note that while reComBat has not been optimized for use on MS data, it is applied here to demonstrate that if sample-specific variation can be removed, FTU-related variation becomes more readily discernable in an unsupervised context.

chemical variation (irrespective of donor origin) is shown color coded for the donor origin (Fig. 4A) and for the FTU type (Fig. 4C). Using the cosine distance, this visualization captures primarily overall variation across all 212 measured features. The presence of sample-specific variation is apparent in Fig. 4A, where certain sample-specific measurement sets tend to cluster away from other measurements. However, Fig. 4C shows that besides sample-specific variation, FTU-specific variation is also clearly present in the data. Specifically, subsets of measurements that share FTU membership tend to cluster together regardless of their donor origin (e.g., note PT concentration in the bottom right corner of Fig. 4C). Figure 4 (A and C) suggests that the measurements collected as part of this atlas effectively capture different types of biological variation. Furthermore, because a rather low-dimensional representation is pursued here, it is expected that the resulting manifold tends to be dominated by the more prominent variabilities captured by the data, which include sample-related and FTU-related chemical variation.

To further our unsupervised exploration toward FTU-specific chemical differentiation, we removed a substantial part of the sample-related variation using the batch normalization method reComBat (*17*) before conducting a secondary UMAP analysis with identical hyperparameters. Although reComBat has not been optimized for use on MS data, it is applied here as a proof-of-concept demonstration of how, in an unsupervised context, FTU-related variation becomes more readily discernable when at least part of the sample-specific variation can be reduced. This effect is apparent in Fig. 4 (B and D). While Fig. 4 (A and B) shares the sample-origin color-coding, in panel (A), certain sample-specific subsets are clearly separated, and in panel (B)'s visualization, it is harder to discern sample-related subsets. This suggests that latent space sample-related variation has been relatively reduced from Fig. 4 (A and B). In Fig. 4 (C and D), the FTU-type color-coding suggests that as sample-related variation is reduced, the FTU-related variation in these measurements tends to come more to the foreground in the chemical variation reported by the cohort dataset. The visualization in Fig. 4D demonstrates that on the basis of the 212 lipid species tracked in this dataset, the 302,142 GL measurements acquired across all tissue experiments (in green) tend to group together, which suggests that these data capture genuine biological variation across different human donors. This is further supported by 462,935 CD pixels (in orange), 253,961 DT pixels (in red), 2,575,639 PT pixels (in purple), and 1,049,658 TAL pixels (in brown) clustering together, regardless of donor. Similar groupings can be observed whether the analysis is done on a per-pixel basis (Fig. 4 and fig. S33) or whether a mean spectrum is calculated for each FTU instance and the unsupervised analysis is performed at the FTU instance level (fig. S34). These observations suggest that the data collected as part of this atlas reports several biologically relevant underlying trends. Some of these chemical variations are sample specific, while other trends are sample agnostic and hold up across the cohort irrespective of the particular donor sampled, supporting the biological validity of the atlas dataset.

Similar observations on chemical variation can be discerned in the positive ionization mode IMS data (Fig. 5 and figs. S35 and S36). This cohort dataset comprises 6,779,166 pixels, each reporting a vector of 211 features that report the peak intensities of a selection of 211 lipid species (table S4). Figure 5 shows that the chemical variation captured by the data tends to group measurements according to the FTU type they represent, regardless of the donor origin, and their chemical differentiation between FTUs. Figure 5D depicts the FTU instance–level grouping of 290,321 GL mean spectra (orange), 524,092 CD mean spectra (green), 271,080 DT mean spectra (red), 2,557,196 PT mean spectra (purple), and 1,020,023 TAL mean spectra (brown). In figs. S33 to S36, we provide colorings for metadata such as sex and BMI in addition to the FTU-type and donor origin labels shown here. However, those visualizations show less clear separation, which is possibly the result of the two-dimensional representation acquired through UMAP being too low-dimensional to accurately represent and separate all these different variation types in one cohesive two-dimensional space. While this is a testament to the richness of the atlas cohort datasets, it also highlights some of the limitations that come with open-ended exploration by unsupervised machine learning for these data and it hints at the need for deeper and more targeted analyses to discern how chemical variation relates to particular donor metadata. To this end, in the next section, a more in-depth and targeted supervised machine learning approach using interpretable machine learning concepts is used in combination with this atlas to uncover potential relationships that are hard to discern through unsupervised means.

Supervised analysis and spatially driven discovery of biomarker candidates

While informative, the unsupervised machine learning analysis of the atlas tends to show the dominant types of variation present in the datasets. This is not necessarily the chemical variation most relevant to certain questions of biological interest. For example, investigations into which lipid species are specific to certain FTUs or which species vary with sex or other donor characteristics require a more narrowly focused, in-depth analysis. To this end, we have developed an interpretable supervised machine learning workflow that enables automated discovery of which molecular species, among the hundreds tracked, are potential biomarker candidates for a specific FTU or which species seem to be informative when differentiating sex or other donor characteristics (*19*). Unlike unsupervised methods, supervised machine learning algorithms are guided by a specific prediction objective and they therefore tend to disregard data variation that is not relevant to the defined classification task. This ability to focus on variation that aids the recognition task at hand, while attenuating other types of variation, makes supervised machine learning—particularly the training of classification models—an effective means of filtering high-dimensional data down to a more compact set of features relevant to a certain biological question of interest. With a deepening of the molecular mapping of the kidney in mind, the following classification tasks were defined and models were built: recognition (one-versus-all classification) of five FTUs, namely GLs, PTs, TAL, DTs, and CDs, with the goal of obtaining lipid marker candidates for each FTU type, binary classification of female versus male donors with the aim of discerning sex-related marker candidates, and binary classification of normal-BMI ($22 < BMI < 25$) versus high-BMI ($BMI > 35$) donors directed at suggesting lipid marker candidates that differentiate these two classes. The interpretability method used, Shapley additive explanations (SHAP) (*20*), is used in each classification task to quantitatively estimate, for each of the hundreds of IMS-detected lipid species, the relevance of a particular molecular species for recognizing the target biological class. This provides, for each task, a ranked list of each molecular species' relative predictive importance for that task, offering an experiment-wide heuristic for a lipid's biomarker potential in a spatially driven manner. While a lipid's relevance in a classification task aimed at recognizing a certain FTU does not necessarily imply that it is a genuine biomarker for that FTU, such a relevance heuristic can help suggest which out of hundreds of lipid species have a potential relationship to the FTU and could merit further investigation. The automated process furthermore avoids the human bias and drift involved with performing such a task manually, and it can be performed at a scale across a whole cohort of datasets, increasing the robustness of the suggested marker candidate panels. For a given classification task, our SHAP workflow quantifies the importance (as reported by a SHAP value) of each molecular species, reporting its donor cohort-wide relevance (i.e., averaged across all donors), its tissue sample-wide relevance (i.e., specific to a donor but averaged across that donor's tissue sample), and its pixel-specific relevance (i.e., specific to a location within a tissue sample). Because the cohort-wide SHAP importance scores report marker candidate potential across many donors, these scores are more robust cues than

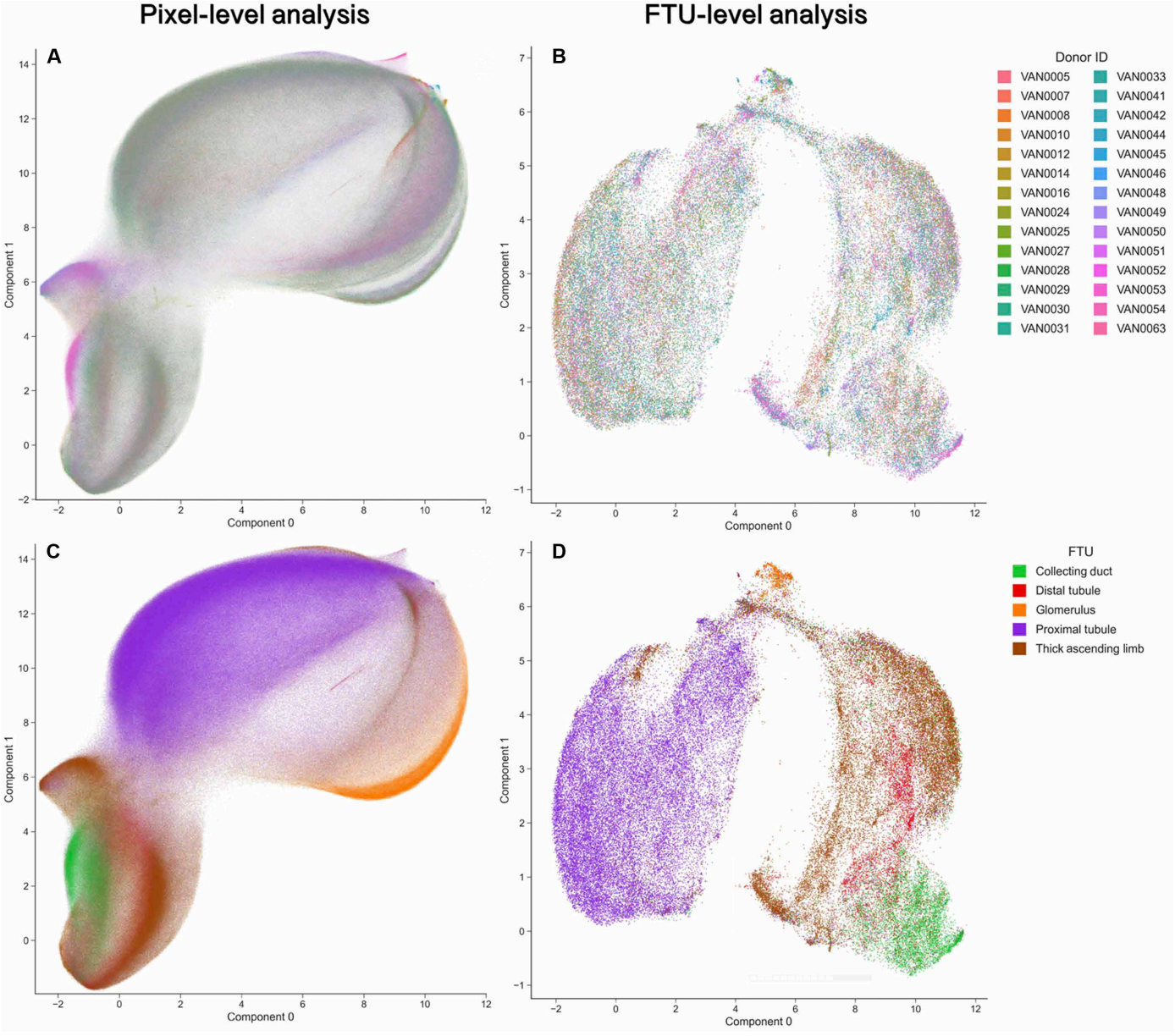

Fig. 5. Pixel-level and FTU instance–level chemical variation in positive ionization mode of IMS measurements across 28 donors. Note that, for consistency, only the 28 samples that had been analyzed by negative ion mode as well were included in this visualization. (**A** and **C**) Two-dimensional visualizations of chemical variation in the positive ionization mode experiment cohort using UMAP to cast a matrix of 6,779,166 observations (i.e., IMS pixels across 28 donor tissues) by 211 features (i.e., lipid species) into a table of 6,779,166 observations by two latent variables while retaining neighborhood relationships between observations as captured by a cosine distance measure. (FTU instance–level column) Two-dimensional visualizations of chemical variation in the positive ionization mode experiment cohort using UMAP to cast a matrix of 75,846 FTU instances (i.e., mean spectrum per FTU instance found across 28 donor tissues) by 211 features (i.e., lipid species) into a table of 74,959 observations by two latent variables while retaining neighborhood relationships between observations as captured by a cosine distance measure. (A) Pixel-level latent space representation of chemical variation after preprocessing and reduction of donor variation by reComBat, with pixels color coded for donor origin. (C) Same pixel-level latent space representation as in (A), with pixels color coded for FTU type (as automatically recognized from microscopy). (**B**) FTU instance–level latent space representation of chemical variation after preprocessing and reduction of donor variation by reComBat, with pixels color coded for donor origin. (**D**) Same FTU instance–level latent space representation as in (B), with pixels color coded for FTU type (as automatically recognized from microscopy). Note that while reComBat has not been optimized for use on MS data, it is applied here to demonstrate that if sample-specific variation can be reduced, FTU-related variation becomes more readily discernable in an unsupervised context.

sample-specific estimates. We therefore use the cohort-wide SHAP scores to extract from the atlas subsets of highly discriminative molecular species that may be biologically relevant marker candidates for each specific FTU, sex, and BMI category. It is noted that the selected comparisons (i.e., FTU, sex, and BMI classifications) are just example comparisons that can be made with this atlas. The assembled and integrated data that provide the foundation of the atlas are a valuable resource that can be mined further to ask other biologically relevant questions and to include additional molecular species detected beyond the 212 (negative ion mode) and 211 (positive ion mode) lipids that have been annotated and examined here.

Lipid profiles of renal FTUs

In figs. S37 to S46, for each FTU, the top 20 cohort-wide SHAP importance scores and the corresponding suggested top 20 biomarker candidates for a specific FTU are provided, both in negative and positive ionization modes. While these plots give a cohort-wide overview of a lipid's relevance to an FTU, an analysis of marker potential down to the sample-specific level is also provided for both modes in the form of bubble plots (fig. S47, negative ion mode; fig. S48, positive ion mode). For simplicity, selected ions that were found to be defining molecular features of one or multiple FTUs are represented in a summary bubble plot in Fig. 6, combining both polarities. The bubble plot data representations provide an effective means of visualizing complex molecular species-FTU relationships in a way that makes trends for specific lipid species across the donor cohort very apparent. Molecular species that are consistent, strong marker candidates for a specific FTU across all donor samples are represented by columns of large bubbles in the chart, and the color of the bubbles indicates whether the molecular marker candidate's signal intensity is positively (red) or negatively (blue) correlated with the FTU. Ion images for the top 10 molecular markers for each FTU for both negative and positive ion modes are provided in figs. S49 to S148.

The most robust glomerular molecule detected is SM(34:1;2O) having the highest SHAP scores in both negative (*m*/*z* 687.545) (Fig. 6B and figs. S37, S47, and S49) and positive ion modes (*m*/*z* 703.575, H^+ adduct) (Fig. 6A and figs. S38, S48, and S59). While previous work has tentatively identified SM(34:1;2O) as having glomerular localization (*21*), we have annotated this lipid by on-tissue MS/MS and correlative LC-MS/MS to confirm its identity as SM(34:1;2O) (Fig. 6B and figs. S37, S47, and S49) and further classified it as SM(18:1;2O/16:0) (Fig. 6A and figs. S38, S48, and S59). Other positive ion mode adducts of SM(18:1;2O/16:0) (Na^+: *m*/*z* 741.531; K^+: *m*/*z* 725.557) were detected with correlation to GLs (Fig. 6A and figs. S38, S48, and S63). SHAP analysis also revealed a negative correlation of SM(34:1;2O) with the TAL (*m*/*z* 687.545), DTs (*m*/*z* 666.434, *m*/*z* 687.545, and *m*/*z* 703.575) and CDs (*m*/*z* 687.545 and *m*/*z* 703.575) (Fig. 6, A and B, figs. S41 and S43 to S46). SM species are enriched in GLs and PTs (Fig. 6, A and B, and figs. S47 and S48), but the distinct spatial distributions of SM(34:1;2O) to GLs (Fig. 6, A and B, and figs. S49, S59, and S63) indicate a functionally specific localization. In addition, we found multiple gangliosides to be specific to GLs, including GA1(18:1;2O/16:0) (*m*/*z* 1225.743) and GA1(18:1;2O/24:1) (*m*/*z* 1335.852) (Fig. 6B and figs. S37, S47, S50, and S54).

Our analysis of the lipid atlas showed that the PT was the most SM-enriched FTU, with nine SM species detected and identified as having a high positive association with PTs based on cohort-wide SHAP importance scores (negative mode: *m*/*z* 769.623 and *m*/*z* 797.654; positive mode: *m*/*z* 701.559, *m*/*z* 729.590, *m*/*z* 757.622, *m*/*z* 785.653, *m*/*z* 811.666, *m*/*z* 813.684, and *m*/*z* 835.666) (Fig. 6, A and B, and figs. S39, S40, S47, S48, S74, S75, S80, S81, S83, S84, S86, and S88). Of those, eight were exclusively discriminative for PTs (negative mode: *m*/*z* 769.623 and *m*/*z* 797.654; positive mode: *m*/*z* 701.559, *m*/*z* 729.590, *m*/*z* 785.653, *m*/*z* 811.666, *m*/*z* 813.684, and *m*/*z* 835.666). The acyl chain identity for five of the eight SM species was revealed to contain long-chain fatty acids (>C18 up to C26 in length) (*m*/*z* 729.590, *m*/*z* 757.622, *m*/*z* 785.653, *m*/*z* 813.684, and *m*/*z* 835.666).

In addition to an SM-rich environment in PTs, phosphatidylserine PS(38:4) (*m*/*z* 810.529) was identified as being a highly discriminative marker for PTs (Fig. 6B and figs. S39, S47, and S73). Also, of the 11 polyunsaturated PEs detected in the atlas (*m*/*z* 528.273, *m*/*z* 556.304, *m*/*z* 714.508, *m*/*z* 722.513, *m*/*z* 738.508, *m*/*z* 750.544, *m*/*z* 762.508, *m*/*z* 764.524, *m*/*z* 766.539, *m*/*z* 778.576, and *m*/*z* 816.530), four were measured with cohort-wide SHAP importance scores that were highly associated with PTs. These are PE(21:4) (*m*/*z* 528.273), PE(36:4) (*m*/*z* 738.508), PE(38:4) (*m*/*z* 766.539), and PE(38:6) (*m*/*z* 762.508) (Fig. 6B and figs. S39, S47, S69, S71, and S76). Of those, PE(38:4) has been shown to directly stimulate Na^+- and K^+-dependent adenosine triphosphatase (Na^+/K^+ ATPase) pump activity (*22*), and this lipid was found to be a notable biomarker candidate of PTs.

The kidney is rich in sulfatides, a class of glycosphingolipids that contain a ceramide bound to a sulfated carbohydrate. Within the kidney, SHexCer are strongly associated with the TAL (Fig. 6B and figs. S41 and S47). In the TAL, six of the seven SHexCer species (*m*/*z* 778.514, *m*/*z* 876.624, *m*/*z* 878.603, *m*/*z* 892.619, *m*/*z* 904.619, and *m*/*z* 906.635) detected had positive discriminatory values for this FTU (fig. S41). Furthermore, among the atlas-measured species, the SHexCers detected at *m*/*z* 876.624, *m*/*z* 892.619, *m*/*z* 904.619, and *m*/*z* 906.635 are the most robustly associated TAL biomarker candidates detected (Fig. 6B and figs. S41, S47, and S89 to S92). In positive ion mode, of the five species with the highest cohort-wide SHAP importance scores (*m*/*z* 758.569, *m*/*z* 760.585, *m*/*z* 801.684, *m*/*z* 811.666, and *m*/*z* 813.684) (fig. S42), none are unique to TAL and they were not considered as discriminatory biomarker candidates of this FTU specifically (Fig. 6A and fig. S48). These observations suggest that the SHexCer signature in the TAL represents a defining feature for this FTU.

The most robust biomarker candidates for DTs include lysophosphatidylcholine (LPC) (*m*/*z* 496.340 and *m*/*z* 522.355) in Fig. 6A (and figs. S44, S48, S119, and S120) and lysophosphatidylethanolamine (LPE) (*m*/*z* 480.310) in Fig. 6B (and figs. S43, S47, and S109). Lysophospholipids are present at low cellular concentrations. However, both LPC(16:0) (*m*/*z* 496.340; fig. S119) and LPE(18:0) (*m*/*z* 480.310; fig. S109) are discriminative for DTs and have the highest cohort-wide SHAP importance scores in positive and negative ion modes, respectively. In addition, LPE(18:1) (*m*/*z* 478.294) and LPE(20:1) (*m*/*z* 506.325) are DT biomarker candidates (figs. S47 and S115). Various SMs (*m*/*z* 731.60, *m*/*z* 801.68, and *m*/*z* 759.63) and SHexCers (*m*/*z* 906.63, *m*/*z* 924.63, and *m*/*z* 890.63) also have high cohort-wide SHAP importance scores for this FTU, but none of these are exclusive to DTs (Fig. 6, A and B, and figs. S47 and S48). The prevalence of lysolipids is a defining feature of DT FTUs.

The lipid profile of the CDs was found to be enriched in PCs (Fig. 6A and figs. S46 and S48) and PEs (Fig. 6B and figs. S45 and S47), both of which serve as discriminatory markers for CDs as per our SHAP analysis. Figures S45 and S46 show that four of the top

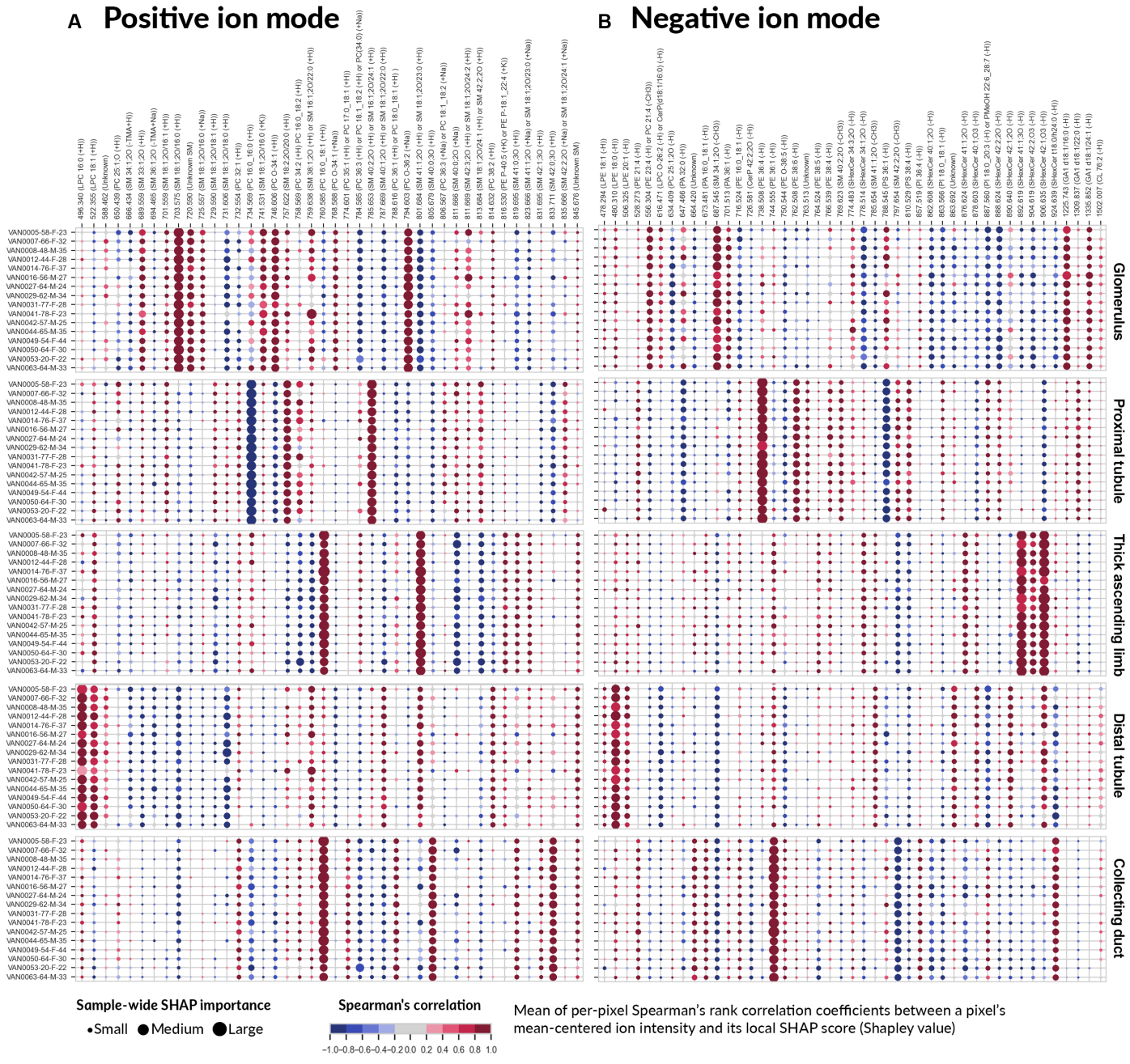

Fig. 6. Summary of biomarker candidates for five FTUs, namely GLs, PTs, DTs, CDs, and the TAL, obtained by applying our SHAP-based workflow to the atlas. The bubble plot reports both positive ion mode (**A**) (left) and negative ion mode (**B**) (right) findings. The columns correspond to a selection of molecular species (in increasing order of *m/z* ratios) that are biomarker candidates for one (or multiple) of the five FTUs under study. The rows correspond to different donors, each row labeled with its donor ID number and followed by the donor's age, sex, and BMI. Each bubble marker is informative of the direction (positive or negative correlation) and magnitude (relatively large or small) of a molecular species' influence on the classification model designed to recognize one of the five FTUs. The marker size represents the magnitude of the molecular species' influence, as measured by its tissue sample-wide SHAP importance score for a given donor sample. The marker color indicates the direction of the molecular species' influence, as measured by Spearman's rank correlation coefficient between the molecular species' mean-centered ion intensity values and its local pixel-specific SHAP scores. A positive Spearman's rank correlation coefficient indicates that a high intensity of the molecular species correlates with the FTU. Conversely, a negative Spearman's rank correlation coefficient indicates that a low intensity of the molecular species correlates with the FTU.

m/*z* markers for CDs were either PCs (*m*/*z* 734.569, *m*/*z* 760.585, and *m*/*z* 788.616) (figs. S139, S142, and S143) or PEs (*m*/*z* 744.555) (fig. S129). PE(36:1) (*m*/*z* 744.555) is the most prominent biomarker candidate of CDs detected in negative ion mode (fig. S45). While not exclusive to CDs, its relative cohort-wide SHAP importance is higher than in DTs and the TAL, where it was determined to be a less prominent positive marker (Fig. 6B and figs. S41, S43, and S45). In positive ion mode, PC(16:0_18:1) (*m*/*z* 760.585) has the highest cohort-wide SHAP importance score for CDs (Fig. 6A and figs. S46 and S48) but is also positively associated with the TAL (Fig. 6A and fig. S42). PC(32:1) (*m*/*z* 732.554), PC(35:1) (*m*/*z* 774.601), and PE (38:1) (*m*/*z* 772.586) are nearly exclusive markers of CDs (Fig. 6, A and B, and figs. S37 to S46, S147, and S148). Three additional PC (*m*/*z* 758.569) (fig. S46) or PE lipids (*m*/*z* 528.273 and *m*/*z* 716.524) (fig. S45) are also positively associated biomarker candidates for CDs (figs. S45 and S46), albeit not as robustly. Although none of these three are exclusively discriminatory, the overall number of positively associated PC and PE biomarker candidates for CDs as determined by our SHAP analysis is 12 in total, the highest number for any FTU (figs. S45 and S46).

The sulfatidyl hexose ceramide SHexCer(t18:0/h24:0) (*m*/*z* 924.639) also has strong positive association with the CD, the second highest positively associated biomarker candidate detected in negative ion mode (Fig. 6B and figs. S45, S47, and S131). It is anticorrelated with the preceding DT segment (Fig. 6B and figs. S43 and S47), suggesting a CD-specific function. SHexCer(t18:0/h24:0) (*m*/*z* 924.639) incorporates a phytosphingosine as the sphingoid base as well as a hydroxy fatty acid. It is of note that it is one of only two phytosphingosine ceramides detected in the dataset and the only one positively correlated with an FTU. SHexCer(t18:0/h24:0) (*m*/*z* 924.639) has been reported to specifically localize to the CD (*23*). Other lipids that are functionally important for the membrane structure were also identified, including the long-chain sphingomyelins SM(40:0;2O) (*m*/*z* 805.679) and SM(42:0;3O) (*m*/*z* 833.711), which are two of the top CD biomarker candidates (Fig. 6 and figs. S45 to 48, S140, and S141).

Sex and BMI comparisons

Using the demographic information provided in the donor metadata of the atlas, we can go beyond the discovery of FTU-level biomarker candidates and explore lipid associations with sex and other important characteristics such as BMI. Binary classification of female versus male donors and binary classification of normal-BMI (22 < BMI < 25) versus high-BMI (BMI > 35) donors, in conjunction with our SHAP-based interpretable machine learning workflow, enables the discovery of lipid biomarker candidates that are associated with these donor categories. The cohort-wide SHAP importance scores, highlighting the top biomarker candidates for the classification of female from male donor tissue overall (i.e., not specific to an FTU) for both negative and positive ion modes, are provided in figs. S149 and S150 and summarized in bubble plots (Fig. 7, A and B). FTU-specific analyses were also performed, differentiating female from male kidney FTUs for both polarities (figs. S151 to S160). While our SHAP approach is inherently multivariate, we can also analyze univariate ion intensity distribution comparisons. These are provided as split violin plots for a selection of biomarker candidates for both polarities (positive ion mode, Fig. 7C and fig. S162; negative ion mode, Fig. 7D and fig. S161). Ion images for the top 10 overall sex biomarker candidates for both negative and positive ion modes are provided in figs. S163 to S182.

Stratifying the dataset for females versus males reveals enrichment in PC and LPC in females. There are nine molecular signals, representing PC species, presenting as positively correlating marker candidates for females (*m*/*z* 572.335, *m*/*z* 650.439, *m*/*z* 728.522, *m*/*z* 732.554, *m*/*z* 754.536, *m*/*z* 758.569, *m*/*z* 786.601, *m*/*z* 810.601, and *m*/*z* 848.557), compared to four species positively correlating with males (*m*/*z* 720.554, *m*/*z* 784.585, *m*/*z* 812.614, and *m*/*z* 822.541) (Fig. 7A and figs. S150, S173, S175, S176, and S178 to S180). For LPCs, none are positively correlated with males, while five (*m*/*z* 496.340 slightly, *m*/*z* 518.322, *m*/*z* 524.371, *m*/*z* 544.33, and *m*/*z* 614.456) show positively correlating discriminative association to females (Fig. 7, A and B, and fig. S150), consistent with the PC pattern. Our analysis suggests that phospholipids containing arachidonic acid LPE(20:4) (*m*/*z* 500.281), PE(O-16:1_20:4) (*m*/*z* 722.513), PC(18:0_20:4) (*m*/*z* 810.601), and PI(18:0_20:4) (*m*/*z* 885.55) are more (positively correlating) associated with females (Fig. 7, A to D, and figs. S149, S150, S163, and S167). Out of the molecular species captured by this atlas, PI(18:0_20:4) (*m*/*z* 885.55) is suggested to be the most robust sex biomarker candidate, positively correlating in females (Fig. 7, B and D).

Figure 7 (B and D) shows that PSs play a consistent role in differentiating females and males, with our analysis reporting that several PS lipids have a negative Spearman's rank correlation coefficient, indicating that low intensity of these molecular species correlates with the female sex and high intensity correlates with males. This is further apparent when evaluated by individual FTU (figs. S151, S153, S155, S157, and S159). In a recent study, it was shown that PSs comprise 60% of the urine lipidome in males and 45% in females, with PS(36:1) (*m*/*z* 788.545) being the most abundant urine lipid. In our analysis, PS(36:1) (*m*/*z* 788.545) was found to be a biomarker candidate for males overall (Fig. 7, B and D, and fig. S149) and an even higher-ranked differentiating marker for the TAL (fig. S155). PS(36:2) was also found to be higher in male urine, and in our atlas, this PS was also found to be an overall molecular marker for male kidney tissue (Fig. 7B and figs. S149 and S171), with PS(36:2) (*m*/*z* 786.529) having a high ranking cohort-wide SHAP importance score in PTs (fig. S153) relative to other FTUs and the overall tissue ranking.

To identify molecular signatures for obesity, 14 donors were selected from the atlas to compare those with a normal BMI range (BMI 20 to 25, seven donors) to those with BMI values classified as obese (BMI > 35, seven donors). The cohort-wide SHAP importance scores highlighting the top biomarker candidates for the classification of obese from normal donor tissue overall (i.e., not specific to an FTU) for both negative and positive ion modes are provided in figs. S183 and S184 and summarized in bubble plots (Fig. 8, A and B). FTU-specific analysis was also performed, differentiating kidney FTUs from normal donors and donors with obesity for both polarities (figs. S185 to S194). Intensity distribution comparisons are provided as split violin plots for a selection of obesity biomarker candidates for both positive ion mode (Fig. 8C and fig. S195) and negative ion mode (Fig. 8D and fig. S196). Ion images for the top 10 overall obesity biomarker candidates for both negative and positive ion modes are provided in figs. S197 to S216.

LPE species, including LPE(18:1) (*m*/*z* 478.284), LPE(18:0) (*m*/*z* 480.310), and *N*-acyl-lysophosphatidylethanolamine LPE-N(FA3:0) (*m*/*z* 554.289), were found to be differentiators and positively

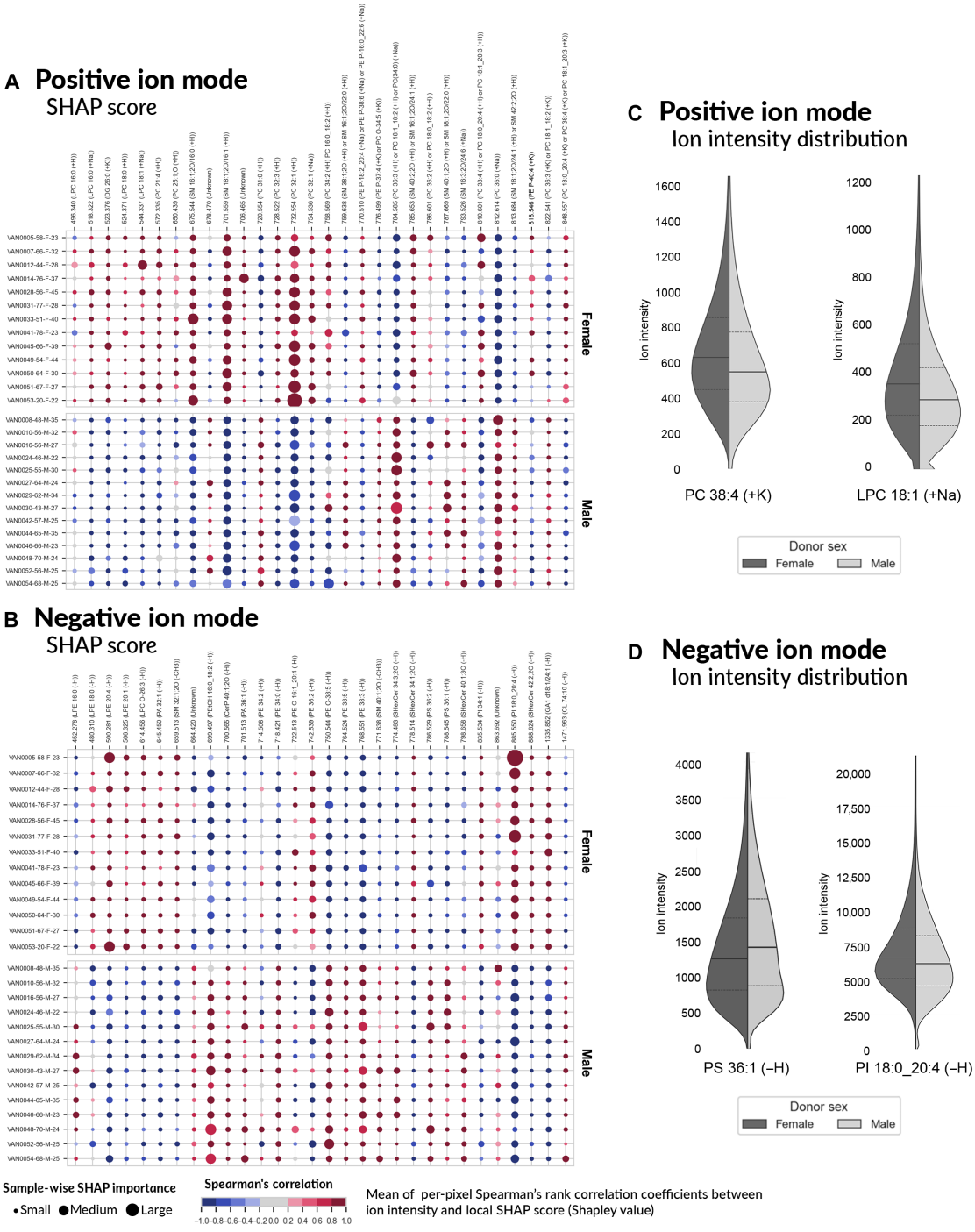

Fig. 7. Summary of biomarker candidates for sex, obtained by applying our SHAP-based workflow to the atlas. The bubble plots report positive ion mode (**A**) and negative ion mode (**B**) findings. The columns correspond to a selection of molecular species (in increasing order of *m/z* ratios) that are biomarker candidates for the female sex. The rows correspond to the different donors. Each bubble marker is informative of the direction and magnitude of a molecular species' influence on the classification model designed to recognize female donor tissue from male donor tissue. A molecular species with a large marker acts as a differentiator between sexes. The marker color indicates the direction of the molecular species' influence, as measured by the Spearman's rank correlation coefficient between the molecular species' mean-centered ion intensity values and its local pixel-specific SHAP scores. Molecular species that are positively correlated with female sex are negatively correlated with male sex, and vice versa. A positive Spearman's rank correlation coefficient indicates that a high intensity of the molecular species correlates with either the female sex (A/B top) or the male sex (A/B bottom). Conversely, a negative Spearman's rank correlation coefficient indicates that a low intensity of the molecular species correlates with the female sex (A/B top) or the male sex (A/B bottom). The split violin plots on the right report the ion intensity distributions of select sex biomarker candidates in positive mode (**C**) and negative mode (**D**), approximated using kernel density estimation. The violin plots are cropped at the 99th percentile of the distribution of one of the two sexes (whichever is larger) to facilitate visual comparison. The full line of each violin plot indicates the median of each class' distribution, whereas the dashed lines indicate its interquartile range.

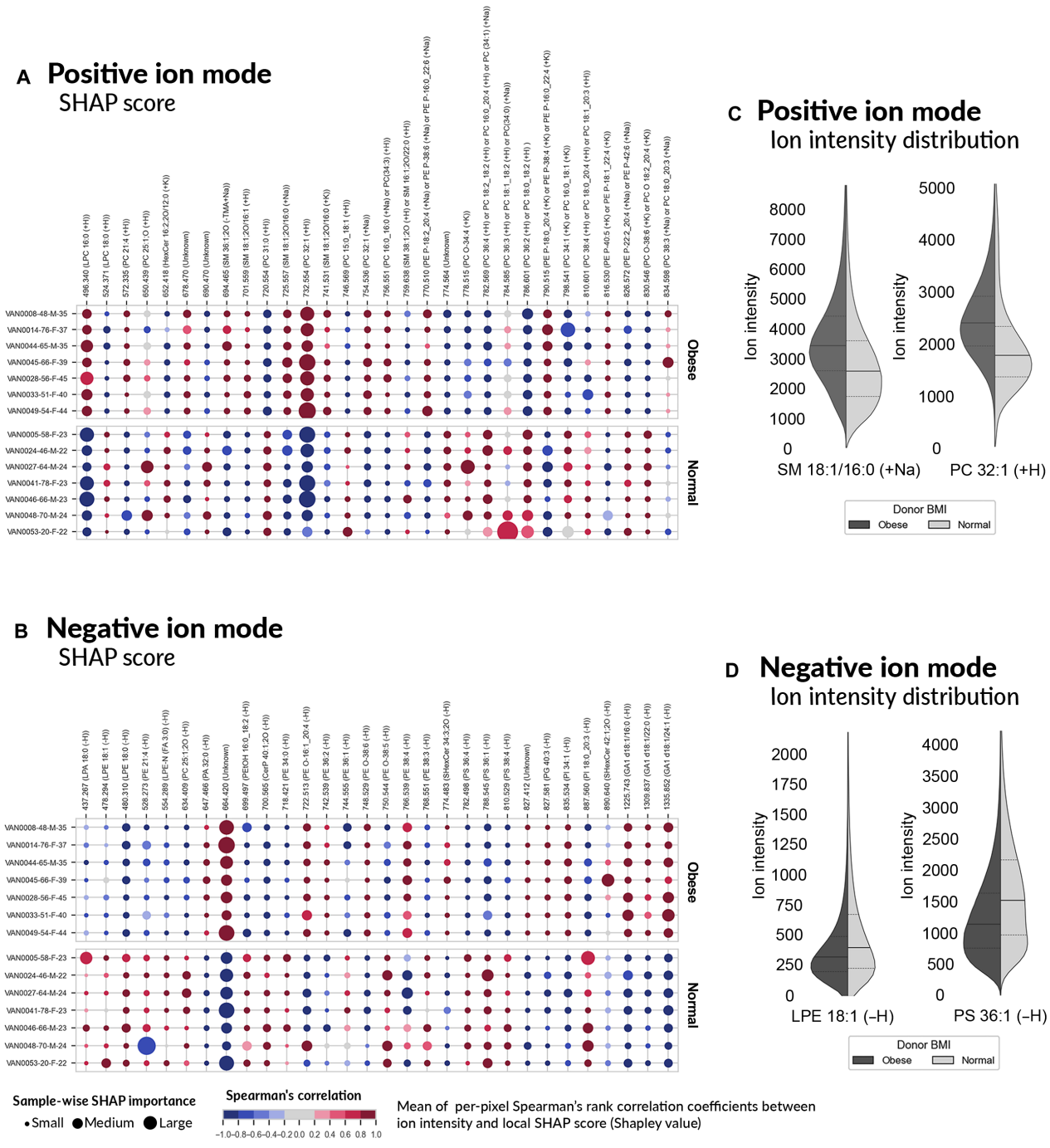

Fig. 8. Summary of biomarker candidates for BMI, obtained by applying our SHAP-based workflow to the atlas. The bubble plots report positive ion mode (**A**) and negative ion mode (**B**) findings. The columns correspond to a selection of molecular species (in increasing order of *m/z* ratios) that are biomarker candidates for the characteristic of obesity. The rows correspond to the different donors. Each bubble marker is informative of the direction and magnitude of a molecular species' influence on the classification model designed to recognize high-BMI donor tissue from normal-BMI donor tissue. A molecular species with a large marker acts as a differentiator between BMI classes. The marker color indicates the direction of the molecular species' influence, as measured by the Spearman's rank correlation coefficient between the molecular species' mean-centered ion intensity values and its local pixel-specific SHAP scores. Molecular species that are positively correlated with high BMI are negatively correlated with normal BMI, and vice versa. A positive Spearman's rank correlation coefficient indicates that a high intensity of the molecular species correlates with either high BMI (A/B top) or normal BMI (A/B bottom). Conversely, a negative Spearman's rank correlation coefficient indicates that a low intensity of the molecular species correlates with either high BMI (A/B bottom) or normal BMI (A/B bottom). The split violin plots on the right report the ion intensity distributions of select obesity biomarker candidates in positive mode (**C**) and negative mode (**D**), approximated using kernel density estimation. The violin plots are cropped at the 99th percentile of the distribution of one of the BMI categories (whichever is larger) to facilitate visual comparison. The full line of each violin plot indicates the median of each class' distribution, whereas the dashed lines indicate its interquartile range.

correlating to normal donors (Fig. 8, B and D, and figs. S183 and S205). Although many PCs were found to be markers for both normal and obese BMI kidney tissues (Fig. 8, A and C, and figs. S183 and S184), PC(32:1) (*m*/*z* 732.554) was found to be the greatest discriminator of these tissues. PC(32:1) (*m*/*z* 732.554) is a positively correlating discriminative marker candidate for obesity and appears negatively correlated to normal tissue, with the greatest cohort-wide SHAP importance score (Fig. 8A and figs. S184 and S207). Phosphatidylcholine PC(36:2) (*m*/*z* 786.601) [most likely PC(18:0_18:2) based on LC-MS/MS data] was found to be one of the most consistent (positively correlating) biomarker candidates for normal BMI, and PC(O-34:4) (*m*/*z* 778.515) and PC(34:1) (*m*/*z* 798.541) suggest less prominent but similar behavior (Fig. 8A and figs. S184, S209, S211, and S214). In negative ion mode, the ion detected at *m*/*z* 664.420 was the most discriminant of BMI, reporting the highest absolute cohort-wide SHAP importance score and correlating positively with obesity and negatively with normal BMI (Fig. 8B and figs. S183 and S197). We were able to confirm the identity of this molecule by MS/MS as a PAz-PC lipid species (i.e., 1-palmitoyl-2-azelaoyl PC), an oxidized PC previously observed in kidney disease (*24*, *25*).

Other marker candidates for obesity included PSs and SMs. Phosphatidylserines PS(36:4) (*m*/*z* 782.498), PS(36:1) (*m*/*z* 788.545), and PS(38:4) (*m*/*z* 810.529) were all (positively correlating) molecular indicators of normal BMI donor tissues (Fig. 8B and figs. S183 and S201). PS(36:1) (*m*/*z* 788.545) was also a positively correlating biomarker candidate for GLs and a negatively correlating marker for PTs (not considering obesity), although it showed a relatively high degree of variability as an FTU indicator (Fig. 6B). When considering the classification of normal versus obese BMI categories, PS(36:1) (*m*/*z* 788.545) was found to be a more consistent marker candidate positively associated with normal BMI donor tissues (Fig. 8, B and D, and figs. S183 and S201). This is an example of how our SHAP analysis of the kidney lipid atlas can differentiate consistent biomarker candidates from those that are rather variable and linked to specific donor comorbidities. The sphingomyelins SM(18:1;2O/16:0) (*m*/*z* 725.557 and *m*/*z* 741.531) and SM(18:1;2O/16:1) (*m*/*z* 701.559) were found to be discriminant for samples from donors with obesity (Fig. 8, A and C, and figs. S184 and S213), correlating positively with high BMI. Given that SM(18:1;2O/16:0) is a robust glomerular biomarker candidate as well (Fig. 6A), the specific SM profile of GLs in donors with obesity versus normal donors was analyzed further. SM(18:1;2O/16:0) (*m*/*z* 725.557) demonstrated the highest cohort-wide SHAP importance score of all measured SM species and showed positive correlation to obesity (fig. S186). This observation appears consistent with the trend of the percentage of glomerular sclerosis found in the histopathological reports of the PAS-stained tissues. The average glomerular sclerosis for donors with obesity was 15.96% compared to 7.8% in normal donor tissues (table S2). The positive correlation of high SM(18:1;2O/16:0) (*m*/*z* 725.557) intensity to GLs in tissues from donors with obesity (fig. S186) and increased glomerular pathology indicates that perturbation of SM(18:1;2O/16:0) (*m*/*z* 725.557) may be a molecular signature of glomerular disease.

DISCUSSION

We have built a multimodal, high-dimensional atlas of the human kidney, including lipid, multicellular FTU, and histological imaging data, as well as pathophysiological information for each human subject. It is constructed from millions of mass spectral measurements with spatial annotations delineating more than 100,000 FTU instances across 29 donor whole slide kidney tissue sections. As part of the NIH HuBMAP (*6*), the data are freely accessible to the greater scientific community (https://portal.hubmapconsortium.org; table S3). As an initial demonstration of the atlas' potential, we highlight lipid differences between specific FTUs of the nephron, between subjects with BMIs classified as normal (BMI 20 to 25) and obese (BMI > 35), and between men and women. Additional insight into the molecular variance associated with demographics may help explain the functional ramifications resulting in differential severity in diseases such as chronic kidney disease, cancer, and diabetes (*26*–*28*). Lipids, particularly in the context of FTUs and cellular neighborhoods, are chronically understudied compared to their transcript and protein counterparts despite their clear involvement in health and disease (*29*, *30*). Many large studies assume that genetic expression is sufficient for understanding disease states. While likely true in some cases, there is a known discordance in gene expression and metabolomic profiles, indicating that further research in this area is required to understand complex physiology. Defining spatially specific molecular profiles is necessary for realizing the promise of personalized medicine and improved therapeutics. As a step toward this goal, our kidney lipid atlas serves to catalog lipid profiles on a multiscale level. Ultimately, multidimensional and multiscale atlas efforts, such as the one discussed here, can provide key findings for a defined, targeted validation and serve as a resource for others to use to generate new hypotheses and corroborate orthogonal studies. Our multimodal lipid atlas and comprehensive analyses, including both unsupervised and supervised interpretable machine learning approaches, can establish the basal lipid signature of normal renal tissue against which diseased tissue can be compared, providing further insight into pathobiological mechanisms.

Extraction of biological insight into FTU lipid function

Tissue atlases depend on harmonization with previous studies to anchor outputs and provide a reliable foundation for new observations. For example, phosphatidic acid PA(36:1) has previously been found to mark GLs using IMS (at a 30-μm pixel size) (*21*). Our higher spatial resolution multimodal imaging workflows and SHAP analysis are consistent with this observation, finding the presence of PA(36:1) to be a marker candidate for GLs (Fig. 6B and figs. S37 and S47). Other lipids localized to GLs include SM(34:1;2O), which was putatively identified in murine and human kidney sections previously (*21*). Specifically, we confirmed the glomerular localization of SM(34:1;2O) (Fig. 6B and figs. S37, S47, and S49) and further identified it as SM(18:1;2O/16:0) (Fig. 6A and figs. S38, S48, and S59), which has previously been shown to be a relevant lipid in plasma and urine studies of chronic kidney disease (*31*, *32*). SM(18:1;2O/16:0) requires the podocyte-specific enzyme ceramide synthase 6 (CERS6) for synthesis. Genetic ablation of CERS6 leads to the loss of F-actin fibers and down-regulation of synaptopodin and CD2AP (CD2-associated protein), which facilitates nephrin binding to the actin cytoskeleton (*33*). The CD2AP, synaptopodin, nephrin, and cytoskeleton complex localizes to lipid rafts to anchor the podocyte slit diaphragm. It is known that lipid raft formation is governed by the interaction of cholesterol with locally available phospholipids with SMs containing 16:0 acyl chains having the highest affinity for cholesterol in ternary PC:SM:cholesterol bilayers, the highest lateral segregation tendency, and the highest thermostability (*34*), suggesting that SM(18:1;2O/16:0) is critical for the formation of lipid rafts.

Given that SM(18:1;2O/16:0) is specifically correlated with the GLs and CERS6 down-regulation destabilizes the slit diaphragm, we propose that SM(18:1;2O/16:0) is mechanistically linked to the slit diaphragm by providing a local lipid environment that stabilizes the architecture of, or regulates the formation of, the specific lipid raft microdomain associated with the slit diaphragm.

In addition to sphingomyelin SM(18:1;2O/16:0) being enriched in GLs, we observed a robust positive association of gangliosides with GLs (Fig. 6B and figs. S37, S47, S50, and S54), consistent with a previous publication (*35*). While GA1 species have not previously been reported to be specifically localized to the GL, products of the GA1 ganglioside biosynthesis pathway, GM1b and GD1a, have been detected as part of the lipid profiles of podocytes (*36*), a unique and important cell type contributing to the filtration barrier in the GL.

Long-chain fatty acid–containing phospholipids combined with cholesterol lead to plasma membranes with high rigidity and higher viscosity (*37*). While SM(18:1;2O/16:0) was determined to be a strong biomarker candidate for GLs (among the species measured in this atlas), a more diverse set of six SMs was found to exhibit positive marker potential for PTs (Fig. 6, A and B, and figs. S39, S40, S47, S48, S74, S75, S80, S81, S83, S84, S86, and S88). The PT, which is the primary site of glucose reabsorption in the nephron, is known to have an unusually rigid apical membrane (*38*). Decreased membrane fluidity is commonly associated with a high ratio of SMs (*37*). Membrane fluidity is a critical regulator of glucose uptake in the PT, as increasing fluidity inhibits glucose reabsorption by sodium-glucose cotransporter-2 (SGLT2) (*39*). This enrichment in SMs, taken together with the tendency for these PT-specific SMs to have very long chain fatty acid tails which drive leaflet interdigitation and a higher propensity for fluid-to-gel phase transition in lipid membranes (*40*), is consistent with the rigid apical membrane and lower membrane fluidity of PTs. Under pathological conditions such as diabetes, membrane fluidity is increased and results in the reduction of glucose transport mediated by the sodium-glucose cotransporter SGLT2 (*38*, *39*, *41*), thereby abrogating a key function of the PT segment of the nephron where the majority of glucose is reabsorbed. This suggests that the maintenance and regulation of membrane fluidity by the local lipid environment are critical for PT function.

In addition to membrane fluidity as a regulator of SGLT2, its function is also regulated by protein kinase C (PKC) (*42*). While PKC binds membranes enriched in SM, it also requires PS for activation in such environments (*43*). Of the two PS species that emerged as biomarker candidates in the atlas, the PT-specific PS(38:4) could be critical for activating PKC in an SM-rich environment. While PKC can bind to SM-rich membranes with high affinity, this lipid microenvironment inhibits PKC activity. This inhibition is relieved by the binding of PS, resulting in full kinase activity (*43*). The phosphatidylserine PS(38:4) is uniquely associated with PTs (Fig. 6B and figs. S39, S47, and S73), which suggests a putative regulatory role for Na^+/K^+ ATPase activity in this FTU.

In PTs, a sodium gradient must be established by the basolateral Na^+/K^+ ATPase pump for the uptake of sodium, and subsequently glucose, from the lumen across the apical membrane via SGLT2. Na^+/K^+ ATPase activity is known to be modified by neutral phospholipids, particularly polyunsaturated PEs. While Na^+/K^+ ATPase and the aforementioned PKC are widely expressed, they have site-specific regulatory mechanisms of SGLT2 in the PT (*42*, *44*). The identification of specific lipids in PTs that can serve as regulators for these proteins suggests that they could be critical for modifying PKC and Na^+/K^+ ATPase activity, which could then fine tune SGLT2 activity and glucose reabsorption. Similar to PKC activation, Na^+/K^+ ATPase is stimulated by certain lipids, including polyunsaturated PEs. PTs had five polyunsaturated PE lipids with (positively correlating) discriminative association (Fig. 6B and figs. S39, S47, S69, S71, and S76), more so than the TAL and DTs, which express higher levels of the sodium pump (*45*). Within this group, PE(38:4) directly stimulates Na^+/K^+ ATPase pump activity (*22*). This enrichment in specific PEs suggests that these may be regulatory molecules that can be leveraged to adjust the sodium gradient by modifying Na^+/K^+ ATPase activity in response to glucose absorption cues from the lumen.

Sulfatides provide a molecular signature for the TAL (Fig. 6B and figs. S41, S47, and S89 to S92). These lipids present an anionic charge on the plasma membrane, where they are available to modulate membrane ion fluxes by binding cations. They have been proposed to act as counterions to interstitial ammonium (NH_4^+), leading to its accumulation in the TAL (*46*). In addition, sulfatides have been proposed to be regulators of Na^+/K^+ ATPase by binding K^+ directly or modulating its localization (*46*). The ceramide moiety of SHexCer is a bioactive second messenger known to activate phosphatases as well as kinases and modulate several TAL-specific ion channels. The ceramide target protein phosphatase 1 (PP1) (*47*) dephosphorylates sodium-hydrogen antiporter 3 (NHE3), alleviating inhibition of this Na^+/H^+ antiporter (*48*). If PP1 activity is inhibited, the dephosphorylation of Na^+/K^+ ATPase is perturbed, and pump activity is negated (*49*). Last, PP1 can suppress epithelial sodium channel activity and regulate sodium reabsorption. The enrichment of SHexCer in the TAL and putative roles of both sulfatide and ceramide in regulating ion balance suggest that SHexCers are mechanistically important for TAL function.

While the nephron as an entire unit is taxed with filtrate reabsorption, the DT FTU is distinct from preceding nephron segments in how it executes reabsorption. The DT relies more on transcellular transport for a tightly controlled approach to regulating water and ion balance. It is critical for absorption of divalent cations, including magnesium and calcium, as well as for maintaining sodium and potassium homeostasis (*50*). The DT is unique in that it is highly plastic and can fine tune ion uptake and secretion in response to luminal ion concentration by modulating ion channel expression and activity. The ability to adjust the activity of ion channels is critical to the function of DTs. Second messengers allow cells to be highly plastic and adapt to environmental cues, such as the change in electrolyte concentration. Reactive oxygen species (ROS) are a class of second messengers that function in signaling cascades that target ion channels and transporters (*51*). LPCs were recognized early on to be an inducer of ROS (*52*). Subsequently, it was discovered that LPC(16:0) is the most potent ROS inducer, followed by LPC(18:0) and LPC(18:1) (*53*). In our atlas, LPC(16:0), LPC(18:1), LPE(18:0), LPE(18:1), and LPE(20:1) were all determined to be (positively correlating) biomarker candidates for DTs (Fig. 6, A and B, and figs. S43, 44, S47, 48, S109, S115, S119, and S120), providing a lipid landscape capable of driving ROS production for modulation of ion channel activity. ROS has been implicated in the up-regulation of the DT sodium-chloride cotransporter (*54*). Also, K^+ channels in the basolateral membrane can be stimulated by prostaglandin F2α in a ROS-dependent mechanism (*54*). Recently, ROS was shown to modulate the DT-specific Mg^{2+} channel TRPM6 (*55*). These observations, coupled with finding LPC and LPE molecular markers in DTs, point to a role for lipid-induced ROS regulation of ion channels in DTs.

The renal filtrate from tubules drains into the CDs, the final stage where nutrients are reabsorbed to adjust the final concentration of the electrolyte and acid-base components of the resulting urine. The CD is responsible for achieving the final electrolyte, pH, and fluid balance of urine before reaching the calyces and renal pelvis. The CD is composed of principal and intercalated cells. Principal cells express the epithelial sodium channel and Na^+/K^+ ATPase to regulate sodium reabsorption and potassium secretion. Intercalated cells primarily participate in acid-base secretion and reabsorption and are part of a class of proton-secreting cells known as mitochondria-rich cells (*56*). Consistent with the presence of mitochondria-rich cells, PCs and PEs are highly enriched in the CD (Fig. 6, A and B, and figs. S45 to S48). PC and PE are the most abundant lipids in the mitochondrial membrane (*57*) with the outer mitochondrial membrane rich in PC and PE, while the inner mitochondrial membrane contains PE (*58*). PC is well suited for the smooth outer mitochondrial membrane as its physical properties make it ideal for forming a smooth bilayer (*58*). The inner mitochondrial membrane lipidome is rich in non–bilayer-forming PE to maintain stability and render it impermeable to solutes (*58*). Given the critical role these phospholipids play in the structure function of mitochondria and their selective enrichment in CDs, their detection as strong biomarker candidates in this atlas is likely aligned with the mitochondria-rich intercalated cells of this FTU.

The phytosphingosine ceramide SHexCer(t18:0/h24:0) also has discriminative positive association with the CD (Fig. 6B and figs. S45, S47, and S131). Phytosphingosine ceramides are unique in that they are 4-hydroxylated molecules. These phytoceramides are found in the skin barrier, where it is proposed that they introduce heterogeneity into the membrane and expand the mechanisms of permeability restrictions (*59*). While their membrane packing is less ordered, the hydrogen bonding of their head group adds structure to the membrane in a manner that is divergent from other ceramides (*60*). Given that the CD must maintain strict permeability regulation, it is intriguing that the same phytoceramide that comprises the skin barrier is detected here. Nakashima *et al.* (*23*) also observe SHexCer(t18:0/h24:0) in CDs and propose that its hydroxylation state is critical for regulating the transport of NH_3 and H^+ by strengthening lateral interactions of the plasma membrane. Long-chain SM species that can stabilize membrane domains (*61*) are also found to be associated with the CDs (Fig. 6 and figs. S45 to 48, S140, and S141). Together, the phytosphingosine ceramide and SM species of the CD could create leakproof but flexible membranes for the regulation of ammonia transport. It is speculated that tuning these lipid components of the CD membrane could allow for strict control of NH_3 and H^+ transport via a transcellular route exclusively.

Molecular characteristics of sexual dimorphism

Our analysis of the atlas suggests that lipidomic signatures of females are enriched with increased PC and LPC compared to males (Fig. 7, A and B, and figs. S150, S173, S175, S176, and S178 to S180). This enhanced association of PCs with females could be attributed to phosphatidylethanolamine-*N*-methyltransferase levels, which are regulated by estrogen (*62*). Our SHAP-based analysis also found arachidonic acid–containing phospholipids to be marker candidates for female donor tissue (Fig. 7, A to D, and figs. S149, S150, S163, and S167). Arachidonic acid is liberated from phospholipids by phospholipases and can be metabolized into prostanoids PGE_2 and PGI_2, which act in the kidney to regulate blood pressure. Studies have shown that females produce higher levels of PGE_2 and that these levels correlate with lower blood pressure compared to age-matched males (*63*). The observation that PCs are key components of the molecular signature of the kidney for females indicates that this important class of lipids should be investigated further related to hormonal regulation of phosphatidylethanolamine-*N*-methyltransferase and the sexual dimorphism of prostanoids and blood pressure regulation.

Lipid signatures in normal donors and donors with obesity

In the comparison of donors classified as those with obesity versus normal based on BMI values, LPEs were one of the primary lipid classes to show an association with normal BMI donor tissues (Fig. 8, B and D, and figs. S183 and S205). This is consistent with previous observations that LPEs are negatively correlated with BMI (*64*). While SHAP analyses of FTU and sex detected differences in overall lipid class enrichments, the lipid profiles when using BMI as a classifier showed a pronounced difference in specific PC and SM molecules. This may be a function of sex and FTU being physiologically stable aspects, while BMI occurs on a spectrum and is more heterogeneous throughout the population. The strongest biomarker for obesity was PC(32:1) (Fig. 8A and figs. S184 and S207), which has previously been shown to be associated with BMI (*65*). SMs also showed differences in their glomerular profiles of donors with obesity compared to normal donors. SM(18:1;2O/16:0)'s discriminative association with obesity (Fig. 8, A and C, and figs. S184 and S213) combined with an elevated detection of glomerular sclerosis in this population suggests a mechanistic link between alterations in SM lipids and the development of glomerular disease. Obesity represents a chronically inflamed state, and under these conditions, coagulation factors are aberrantly expressed or regulated, promoting increased prothrombotic risk (*66*, *67*). The oxidized phospholipid, PAz-PC (Fig. 8B and figs. S183 and S197), could be an important molecule in the feedback loop that exists between obesity and risk factors for cardiovascular disease as PAz-PC inhibits the tissue factor pathway inhibitor (*68*). The tissue factor pathway inhibitor inhibits the tissue factor-dependent, or extrinsic coagulation, pathway by directly and indirectly perturbing the activation of the serine protease factor Xa that interacts with fibrin, a marker of obesity (*69*) and linked to cardiovascular disease (*70*).

Future directions and refinements

The kidney lipid atlas that we have constructed provides a publicly available, comprehensive foundation for exploring the relationships between spatiomolecular information, defined FTUs of the nephron, histopathological data, and key donor characteristics. Now, these analyses provide detailed lipid information with FTU-level specificity. This work is limited in molecular breadth and structural specificity, as well as cohort diversity. Future studies will expand the cohort with additional and more varied demographics including age, race, sex, and BMI. Furthermore, workflows are being developed to increase the molecular coverage and structural specificity through multimodal and multiomic assay integration (*14*). The ability to connect deep molecular information to specific tissue features with improved granularity, from FTUs to specific cell types and, eventually, single cells, is necessary for building multiscale molecular atlases of human organs associated with normal aging and disease. This goal is now central to the efforts of many large-scale research consortia, such as the HuBMAP (*6*), Kidney Precision Medicine Project (*9*), Human Tumor Atlas Network (*10*), and others. The broader goals of

these research efforts and the focus of future releases of this kidney lipid atlas are the expansion of our MALDI IMS lipid coverage and integration of additional modalities to capture cellular organization and additional molecular classes more thoroughly. For example, multiplexed immunofluorescence microscopy assays use dozens of well-characterized protein markers to identify distinct cell types and their organization at cellular resolution. On the other hand, transcriptomics and MS-based proteomics assays are often collected from bulk samples (no spatial information) or in a profiling mode with limited spatial resolution targeting anatomical features or specific FTUs. High-spatial-resolution approaches to transcriptomic and MS-based proteomics are emerging but with limited molecular coverage. An integrated approach, building on what we have developed as part of the current version of the kidney atlas, will be necessary to refine the atlas in later iterations and leverage the strengths of many individual assays to provide a comprehensive picture of how molecular distributions relate to cellular organization and localized changes in biomolecular pathways in normal and diseased kidney tissue.

MATERIALS AND METHODS

Materials

Acetone, isopentane, tetrahydrofuran, acetonitrile, and methanol were purchased from Thermo Fisher Scientific (Pittsburgh, PA). 1,5-Diaminonaphthalene and carboxymethylcellulose were purchased from Sigma-Aldrich Chemical Company (St. Louis, MO).

Sample preparation

Human kidney tissue was surgically removed during a full nephrectomy, and remnant tissue was processed for research purposes by the Cooperative Human Tissue Network at Vanderbilt University Medical Center. Remnant biospecimens were collected in compliance with the Cooperative Human Tissue Network standard protocols and National Cancer Institute's Best Practices for the procurement of remnant surgical research material. Participants were consented for remnant tissue collection in accordance with Institutional Review Board policies. The study received ethical approval from the Vanderbilt University's Institutional Review Board (no. 210190). This committee is responsible for reviewing and approving research involving human subjects to ensure that ethical standards are met. Donor demographics including age, sex, race, weight, height, BMI, and comorbidities are recorded in table S1. Half of the excised tissue was flash frozen over an isopentane-dry ice slurry, embedded in carboxymethylcellulose, and stored at −80°C until use. Kidney tissues were cryosectioned to a 10-μm thickness and thaw mounted onto indium tin oxide–coated glass slides (Delta Technologies, Loveland, CO) for IMS analysis or regular glass slides for histological staining. Slides were stored at −80°C and returned to ~20°C within a vacuum desiccator before further processing. The remaining excised tissue was formalin fixed and paraffin embedded. Formalin-fixed and paraffin-embedded tissue was sectioned to a 10-μm thickness and stained using PAS staining methods for pathological assessment as done previously (*14*). Briefly, the percentage of cortex, medulla, and other structures was determined within the tissue section. In addition, a histological assessment is reported in table S2.

AF microscopy and tissue segmentation

AF microscopy images were acquired on each tissue before analysis using 4′,6-diamidino-2-phenylindole, enhanced green fluorescent protein, and DsRed filters on a AxioScan Z1 slide scanner (Carl Zeiss Microscopy GmbH, Oberkochen, Germany) (*71*). After IMS analysis, an additional AF image was acquired using bright-field and enhanced green fluorescent protein filters to locate MALDI laser ablation marks for image registration between the marks and their corresponding MALDI IMS pixels. Segmentation of FTUs based on AF images was performed as described previously (*15*).

Imaging mass spectrometry

Samples for IMS analysis were coated with a solution (20 mg/ml) of 1,5-diaminonaphthalene dissolved in tetrahydrofuran using a TM Sprayer M3 (HTX Technologies, LLC, Chapel Hill, NC), yielding a 1.67 mg/cm^2 coating (0.05 ml/hour, four passes, 40°C spray nozzle). Tissue samples were imaged immediately after matrix deposition. MALDI IMS was performed on a timsTOF fleX mass spectrometer (Bruker Daltonics, Bremen, Germany). The ion images were collected in positive and negative ion modes at a 10-μm pixel size with the beam scan set to 6 μm^2 using 150 laser shots per pixel and 18.6% laser power (30% global attenuator and 62% local laser power) at 10 kHz. A total of 75,150 to 400,250 pixels was acquired from each tissue. Data were collected from *m*/*z* 150 to 2000 for lipid analysis. Imaging data were visualized using SCiLS Lab Version 2019 (Bruker Daltonics, Bremen, Germany) or custom software described below. Lipids were identified using a combination of mass accuracy (≤5 ppm), orthogonal LC-MS experiments, and LIPIDMAPS (*72*, *73*) database searching. Approximately 200 lipids were identified from each patient sample.

LC-MS/MS

Three to five 10-μm sections were collected per patient into Eppendorf tubes. Samples were transferred into glass vials with 200 μl of methanol. Lipids were extracted using a modified methyl-*tert*-butyl ether protocol (*74*). Two hundred microliters of cold methanol was added to samples on ice and vortexed for 1 min. Eight hundred microliters of cold methyl-*tert*-butyl ether was added, and samples were homogenized by sonication in four rounds of 10 min each on ice. Two hundred microliters of cold water was then added, and samples were allowed to rest for 10 min. Samples were centrifuged at 1000*g* for 10 min at 4°C, and the upper organic layer was removed and evaporated to dryness. Samples were reconstituted in 500 μl of methanol for analysis with 500–parts per billion Avanti Equisplash internal standards. A Thermo Fisher Scientific Q Exactive HF with a Vanquish UHPLC+ and a Bruker TimsTof FleX mass spectrometer with a Waters Premier UHPLC were used for analysis with a Waters Premier reversed-phase C18 BEH column (2.1 mm by 100 mm) at 50°C. Lipids were separated with a binary gradient of (A) 10 mM ammonium acetate and 0.1% formic acid in water/acetonitrile (4:6) and (B) 10 mM ammonium acetate and 0.1% formic acid in 2-propanol/acetonitrile (9:1) with a flow rate of 250 μl/min for 15 min on the Thermo system and 30 min on the Bruker system. Ten microliters was injected in both negative and positive ion modes. Samples were pooled to create a reference sample for quality control. Data were analyzed and annotated in MS-DIAL version 4.90 (*75*).

MALDI IMS data preprocessing

MALDI IMS data were processed using in-house developed software. All datasets were first converted into a custom binary format optimized for storage and speed of analysis of IMS data. Each spectrum was *m*/*z* aligned using between 6 and 10 alignment peaks that

were automatically selected for each dataset. The alignment peaks are selected by their frequency of occurrence throughout the dataset and ensure that several anchor points are available for each spectrum. This is in contrast to using a predefined set of peaks that might not be present in every spectrum, hindering the alignment process. The alignment process was accomplished using the Python msalign library (version 0.2.0) (*76*, *77*). Subsequently, the spectra were calibrated using at least four calibration points in both polarities. Two vectors of normalization factors were calculated for each dataset, namely the total ion current (TIC) and an outlier-insensitive TIC variant that only includes intensity data lying between the 5th and 95th percentiles of each mass spectrum (5/95% TIC). Normalization aims to counteract noise factors and to project all mass spectra onto a consensus intensity scale so that intensities can be compared between spectra. Because the subsequent analyses compare multiple datasets from different patients, the normalization factors of a single dataset have no knowledge of the intensity scales another dataset has, which means that if normalization is performed on a dataset-by-dataset basis, ion intensities in one dataset might not be comparable to ion intensity of another. To counteract this, all normalization vectors were concatenated (TIC and 5/95% TIC were processed separately) into one large vector, which was divided by its median factor. This ensures that intensities are comparable between datasets. At this point, rather than creating a copy of the data for both normalization factors, we keep the IMS data in the *m*/*z*-aligned and calibrated state and apply the normalization factors when needed. Following the preprocessing, we calculate the average mass spectrum for each dataset, which is subsequently annotated using in-house–developed software (see below). The unsupervised and classification workflows (discussed in depth below) require that a common set of features (peaks) is present for each analyzed dataset; hence, a set of annotated peaks was used to extract ion images from the processed binary data. The data were extracted into a matrix of size ($N \times M$) where N is the number of pixels and M is the number of identified peaks. An image is created by summing the intensity of a specified peak within a ±3- to 5-ppm extraction window. Before further analysis, the image intensity matrix is normalized using one of the available normalization factors.

MALDI IMS tentative identification

The average mass spectrum was scaled between 0 and 1, peak picked, deisotoped, and filtered using a signal-to-noise ratio threshold of 0.03 (calculated as the fraction of the base peak of the mass spectrum). Features below this threshold were excluded from the identification process. IMS identification was performed using in-house–developed software that associated detected peaks with tentative lipids from the LIPID MAPS Structure Database and custom LC-MS/MS database (discussed above). Parameters for annotating peaks include $[M + H]^+$, $[M + Na]^+$, and $[M + K]^+$ adducts in positive mode and $[M - H]^-$ and $[M - CH_3]^-$ in negative mode and a search window of ±5 ppm. To provide more confident identification, IMS identification was subsequently associated with LC-MS/MS identifications.

MALDI IMS-microscopy and microscopy-microscopy image registration

MALDI IMS pixels were registered to microscopy using IMS MicroLink, an in-house–developed, open-source plug-in developed for the napari image viewer. Within IMS MicroLink, the theoretical coordinate (i.e., the *x*,*y* integer coordinates) of each IMS pixel was extracted from the IMS metadata and visualized as an IMS pixel map with randomized intensities and the post-AF image was registered to the IMS pixel map by selecting five corresponding fiducials where a fiducial is an IMS pixel in the IMS pixel map image and a laser ablation mark in the post-AF image. This creates an exact registration of the IMS pixel to its origin in microscopy coordinates, the laser ablation mark. After the alignment of the post-AF image to IMS, the preacquisition AF [pre-AF (IMS)] image is automatically registered to the registered post-AF image [post-AF (IMS)] using wsireg, an in-house–developed, open-source Python package to perform whole slide image registration. The CODEX image is also automatically registered with wsireg to the post-AF (IMS) image by concatenating transformations from modality to modality. After all registrations and alignments are completed, all microscopy images and IMS are sampled into a common coordinate space.

Unsupervised machine learning

UMAP was used to uncover molecular trends across all datasets. UMAP aims to preserve the global neighborhood structure of the data while producing low-dimensional embedding. UMAP embedding was performed at the pixel level where each pixel across all samples was considered as an observation (6,779,166 in positive mode and 6,568,017 in negative mode) and at the FTU level where an average profile was obtained for each detected FTU (78,190 in positive mode and 74,959 in negative mode). In both cases, 211 and 212 features were used in positive and negative modes, respectively.

UMAP projection was performed in Python using the umap-learn (version 0.5.2) (*18*, *78*) and scikit-learn (version 1.0.2) (*79*) libraries. In our analysis, the purpose of UMAP was to identify a low-dimensional projection; hence, the number of target dimensions was limited to 2 (n_components = 2). Several other parameters were adjusted, including the distance metric (metric = 'cosine'), minimum number of neighbors (n_neighbors = 250), and the minimal distance (min_dist = 0). The random seed was also specified to allow us to reproduce the results.

Because of the multidataset nature of the data, two normalization approaches were explored. First, standard TIC normalization was used, which results in internormalized data. The second approach was to remove inherent batch effects caused by the fact that each sample came from different patients with vastly different BMI, age, or gender. Batch effects were removed using the reComBat algorithm (*17*).

Supervised machine learning and Shapley additive explanations

The identification of biomarker candidates was performed computationally following a workflow similar to that previously presented (*19*). The first step of the workflow consists in learning a classification model from a set of labeled IMS datasets via supervised machine learning (28 datasets in negative ion mode and 29 in positive ion mode). The following classification tasks were performed: recognition (one-versus-all classification) of five FTUs (GLs, DTs, PTs, CDs, and TAL), binary classification of male versus female donors, and binary classification of normal-BMI ($22 < BMI < 25$) versus high-BMI ($BMI > 35$) donors. In the case of the FTU recognition task, segmentation by AF provided pixel-wise labeling of the IMS data. In the case of the sex and BMI classification, each IMS dataset was labeled using its corresponding donor metadata (female or

male; high BMI or low BMI). We choose eXtreme Gradient Boosting (XGBoost) models (*80*) for the classification of IMS data for the following reasons: XGBoost can encode nonlinear dependencies between molecular species, does not make assumptions about the data distribution, does not depend on measuring distances between high-dimensional mass spectra, and does not require the scaling of ion intensity values. Rather than use one XGBoost model, as per Tideman *et al.* (*19*), we use an ensemble of XGBoost models for each classification task. The reason for using an ensemble of XGBoost models is to ensure that our SHAP-based biomarker candidate discovery workflow is robust to correlated inputs (e.g., molecular species with similar ion images). The problem posed by correlated inputs is that multiple models may fit the data equally well without necessarily being faithful to the true biological relationships. There is therefore a risk that spurious correlations influence the results produced by SHAP on one single model and that the results vary from one model to the next. Given that our aim is to select a shortlist of biologically relevant candidate biomarkers, we cannot allow for correlation between molecular species to cause inconsistencies. The solution we propose is to stabilize the output of SHAP by applying it to an ensemble of 10 XGBoost models. The training process of each of these XGBoost models is initialized using a different random seed, and the models are trained on slightly different training datasets (the split between training and testing datasets is done by sampling without replacement with 10 different random seeds). Spurious patterns because of correlated inputs cancel out across the different XGBoost models, whereas the true correlations are reinforced by the ensembling process. The patterns that dominate the decision-making process (and the ensuing SHAP explanations) of the ensemble are therefore likely to be biologically relevant.

The second step of the workflow consists in quantifying the importance of each molecular species with respect to a given classification task using SHAP (*20*). We apply SHAP to the ensemble of 10 XGBoost models obtained in the previous step. The Shapley value, or local SHAP importance score, of a molecular species for a given mass spectrum is the mean of the Shapley values obtained by applying SHAP to the 10 classification models making up the ensemble. As discussed by Tideman *et al.* (*19*), we have one Shapley value per molecular species and per pixel: The Shapley value of a molecular species measures the magnitude and direction of its effect on the classification of one spatially localized mass spectrum. We have multiple IMS datasets, each of which corresponds to the renal tissue sample of a different donor. The datasets are labeled according to the donor's metadata: VAN - donor ID - donor age - donor sex - donor BMI. The global SHAP importance score is a summary heuristic for a molecular species' biomarker potential that is computed by taking the mean of the magnitude of its Shapley values across all mass spectra in a given dataset. The global SHAP importance score is a sample-specific measure of a molecular species' relevance to a given classification task. In the summary bubble plots (Figs. 6 to 8), the size of each marker corresponds to the global SHAP importance score of a given molecular species (column) for a given dataset (row). Last, the total SHAP score of a molecular species is an experiment-wide summary heuristic for its biomarker potential: It is computed by taking the mean of the molecular species' global SHAP scores across all datasets. The molecular species are ranked in descending order of total SHAP score to facilitate the selection of a shortlist of molecular species that are likely to be useful biomarker candidates for a given FTU, sex, or BMI.

The next question is whether a biomarker candidate is positively or negatively correlated with a given FTU, sex, or BMI. We measure the direction and magnitude of the relationship by computing the Spearman rank-order correlation coefficient $\boldsymbol{\rho}$ between the mean-centered intensity and the Shapley values of a given molecular species. The Spearman rank-order correlation coefficient $\boldsymbol{\rho}$ also provides a way of assessing the statistical significance of the relationship: $\boldsymbol{\rho}$ ranges from −1 to 1, and we consider $\boldsymbol{\rho}$ to be significant if its magnitude exceeds 0.2. In the summary bubble plots (Figs. 6 to 8), the marker color corresponds to the Spearman rank-order correlation coefficient per molecular species (column) and per dataset (row). We distinguish between the following three scenarios:

1) Positive Spearman rank-order correlation coefficient ($\boldsymbol{\rho} > 0.2$)—Red marker: Pixels with high (above the mean) ion intensity have positive Shapley values, and conversely, pixels with low (below the mean) ion intensity have negative Shapley values. A pixel's likelihood of belonging to the target class increases monotonically as the ion intensity of the molecular species increases. A high intensity of the molecular species is indicative of the target characteristic.

2) Negative Spearman rank-order correlation coefficient ($\boldsymbol{\rho} < -0.2$)—Blue marker: Pixels with low (below the mean) ion intensity have positive Shapley values, and conversely, pixels with high (above the mean) ion intensity have negative Shapley values. A pixel's likelihood of belonging to the target class decreases monotonically as the ion intensity of the molecular species increases. A low intensity of the molecular species is indicative of the target characteristic.

3) Nonsignificant Spearman rank-order correlation coefficient ($-0.2 < \boldsymbol{\rho} < 0.2$)—Gray marker: The relationship between the molecular species and the target class is either monotonic but nonsignificant or nonmonotonic (and therefore not well-captured by the Spearman rank-order correlation coefficient).

For example, we consider the BMI classification task, where an ensemble of XGBoost classification models is made to differentiate between high-BMI donors, donors with obesity, and normal-BMI donors. The characteristic of interest, for which we need biomarkers, is obesity. A molecular species with a high total SHAP score is one whose ion intensity distribution varies from one BMI category to the other, which enables classification. It may therefore be a useful biomarker for obesity and should be investigated further. The next step is to determine whether the molecular species is positively or negatively correlated with our characteristic of interest. If the ion intensity of a given molecular species is higher in high-BMI donors, the relationship between the ion intensity and the target characteristic is positive: A high ion intensity of the molecular species is a marker of obesity. Conversely, if it is lower in high-BMI donors, the relationship is negative: A low ion intensity of the molecular species is a marker of obesity.

Supplementary Materials

This PDF file includes:

Figs. S1 to S216
Tables S1 to S4

REFERENCES AND NOTES

1. C. D. Conde, C. Xu, L. B. Jarvis, D. B. Rainbow, S. B. Wells, T. Gomes, S. K. Howlett, O. Suchanek, K. Polanski, H. W. King, L. Mamanova, N. Huang, P. A. Szabo, L. Richardson, L. Bolt, E. S. Fasouli, K. T. Mahbubani, M. Prete, L. Tuck, N. Richoz, Z. K. Tuong, L. Campos, H. S. Mousa, E. J. Needham, S. Pritchard, T. Li, R. Elmentaite, J. Park, E. Rahmani, D. Chen, D. K. Menon, O. A. Bayraktar, L. K. James, K. B. Meyer, N. Yosef, M. R. Clatworthy, P. A. Sims,

D. L. Farber, K. Saeb-Parsy, J. L. Jones, S. A. Teichmann, Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).
2. The Tabula Sapiens Consortium, R. C. Jones, J. Karkanias, M. A. Krasnow, A. O. Pisco, S. R. Quake, J. Salzman, N. Yosef, B. Bulthaup, P. Brown, W. Harper, M. Hemenez, R. Ponnusamy, A. Salehi, B. A. Sanagavarapu, E. Spallino, K. A. Aaron, W. Concepcion, J. M. Gardner, B. Kelly, N. Neidlinger, Z. Wang, S. Crasta, S. Kolluru, M. Morri, S. Y. Tan, K. J. Travaglini, C. Xu, M. Alcántara-Hernández, N. Almanzar, J. Antony, B. Beyersdorf, D. Burhan, K. Calcuttawala, M. M. Carter, C. K. F. Chan, C. A. Chang, S. Chang, A. Colville, R. N. Culver, I. Cvijović, G. D'Amato, C. Ezran, F. X. Galdos, A. Gillich, W. R. Goodyer, Y. Hang, A. Hayashi, S. Houshdaran, X. Huang, J. C. Irwin, S. Jang, J. V. Juanico, A. M. Kershner, S. Kim, B. Kiss, W. Kong, M. E. Kumar, A. H. Kuo, R. Leylek, B. Li, G. B. Loeb, W.-J. Lu, S. Mantri, M. Markovic, P. L. McAlpine, A. de Morree, K. Mrouj, S. Mukherjee, T. Muser, P. Neuhöfer, T. D. Nguyen, K. Perez, R. Phansalkar, N. Puluca, Z. Qi, P. Rao, H. Raquer-McKay, N. Schaum, B. Scott, B. Seddighzadeh, J. Segal, S. Sen, S. Sikandar, S. P. Spencer, L. C. Steffes, V. R. Subramaniam, A. Swarup, M. Swift, W. Van Treuren, E. Trimm, S. Veizades, S. Vijayakumar, K. C. Vo, S. K. Vorperian, W. Wang, H. N. W. Weinstein, J. Winkler, T. T. H. Wu, J. Xie, A. R. Yung, Y. Zhang, A. M. Detweiler, H. Mekonen, N. F. Neff, R. V. Sit, M. Tan, J. Yan, G. R. Bean, V. Charu, E. Forgó, B. A. Martin, M. G. Ozawa, O. Silva, A. Toland, V. N. P. Vemuri, S. Afik, K. Awayan, O. B. Botvinnik, A. Byrne, M. Chen, R. Dehghannasiri, A. Gayoso, A. A. Granados, Q. Li, G. Mahmoudabadi, A. McGeever, J. E. Olivieri, M. Park, N. Ravikumar, G. Stanley, W. Tan, A. J. Tarashansky, R. Vanheusden, P. Wang, S. Wang, G. Xing, L. Dethlefsen, P.-Y. Ho, S. Liu, J. S. Maltzman, R. J. Metzger, K. Sasagawa, R. Sinha, H. Song, B. Wang, S. E. Artandi, P. A. Beachy, M. F. Clarke, L. C. Giudice, F. W. Huang, K. C. Huang, J. Idoyaga, S. K. Kim, C. S. Kuo, P. Nguyen, T. A. Rando, K. Red-Horse, J. Reiter, D. A. Relman, J. L. Sonnenburg, A. Wu, S. M. Wu, T. Wyss-Coray, The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
3. G. Eraslan, E. Drokhlyansky, S. Anand, E. Fiskin, A. Subramanian, M. Slyper, J. Wang, N. Van Wittenberghe, J. M. Rouhana, J. Waldman, O. Ashenberg, M. Lek, D. Dionne, T. S. Win, M. S. Cuoco, O. Kuksenko, A. M. Tsankov, P. A. Branton, J. L. Marshall, A. Greka, G. Getz, A. V. Segrè, F. Aguet, O. Rozenblatt-Rosen, K. G. Ardlie, A. Regev, Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science* **376**, eabl4290 (2022).
4. J. Hansen, R. Sealfon, R. Menon, M. T. Eadon, B. B. Lake, B. Steck, K. Anjani, S. Parikh, T. K. Sigdel, G. Zhang, D. Velickovic, D. Barwinska, T. Alexandrov, D. Dobi, P. Rashmi, E. A. Otto, M. Rivera, M. P. Rose, C. R. Anderton, J. P. Shapiro, A. Pamreddy, S. Winfree, Y. Xiong, Y. He, I. H. de Boer, J. B. Hodgin, L. Barisoni, A. S. Naik, K. Sharma, M. M. Sarwal, K. Zhang, J. Himmelfarb, B. Rovin, T. M. El-Achkar, Z. Laszik, J. C. He, P. C. Dagher, M. T. Valerius, S. Jain, L. M. Satlin, O. G. Troyanskaya, M. Kretzler, R. Iyengar, E. U. Azeloglu, Kidney Precision Medicine Project, A reference tissue atlas for the human kidney. *Sci. Adv.* **8**, eabn4965 (2022).
5. B. B. Lake, R. Menon, S. Winfree, Q. Hu, R. M. Ferreira, K. Kalhor, D. Barwinska, E. A. Otto, M. Ferkowicz, D. Diep, N. Plongthongkum, A. Knoten, S. Urata, L. H. Mariani, A. S. Naik, S. Eddy, B. Zhang, Y. Wu, D. Salamon, J. C. Williams, X. Wang, K. S. Balderrama, P. J. Hoover, E. Murray, J. L. Marshall, T. Noel, A. Vijayan, A. Hartman, F. Chen, S. S. Waikar, S. E. Rosas, F. P. Wilson, P. M. Palevsky, K. Kiryluk, J. R. Sedor, R. D. Toto, C. R. Parikh, E. H. Kim, R. Satija, A. Greka, E. Z. Macosko, P. V. Kharchenko, J. P. Gaut, J. B. Hodgin, KPMP Consortium, M. T. Eadon, P. C. Dagher, T. M. El-Achkar, K. Zhang, M. Kretzler, S. Jain, An atlas of healthy and injured cell states and niches in the human kidney. *Nature* **619**, 585–594 (2023).
6. S. Jain, L. Pei, J. M. Spraggins, M. Angelo, J. P. Carson, N. Gehlenborg, F. Ginty, J. P. Gonçalves, J. S. Hagood, J. W. Hickey, N. L. Kelleher, L. C. Laurent, S. Lin, Y. Lin, H. Liu, A. Naba, E. S. Nakayasu, W.-J. Qian, A. Radtke, P. Robson, B. R. Stockwell, R. van de Plas, I. S. Vlachos, M. Zhou, HuBMAP Consortium, K. Börner, M. P. Snyder, Advances and prospects for the Human BioMolecular Atlas Program (HuBMAP). *Nat. Cell Biol.* **25**, 1089–1100 (2023).
7. O. Rozenblatt-Rosen, M. J. T. Stubbington, A. Regev, S. A. Teichmann, The Human Cell Atlas: from vision to reality. *Nature* **550**, 451–453 (2017).
8. C. I. Bargmann, W. T. Newsome, The Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) initiative and neurology. *JAMA Neurol.* **71**, 675–676 (2014).
9. I. H. de Boer, C. E. Alpers, E. U. Azeloglu, U. G. J. Balis, J. M. Barasch, L. Barisoni, K. N. Blank, A. S. Bomback, K. Brown, P. C. Dagher, A. L. Dighe, M. T. Eadon, T. M. El-Achkar, J. P. Gaut, N. Hacohen, Y. He, J. B. Hodgin, S. Jain, J. A. Kellum, K. Kiryluk, R. Knight, Z. G. Laszik, C. Lienczewski, L. H. Mariani, R. L. Mc Clelland, S. Menez, D. G. Moledina, S. D. Mooney, J. F. O'Toole, P. M. Palevsky, C. R. Parikh, E. D. Poggio, S. E. Rosas, M. R. Rosengart, M. M. Sarwal, J. A. Schaub, J. R. Sedor, K. Sharma, B. Steck, R. D. Toto, O. G. Troyanskaya, K. R. Tuttle, M. A. Vazquez, S. S. Waikar, K. Williams, F. P. Wilson, K. Zhang, R. Iyengar, M. Kretzler, J. Himmelfarb, Kidney Precision Medicine Project, Rationale and design of the Kidney Precision Medicine Project. *Kidney Int.* **99**, 498–510 (2021).
10. O. Rozenblatt-Rosen, A. Regev, P. Oberdoerffer, T. Nawy, A. Hupalowska, J. E. Rood, O. Ashenberg, E. Cerami, R. J. Coffey, E. Demir, L. Ding, E. D. Esplin, J. M. Ford, J. Goecks, S. Ghosh, J. W. Gray, J. Guinney, S. E. Hanlon, S. K. Hughes, E. S. Hwang, C. A. Iacobuzio-Donahue, J. Jané-Valbuena, B. E. Johnson, K. S. Lau, T. Lively, S. A. Mazzilli, D. Pe'er, S. Santagata, A. K. Shalek, D. Schapiro, M. P. Snyder, P. K. Sorger, A. E. Spira, S. Srivastava, K. Tan, R. B. West, E. H. Williams, Human Tumor Atlas Network, The Human Tumor Atlas Network: Chartingtumor transitions across space and time at single-cell resolution. *Cell* **181**, 236–249 (2020).
11. M. E. Colley, A. B. Esselman, C. F. Scott, J. M. Spraggins, High-specificity imaging mass spectrometry. *Annu. Rev. Anal. Chem.* **17**, 1–24 (2024).
12. K. V. Djambazova, J. M. Van Ardenne, J. M. Spraggins, Advances in imaging mass spectrometry for biomedical and clinical research. *Trends Analyt. Chem.* **169**, 117344 (2023).
13. A. B. Esselman, N. H. Patterson, L. G. Migas, M. Dufresne, K. V. Djambazova, M. E. Colley, R. Van de Plas, J. M. Spraggins, Microscopy-directed imaging mass spectrometry for rapid high spatial resolution molecular imaging of glomeruli. *J. Am. Soc. Mass Spectrom.* **34**, 1305–1314 (2023).
14. A. B. Esselman, F. A. Moser, L. Tideman, L. G. Migas, K. V. Djambazova, M. E. Colley, E. L. Pingry, N. H. Patterson, M. A. Farrow, H. Yang, A. B. Fogo, M. de Caestecker, R. Van de Plas, J. M. Spraggins, In situ molecular profiles of glomerular cells by integrated imaging mass spectrometry and multiplexed immunofluorescence microscopy. *Kidney Int.* **107**, 3320–3337 (2024).
15. N. H. Patterson, E. K. Neumann, K. Sharman, J. Allen, R. Harris, A. B. Fogo, M. de Caestecker, R. M. Caprioli, R. Van de Plas, J. M. Spraggins, Autofluorescence microscopy as a label-free tool for renal histology and glomerular segmentation. bioRxiv 452703 [Preprint] (2021).
16. C. D. Fryar, M. D. Carroll, Q. Gu, J. Afful, C. L. Ogden, Anthropometric reference data for children and adults: United States, 2015–2018. *Vital Health Stat.* **3**, 1–44 (2021).
17. M. F. Adamer, S. C. Brüningk, A. Tejada-Arranz, F. Estermann, M. Basler, K. Borgwardt, reComBat: Batch-effect removal in large-scale multi-source gene-expression data integration. *Bioinform. Adv.* **2**, vbac071 (2022).
18. L. McInnes, J. Healy, J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426 (2018).
19. L. E. M. Tideman, L. G. Migas, K. V. Djambazova, N. H. Patterson, R. M. Caprioli, J. M. Spraggins, R. Van de Plas, Automated biomarker candidate discovery in imaging mass spectrometry data through spatially localized Shapley additive explanations. *Anal. Chim. Acta* **1177**, 338522 (2021).
20. S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
21. S. Miyamoto, C.-C. Hsu, G. Hamm, M. Darshi, M. Diamond-Stanic, A.-E. Declèves, L. Slater, S. Pennathur, J. Stauber, P. C. Dorrestein, K. Sharma, Mass spectrometry imaging reveals elevated glomerular ATP/AMP in diabetes/obesity and identifies sphingomyelin as a possible mediator. *EBioMedicine* **7**, 121–134 (2016).
22. M. Habeck, H. Haviv, A. Katz, E. Kapri-Pardes, S. Ayciriex, A. Shevchenko, H. Ogawa, C. Toyoshima, S. J. D. Karlish, Stimulation, inhibition, or stabilization of Na,K-ATPase caused by specific lipid interactions at distinct sites. *J. Biol. Chem.* **290**, 4829–4842 (2015).
23. K. Nakashima, Y. Hirahara, T. Koike, S. Tanaka, K. Gamo, S. Oe, S. Hayashi, R. Seki-Omura, Y. Nakano, C. Ohe, T. Yoshida, Y. Kataoka, M. Tsuda, T. Yamashita, K. Honke, M. Kitada, Sulfatide with ceramide composed of phytosphingosine (t18:0) and 2-hydroxy FAs in renal intercalated cells. *J. Lipid Res.* **63**, 100210 (2022).
24. Z. Solati, A. L. Edel, Y. Shang, K. O, A. Ravandi, Oxidized phosphatidylcholines are produced in renal ischemia reperfusion injury. *PLOS ONE* **13**, e0195172 (2018).
25. R. G. Salomon, Structural identification and cardiovascular activities of oxidized phospholipids. *Circ. Res.* **111**, 930–946 (2012).
26. A. Eberhard, S. Kahlert, V. Goede, B. Hemmerlein, K. H. Plate, H. G. Augustin, Heterogeneity of angiogenesis and blood vessel maturation in human tumors: Implications for antiangiogenic tumor therapies. *Cancer Res.* **60**, 1388–1393 (2000).
27. M. Battaglia, S. Ahmed, M. S. Anderson, M. A. Atkinson, D. Becker, P. J. Bingley, E. Bosi, T. M. Brusko, L. A. DiMeglio, C. Evans-Molina, S. E. Gitelman, C. J. Greenbaum, P. A. Gottlieb, K. C. Herold, M. J. Hessner, M. Knip, L. Jacobsen, J. P. Krischer, S. A. Long, M. Lundgren, E. F. McKinney, N. G. Morgan, R. A. Oram, T. Pastinen, M. C. Peters, A. Petrelli, X. Qian, M. J. Redondo, B. O. Roep, D. Schatz, D. Skibinski, M. Peakman, Introducing the endotype concept to address the challenge of disease heterogeneity in type 1 diabetes. *Diabetes Care* **43**, 5–12 (2020).
28. R. A. Abeysekera, H. G. Healy, Z. Wang, A. L. Cameron, W. E. Hoy, Heterogeneity in patterns of progression of chronic kidney disease. *Intern. Med. J.* **51**, 220–228 (2021).
29. X.-G. Du, X.-Z. Ruan, Lipid metabolism disorder and renal fibrosis. *Adv. Exp. Med. Biol.* **1165**, 525–541 (2019).
30. C. K. Abrass, Cellular lipid metabolism and the role of lipids in progressive renal disease. *Am. J. Nephrol.* **24**, 46–53 (2004).
31. J. H. Lluesa, L. C. López-Romero, J. J. B. Monzó, M. R. Marugán, I. V. Boyano, D. Rodríguez-Espinosa, A. Gómez-Bori, A. S. Orient, R. D. Such, P. S. Perez, J. H. Jaras, Lipidic profiles of patients starting peritoneal dialysis suggest an increased cardiovascular risk beyond classical dyslipidemia biomarkers. *Sci. Rep.* **12**, 16394 (2022).

32. W.-L. Yang, Q. Bai, D.-D. Li, T.-L. A, S. Wang, R.-S. Zhao, H.-G. Nie, A.-H. Zhang, T. Wang, M.-H. Fan, Changes of urinary phospholipids in the chronic kidney disease patients. *Biomarkers* **18**, 601–606 (2013).
33. Y. Lu, Y. Ye, W. Bao, Q. Yang, J. Wang, Z. Liu, S. Shi, Genome-wide identification of genes essential for podocyte cytoskeletons based on single-cell RNA sequencing. *Kidney Int.* **92**, 1119–1129 (2017).
34. O. Engberg, K.-L. Lin, V. Hautala, J. P. Slotte, T. K. M. Nyholm, Sphingomyelin acyl chains influence the formation of sphingomyelin- and cholesterol-enriched domains. *Biophys. J.* **119**, 913–923 (2020).
35. K. J. Grove, P. A. Voziyan, J. M. Spraggins, S. Wang, P. Paueksakon, R. C. Harris, B. G. Hudson, R. M. Caprioli, Diabetic nephropathy induces alterations in the glomerular and tubule lipid profiles. *J. Lipid Res.* **55**, 1375–1385 (2014).
36. M. Valsecchi, V. Cazzetta, F. Oriolo, X. Lan, R. Piazza, M. A. Saleem, P. C. Singhal, D. Mavilio, J. Mikulak, M. Aureli, APOL1 polymorphism modulates sphingolipid profile of human podocytes. *Glycoconj. J.* **37**, 729–744 (2020).
37. J. Bernardino de la Serna, G. J. Schütz, C. Eggeling, M. Cebecauer, There is no simple model of the plasma membrane organization. *Front. Cell Dev. Biol.* **4**, 106 (2016).
38. G. Friedlander, M. Shahedi, C. Le Grimellec, C. Amiel, Increase in membrane fluidity and opening of tight junctions have similar effects on sodium-coupled uptakes in renal epithelial cells. *J. Biol. Chem.* **263**, 11183–11188 (1988).
39. M. F. A. Borghese, M. P. Majowicz, M. C. Ortiz, M. del Rosario Passalacqua, N. B. S. Speziale, N. A. Vidal, Expression and activity of SGLT2 in diabetes induced by streptozotocin: Relationship with the lipid environment. *Nephron Physiol.* **112**, p45–p52 (2009).
40. J. E. Markham, D. Molino, L. Gissot, Y. Bellec, K. Hématy, J. Marion, K. Belcram, J.-C. Palauqui, B. Satiat-JeuneMaître, J.-D. Faure, Sphingolipids containing very-long-chain fatty acids define a secretory pathway for specific polar plasma membrane protein targeting in *Arabidopsis*. *Plant Cell* **23**, 2362–2378 (2011).
41. H. Ebel, A. Fromm, D. Günzel, M. Fromm, J. D. Schulzke, Phospholipid effects on SGLT1-mediated glucose transport in rabbit ileum brush border membrane vesicles. *Biochim. Biophys. Acta Biomembr.* **1861**, 182985 (2019).
42. C. Ghezzi, E. M. Wright, Regulation of the human Na^+-dependent glucose cotransporter hSGLT2. *Am. J. Physiol. Cell Physiol.* **303**, C348–C354 (2012).
43. Y. Jiang, Z. Pan, J. W. Chen, Interaction between protein kinase C and sphingomyelin/cholesterol. *Cell Biol. Int.* **23**, 457–463 (1999).
44. V. Vallon, Glucose transporters in the kidney in health and disease. *Pflugers Arch.* **472**, 1345–1370 (2020).
45. V. Walker, Phosphaturia in kidney stone formers: Still an enigma. *Adv. Clin. Chem.* **90**, 133–196 (2019).
46. P. Stettner, S. Bourgeois, C. Marsching, M. Traykova-Brauch, S. Porubsky, V. Nordström, C. Hopf, R. Koesters, R. Sandhoff, H. Wiegandt, C. A. Wagner, H.-J. Gröne, R. Jennemann, Sulfatides are required for renal adaptation to chronic metabolic acidosis. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 9998–10003 (2013).
47. C. E. Chalfant, K. Kishikawa, M. C. Mumby, C. Kamibayashi, A. Bielawska, Y. A. Hannun, Long chain ceramides activate protein phosphatase-1 and protein phosphatase-2A. Activation is stereospecific and regulated by phosphatidic acid. *J. Biol. Chem.* **274**, 20313–20317 (1999).
48. R. O. Crajoinas, T. D. Pessoa, M. V. Rodrigues, G. Malnic, A. C. C. Girardi, Changes in the activity and expression of protein phosphatase-1 accompany the differential regulation of NHE3 before and after the onset of hypertension in spontaneously hypertensive rats. *Acta Physiol.* **211**, 395–408 (2014).
49. D. Li, A. Aperia, G. Celsi, E. F. da Cruz e Silva, P. Greengard, B. Meister, Protein phosphatase-1 in the kidney: Evidence for a role in the regulation of medullary Na(+)-K(+)-ATPase. *Am. J. Physiol.* **269**, F673–F680 (1995).
50. A. R. Subramanya, D. H. Ellison, Distal convoluted tubule. *Clin. J. Am. Soc. Nephrol.* **9**, 2147–2163 (2014).
51. J. Zhang, X. Wang, V. Vikash, Q. Ye, D. Wu, Y. Liu, W. Dong, ROS and ROS-mediated cellular signaling. *Oxid. Med. Cell. Longev.* **2016**, 4350965 (2016).
52. S. Takeshita, N. Inoue, D. Gao, Y. Rikitake, S. Kawashima, R. Tawa, H. Sakurai, M. Yokoyama, Lysophosphatidylcholine enhances superoxide anions production via endothelial NADH/NADPH oxidase. *J. Atheroscler. Thromb.* **7**, 238–246 (2000).
53. Y.-J. Im, Y.-K. Lee, H.-Y. Chung, D.-S. Im, Multiple actions of lysophosphatidylcholine in human Jurkat T cells1. *Acta Pharmacol Sin.* **27**, 700–707 (2006).
54. A. Gonzalez-Vicente, N. Hong, J. L. Garvin, Effects of reactive oxygen species on renal tubular transport. *Am. J. Physiol. Renal Physiol.* **317**, F444–F455 (2019).
55. C. Hirota, Y. Takashina, Y. Yoshino, H. Hasegawa, E. Okamoto, T. Matsunaga, A. Ikari, Reactive oxygen species downregulate transient receptor potential melastatin 6 expression mediated by the elevation of miR-24-3p in renal tubular epithelial cells. *Cells* **10**, 1893 (2021).
56. S. Ghazi, S. Bourgeois, A. Gomariz, M. Bugarski, D. Haenni, J. R. Martins, C. Nombela-Arrieta, R. J. Unwin, C. A. Wagner, A. M. Hall, E. Craigie, Multiparametric imaging reveals that mitochondria-rich intercalated cells in the kidney collecting duct have a very high glycolytic capacity. *FASEB J.* **34**, 8510–8525 (2020).
57. S. E. Horvath, G. Daum, Lipids of mitochondria. *Prog. Lipid Res.* **52**, 590–614 (2013).
58. A. Poulaki, S. Giannouli, Mitochondrial lipids: From membrane organization to apoptotic facilitation. *Int. J. Mol. Sci.* **23**, 3738 (2022).
59. B. Školová, A. Kováčik, O. Tesař, L. Opálka, K. Vávrová, Phytosphingosine, sphingosine and dihydrosphingosine ceramides in model skin lipid membranes: Permeability and biophysics. *Biochim. Biophys. Acta Biomembr.* **1859**, 824–834 (2017).
60. S. Guo, T. C. Moore, C. R. Iacovella, L. A. Strickland, C. McCabe, Simulation study of the structure and phase behavior of ceramide bilayers and the role of lipid headgroup chemistry. *J. Chem. Theory Comput.* **9**, 5116–5126 (2013).
61. I. de la Arada, E. J. González-Ramírez, A. Alonso, F. M. Goñi, J.-L. R. Arrondo, Exploring polar headgroup interactions between sphingomyelin and ceramide with infrared spectroscopy. *Sci. Rep.* **10**, 17606 (2020).
62. L. M. Fischer, K.-A. da Costa, L. Kwock, J. Galanko, S. H. Zeisel, Dietary choline requirements of women: Effects of estrogen and genetic variation. *Am. J. Clin. Nutr.* **92**, 1113–1119 (2010).
63. J. C. Sullivan, J. M. Sasser, D. M. Pollock, J. S. Pollock, Sexual dimorphism in renal production of prostanoids in spontaneously hypertensive rats. *Hypertension* **45**, 406–411 (2005).
64. H. B. Beyene, G. Olshansky, A. A. T. Smith, C. Giles, K. Huynh, M. Cinel, N. A. Mellett, G. Cadby, J. Hung, J. Hui, J. Beilby, G. F. Watts, J. S. Shaw, E. K. Moses, D. J. Magliano, P. J. Meikle, High-coverage plasma lipidomics reveals novel sex-specific lipidomic fingerprints of age and BMI: Evidence from two large population cohort studies. *PLoS Biol.* **18**, e3000870 (2020).
65. P. E. N. R. Bellot, M. N. Moia, B. Z. Reis, L. F. C. Pedrosa, L. Tasic, F. Barbosa Jr., K. C. M. Sena-Evangelista, Are phosphatidylcholine and lysophosphatidylcholine body levels potentially reliable biomarkers in obesity? A review of human studies. *Mol. Nutr. Food Res.* **67**, e2200568 (2023).
66. F. Samad, W. Ruf, Inflammation, obesity, and thrombosis. *Blood* **122**, 3415–3422 (2013).
67. A. Miszta, H. K. Ahmadzia, N. L. C. Luban, S. Li, D. Guo, L. A. Holle, J. S. Berger, A. H. James, J. V. S. Gobburu, J. van den Anker, B. de Laat, A. S. Wolberg, Application of a plasmin generation assay to define pharmacodynamic effects of tranexamic acid in women undergoing cesarean delivery. *J. Thromb. Haemost.* **19**, 221–232 (2021).
68. V. Bochkov, B. Gesslbauer, C. Mauerhofer, M. Philippova, P. Erne, O. V. Oskolkova, Pleiotropic effects of oxidized phospholipids. *Free Radic. Biol. Med.* **111**, 6–24 (2017).
69. A. K. Kopec, S. R. Abrahams, S. Thornton, J. S. Palumbo, E. S. Mullins, S. Divanovic, H. Weiler, A. P. Owens, N. Mackman, A. Goss, J. van Ryn, J. P. Luyendyk, M. J. Flick, Thrombin promotes diet-induced obesity through fibrin-driven inflammation. *J. Clin.* **127**, 3152–3166 (2017).
70. M. Ząbczyk, J. Natorska, A. Undas, Fibrin clot properties in atherosclerotic vascular disease: From pathophysiology to clinical outcomes. *J. Clin. Med.* **10**, 2999 (2021).
71. N. H. Patterson, M. Tuck, A. Lewis, A. Kaushansky, J. L. Norris, R. Van de Plas, R. M. Caprioli, Next generation histology-directed imaging mass spectrometry driven by autofluorescence microscopy. *Anal. Chem.* **90**, 12404–12413 (2018).
72. M. J. Conroy, R. M. Andrews, S. Andrews, L. Cockayne, E. A. Dennis, E. Fahy, C. Gaud, W. J. Griffiths, G. Jukes, M. Kolchin, K. Mendivelso, A. F. Lopez-Clavijo, C. Ready, S. Subramaniam, V. B. O'Donnell, LIPID MAPS: Update to databases and tools for the lipidomics community. *Nucleic Acids Res.* **52**, D1677–D1682 (2024).
73. M. Sud, E. Fahy, D. Cotter, A. Brown, E. A. Dennis, C. K. Glass, A. H. Merrill, R. C. Murphy, C. R. H. Raetz, D. W. Russell, S. Subramaniam, LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **35**, D527–D532 (2007).
74. V. Matyash, G. Liebisch, T. V. Kurzchalia, A. Shevchenko, D. Schwudke, Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.* **49**, 1137–1146 (2008).
75. H. Tsugawa, K. Ikeda, M. Takahashi, A. Satoh, Y. Mori, H. Uchino, N. Okahashi, Y. Yamada, I. Tada, P. Bonini, Y. Higashi, Y. Okazaki, Z. Zhou, Z.-J. Zhu, J. Koelmel, T. Cajka, O. Fiehn, K. Saito, M. Arita, M. Arita, A lipidome atlas in MS-DIAL 4. *Nat. Biotechnol.* **38**, 1159–1163 (2020).
76. P. Monchamp, L. Cetto, J. Y. Zhang, R. Henson, "Signal processing methods for mass spectrometry" in *Systems Bioinformatics: An Engineering Case-Based Approach* (Artech, 2007), pp. 101–124.
77. L. Migas, "msalign"; msalign.lukasz-migas.com/.
78. L. McInnes "Uniform manifold approximation and projection"; https://github.com/lmcinnes/umap.
79. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
80. T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system," in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, 2016).

Acknowledgments: We thank the Hubmap data curation and ingest team, particularly B. Honick, P. Blood, and J. Silverstein, for the support. **Funding:** This work was supported by the following: National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) U54DK120058 (to J.M.S., R.M.C., and R.V.d.P.), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) U54DK134302 (to J.M.S., R.M.C., and R.V.d.P.), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) U01DK133766 (to J.M.S., R.M.C., and R.V.d.P.), National Eye Institute U54EY032442 (to J.M.S., R.M.C., and R.V.d.P.), National Institute of Allergy and Infectious Diseases (NIAID) R01AI138581 (to J.M.S. and R.V.d.P.), National Institute of Allergy and Infectious Diseases (NIAID) R01AI145992 (to J.M.S. and R.V.d.P.), National Institute on Aging (NIA) under award number R01AG078803 (to J.M.S. and R.V.d.P.), National Science Foundation Major Research Instrument Program CBET 1828299 (to J.M.S. and R.M.C.), Chan Zuckerberg Initiative DAF grant number 2021-240339 (to L.G.M. and R.V.d.P.), and Chan Zuckerberg Initiative DAF grant number 2022-309518 (to L.G.M. and R.V.d.P.). **Author contributions:** Conceptualization: M.A.F., E.K.N., N.H.P., R.C.H., R.M.C., R.V.d.P., and J.M.S. Methodology: M.A.F., L.E.M.T., E.K.N., N.H.P., M.E.C., J.L.A., E.L.P., M.D., M.B., M.P.d.C., R.M.C., R.V.d.P., and J.M.S. Investigation: M.A.F., E.K.N., M.E.C., J.L.A., H.Y., C.E.R., K.D., A.R.S.K., R.M.C., and J.M.S. Formal analysis: M.A.F., L.E.M.T., E.K.N., L.G.M., N.H.P., M.E.C., H.Y., R.V.d.P., and J.M.S. Resources: M.A.F., E.K.N., M.E.C., L.G.M., J.L.A., M.D., M.B., A.B.F., M.P.d.C., R.M.C., R.V.d.P., and J.M.S. Funding acquisition: M.A.F., R.C.H., M.P.d.C., R.M.C., R.V.d.P., and J.M.S. Project administration: M.A.F., E.K.N., M.P.d.C., R.M.C., R.V.d.P., and J.M.S. Supervision: M.A.F., E.K.N., M.P.d.C., R.M.C., R.V.d.P., and J.M.S. Visualization: M.A.F., L.E.M.T., L.G.M., N.H.P., M.E.C., J.L.A., K.S., M.P.d.C., and J.M.S. Data curation: M.A.F., L.E.M.T., L.G.M., N.H.P., M.E.C., H.Y., E.S.R., and J.M.S. Validation: M.A.F., N.H.P., M.E.C., M.D., M.B., M.P.d.C., and J.M.S. Software: L.E.M.T., L.G.M., N.H.P., and E.L.P. Writing—original draft: M.A.F., L.E.M.T., E.K.N., N.H.P., K.D., and J.M.S. Writing—review and editing: M.A.F., L.E.M.T., L.G.M., M.E.C., M.D., H.Y., K.D., D.B.G., R.C.H., A.B.F., M.P.d.C., R.V.d.P., and J.M.S. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 23 November 2024
Accepted 5 May 2025
Published 11 June 2025
10.1126/sciadv.adu3730