Data in Brief 54 (2024) 110431

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Spatially resolved metabolomic dataset of distinct human kidney anatomic regions

Haikuo Li[a], Benjamin D. Humphreys[a,b,*]

[a] Division of Nephrology, Department of Medicine, Washington University in St. Louis, St. Louis, MO, USA
[b] Department of Developmental Biology, Washington University in St. Louis, St. Louis, MO, USA

ARTICLE INFO

Article history:
Received 22 March 2024
Revised 9 April 2024
Accepted 11 April 2024
Available online 16 April 2024

Dataset link: A dataset of spatially resolved metabolomics on human kidney anatomic regions (Original data)
Dataset link: Online visualizer of spatially resolved metabolomics data (Original data)

Keywords:
MALDI
Imaging mass spectrometry
Kidney anatomy
Nephrology
Metabolism
Omics

ABSTRACT

Cortex, medulla and papilla are three major human kidney anatomic structures and they harbour unique metabolic functions, but the underlying metabolomic profiles are largely unknown at spatial resolution. Here, we generated a spatially resolved metabolomics dataset on human kidney cortex, medulla and papilla tissues dissected from the same donor. Matrix-Assisted Laser Desorption/Ionization-Imaging Mass Spectrometry (MALDI-IMS) was used to detect metabolite species over mass-to-charge ratios of 50 -1500 for each section at a resolution of 10 × 10 μm^2 pixel size. We present raw data matrix of each sample, feature annotations, raw AnnData merged from three samples and processed AnnData files after quality control, dimensional reduction and data integration, which contains a total of 170,459 spatially resolved metabolomes with 562 features detected. This dataset can be either visualized through an interactive browser or further analyzed to study metabolomic heterogeneity across regional human kidney anatomy.

* Corresponding author at: Division of Nephrology, Department of Medicine, Washington University in St. Louis, St. Louis, MO, USA.
E-mail address: humphreysbd@wustl.edu (B.D. Humphreys).
Social media: @HumphreysLab (B.D. Humphreys)

https://doi.org/10.1016/j.dib.2024.110431

Specifications Table

Subject	Biological Sciences: Omics (General)
Specific subject area	Spatially resolved metabolomics by Matrix-Assisted Laser Desorption/Ionization-Imaging Mass Spectrometry (MALDI-IMS) on human kidney cortex, medulla and papilla
Data format	Raw, Analyzed
Type of data	Count matrix, AnnData
Data collection	Kidney cortex, medulla and papilla were dissected from the same donor and one 10-µm section for each tissue was selected for MALDI-IMS analysis. Data acquisition was performed with the positive ion mode at a pixel size of 10 × 10 µm^2 over m/z range 50–1500. The matrix used was 2,5-dihydroxybenzoic acid (DHB). Matrix application was performed with the HTX M5 Sprayer and data acquisition was performed on the timsTOF fleX MALDI-2 instrument. Data were uploaded to METASPACE for annotation with the CoreMetabolome database. Downstream analysis was performed based on our recently described package MALDIpy.
Data source location	Institution: Washington University in St. Louis City/Town/Region: St. Louis, Missouri Country: USA Latitude and longitude (and GPS coordinates) for collected samples/data: 38.63443, -90.26293 (38° 38′ 3.948″ N, 90° 15′ 46.548″ W)
Data accessibility	Repository name: METASPACE; Mendeley DataData identification number: 1. METASPACE: https://metaspace2020.eu/project/human_kidney_region 2. Mendeley Data: https://doi.org/10.17632/hss5zczhrk Direct URL to data: 1. METASPACE: https://metaspace2020.eu/project/human_kidney_region 2. Mendeley Data: https://data.mendeley.com/datasets/hss5zczhrk/1
Related research article	Li et al. [1]

1. Value of the Data

- We generated spatially resolved metabolomics data with MALDI-IMS on different human kidney anatomic regions (cortex, medulla, papilla) dissected from the same donor. This dataset allows researchers to investigate the metabolomic signatures underlying regional human kidney anatomy.
- We analyzed a total of 170,459 spatially resolved metabolomes (10 × 10 µm^2 pixels) across the three kidney anatomic regions after quality control, with a total of 562 features (metabolites and small molecules) detected.
- With our recently described MALDI-IMS analytical package [1], we present both raw and processed metabolomic data matrices in the AnnData format, which allow researchers to perform customized downstream analysis. The processed AnnData contains spatial information (x/y coordinates), uniform manifold approximation and projection (UMAP) coordinates and Leiden annotations for all metabolomes, as well as mass-to-charge (m/z) ratios, chemical formulas and molecule identities for all features.
- The dataset is available on the METASPACE browser [2], allowing researchers to interactively visualize any metabolites of interest in each human kidney anatomic region.

2. Background

The human kidneys remove waste products from the blood and regulate electrolyte balance, blood pressure and various metabolic processes [3]. Nephron is the kidney functional unit and it spans cortex, medulla and papilla, the three major anatomic regions of the human kidney, but the underlying metabolomic signature across kidney anatomic regions is still incompletely understood. Although several metabolomics analyses have been conducted on primary kidney

Table 1
Number of detected features in the three kidney anatomic regions. Putative annotations refer to all annotations derived from the CoreMetabolome database. The list of unique species excludes metabolite species with multiple adducts (e.g., H+, Na+, K+).

Sample	Feature number	
Cortex	Number of putative annotations	311
	Number of unique species	217
Medulla	Number of putative annotations	274
	Number of unique species	205
Papilla	Number of putative annotations	254
	Number of unique species	184

samples [4], spatial information is lost in these bulk assays since the whole tissue is completed lysed [5]. The MALDI-IMS technology is the current widely adopted solution to spatially resolved metabolomics analysis and has been used in multiple biomedicine disciplines [6–10]. Although MALDI-IMS analysis on mouse kidneys [11,12], kidney organoids [13] and a human kidney cortex sample [14] have been recently reported, a spatially resolved metabolomics dataset on the three human kidney anatomic regions dissected from the same donor is still highly demanded. Furthermore, we recently described an analytical package for MALDI-IMS analysis, which we called MALDIpy [1], but additional publicly available datasets are needed for further evaluation of the efficacy of this package.

3. Data Description

For each sample, the raw matrix of spatially resolved metabolomics data, with false discovery rate (FDR) $<$ 20 % and with total ion count (TIC) normalization, is presented (compressed in a .zip file), where each row indicates a feature (annotation with the CoreMetabolome database v3 [2]) and each column indicates a x/y coordinate pair. Researchers can also export customized data matrices (e.g., choosing different FDR thresholds or no TIC normalization) through the METASPACE browser with the link provided above.

Feature annotations are presented in a spreadsheet (.xlsx file), which describes metabolite species detected in each sample, including their m/z ratios, chemical formulas, ion adducts and molecule identities. Table 1 summarizes the number of detected features in each MALDI-IMS sample.

A raw Anndata (.h5ad file) for merged three MALDI-IMS data matrices is presented, containing a total of 370,662 metabolomes (i.e., 10 $\times$ 10 μm^2 pixels) and 562 features. Cortex, medulla and papilla samples contain 177,870, 92,120 and 100,672 metabolomes, respectively. With this raw data, researchers can visualize any metabolite species of interest across the three tissue sections with the MALDIpy package. For example, Fig. 1A presents a sphingomyelin (d18:0/16:1(9Z)) (C39H79N2O6PNa) which is more specific to the kidney glomerulus structure in the cortex, and Fig. 1B presents a phosphatidylcholine (16:0/18:1(11Z)) (C42H82NO8PNa) which indicates tubular epithelia that are more abundant in the medulla and papilla. Of note, this raw Anndata contains significant matrix background and low-quality metabolomes which are removed in the processed Anndata described below.

A processed Anndata (.h5ad file) is presented, containing a total of 170,459 metabolomes and 562 features. Cortex, medulla and papilla samples have 79,814, 49,771 and 40,874 metabolomes included in this processed data, respectively (Fig. 2A). All metabolomes in this processed file have a minimum count number of 40,000 and a minimum feature number of 30 detected (Fig. 2B). Other quality control procedures, including removal of matrix background observations and low-quality metabolomes, were performed (see Methods). MALDIpy-based downstream data processing, multi-sample integration and dimensional reduction were also performed to gen-

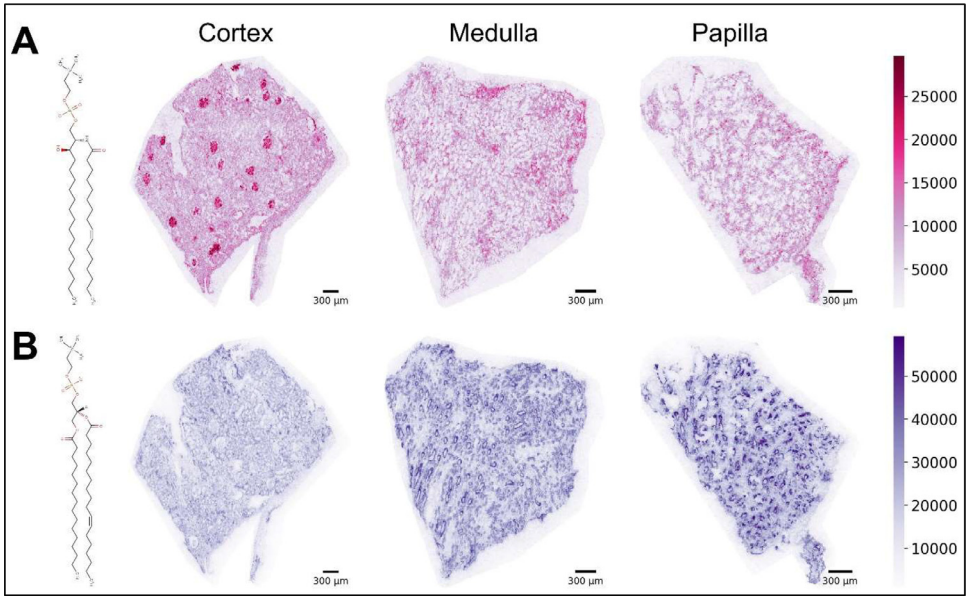

Fig. 1. Spatial feature plot of two metabolite species across human kidney anatomic regions. (A) MALDI-IMS intensities of sphingomyelin (d18:0/16:1(9Z)) with its chemical structure shown on the left. (B) MALDI-IMS intensities of phosphatidylcholine (16:0/18:1(11Z)) with its chemical structure shown on the left.

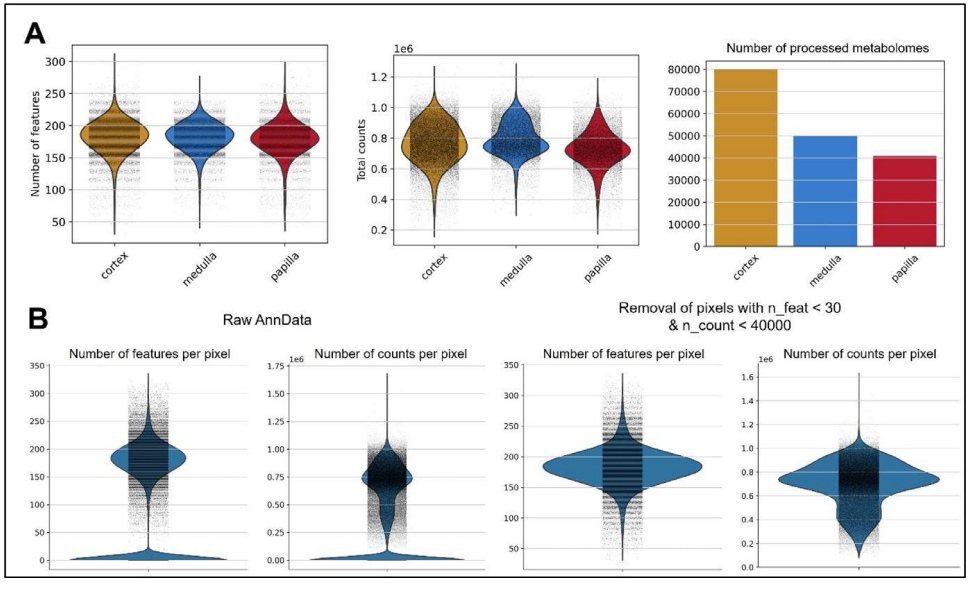

Fig. 2. Quality control metrics of the presented AnnData. (A) Number of features per pixel (left), number of counts per pixel (middle) and total number of metabolomes of each kidney anatomic region (right) in the processed AnnData. (B) Number of features and counts per pixel in the raw AnnData (left) and number of features and counts per pixel after quality control.

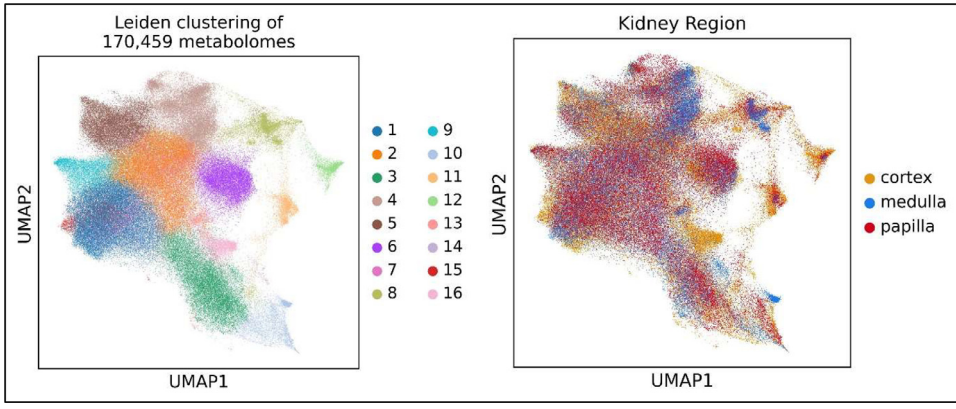

Fig. 3. UMAP presentation of 170,459 spatially resolved metabolomes, colored by curated Leiden clusters (left) or kidney anatomic regions (right).

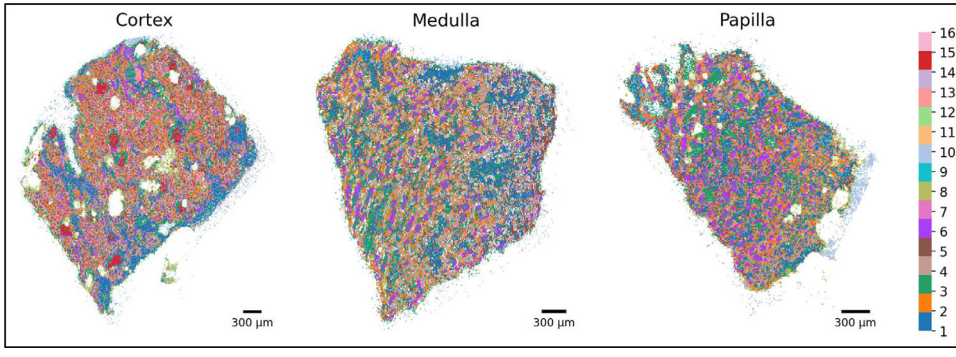

Fig. 4. Projection of the 16 Leiden clusters (presented in Fig. 3) onto sections of the three human kidney anatomic regions.

erate this data. A total of 16 clusters were identified in Leiden clustering analysis. Cluster identities and anatomic region origins are included in the Anndata.obs columns (Fig. 3). With this processed data, researchers can analyze cluster distributions across the three kidney anatomic regions. As shown in Fig. 4, Cluster #15 metabolomes (colored in red) specifically mark kidney glomeruli in the cortex, while Cluster #6 metabolomes (colored in purple) indicate tubular epithelia that are more abundant in the medulla and papilla.

4. Experimental Design, Materials and Methods

4.1. Sample collection

Human kidney cortical, medullary and papillary tissues were dissected from the same donor. This deceased organ donor is a 67-year-old female with normal kidney functions (with creatinine levels at sampling: 0.6 mg/dL and with mild interstitial fibrosis). Fat and the renal capsule were first removed and the kidney was cut sagittally with a trimming blade. Then, kidney tissues of different anatomic regions were dissected with a scalpel. Each tissue was frozen in a cryotube

with liquid nitrogen immediately after dissection and stored at -80°C before the MALDI-IMS experiment.

4.2. MALDI-IMS data acquisition

The three tissues were proceeded with unbiased MALDI-IMS profiling with a publicly established protocol at the Mass Spectrometry Technology Access Center at Washington University School of Medicine. Here, the tissues were sectioned at 10 μm thickness and sections of use were applied on one single MALDI IntelliSlide (Bruker) to reduce technical batch effects. We used 2,5-dihydroxybenzoic acid (DHB) as the MALDI matrix and matrix application was performed with the HTX M5 Sprayer (HTX Technologies). The timsTOF fleX MALDI-2 instrument (Bruker) was used for MALDI-IMS data acquisition. The laser was rastered over a tissue section and mass spectra were recorded in each spot. Spraying parameters were: 60 °C nozzle temperature, a flow rate of 0.1 mL/min, 1200 mm/min velocity, a track spacing of 3 mm, moving pattern CC, 14 passes and a N2 pressure of 10 psi. In data acquisition, the positive ion mode was chosen at a pixel size of $10 \times 10\ \mu m^2$ over m/z range 50–1500.

4.3. MALDI-IMS data processing

Metabolite annotation was performed with the METASPACE database of core mammalian metabolites and lipids (CoreMetabolome database v3) (https://metaspace2020.eu/)[2] and a data matrix with FDR < 20 % and TIC normalization was exported for each sample. Next, we created a MALDIpy object for each sample with the msi_data function (MALDIpy version 0.1.5; https://pypi.org/project/MALDIpy/). We then converted each MALDIpy object to an AnnData with the to_adata function (add_meta=True), in which each metabolome was considered as an observation and each feature was considered as a variable. This enabled us to perform downstream data quality control, normalization, dimensional reduction and clustering analysis with Scanpy [15]. Quality control metrics were computed with the scanpy.pp.calculate_qc_metrics function (percent_top=None, log1p=False, inplace=True). As shown in Fig. 2B, during quality control, we found artefactual metabolomes could be efficiently removed with the (n_feat < 30, n_count < 40,000) thresholds. After normalization with the scanpy.pp.normalize_total function, the effect of total counts was regressed out and the data was scaled [MALDIpy.single_cell.maldi_norm function, (regress_out_key='total_counts')).

4.4. MALDI-IMS data downstream analysis

Dimensional reduction, multi-sample integration and clustering analysis were performed with the MALDIpy.single_cell.maldi_clustering function, in which we used Harmony (Harmonypy v0.0.6) [16] for data integration and batch effect elimination. During this pipeline, we computed a neighborhood graph of all metabolomes using the scanpy.pp.neighbors function with the number of neighbors set as 30 and the number of components set as 30. Then we calculated the UMAP space with the effective minimum distance between embedded points set as 0.2. Leiden clustering was performed with a resolution of 0.8. We used the MALDIpy.projection.project_cluster_in_groups function to visualize the Leiden clusters across the three tissue sections [17]. Next, we removed away clusters that indicate MALDI background artefacts and section edge artefacts, and re-ran the same analytical pipeline on a total of 170,459 metabolomes, with the final curated Leiden clustering presented in Fig. 3A. Next, we used the scanpy.tl.rank_genes_groups function (method = 'wilcoxon') to identify differential marker features of each cluster. We used the MALDIpy.projection.project3tissues function to visualize the

clusters across the three tissue sections. For Fig. 1, the MALDIpy.multi_sample.plot3tissues function was used for visualization of a feature across all three tissue sections concurrently, in which a linear normalization based on total count number of each sample was included.

Limitations

We provide a spatially resolved metabolomic dataset obtained from kidney samples dissected from a single donor. A larger sample cohort would be required to conduct a comprehensive survey of metabolomics across the human kidney anatomy. In addition, metabolite species outside the m/z range of 50–1500 cannot be detected by this method.

Ethics Statement

This research complies with all relevant ethical regulations and has been approved by the Washington University Institutional Review Board. Discarded human kidney samples were obtained from deceased the organ donor under an established Institutional Review Board (protocol 201601020) approved by Washington University in St. Louis.

Data Availability

A dataset of spatially resolved metabolomics on human kidney anatomic regions (Original data) (Mendeley Data).

Online visualizer of spatially resolved metabolomics data (Original data) (METASPACE).

CRediT Author Statement

Haikuo Li: Conceptualization, Methodology, Data curation, Writing – original draft; **Benjamin D. Humphreys:** Supervision, Writing – review & editing.

Acknowledgments

These experiments were funded by NIH grants DK103740, U54DK137332 and UC2DK126024 to BDH. The authors acknowledge the Washington University Mass Spectrometry Technology Access Center for imaging mass spectrometry technology support.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Li, D. Li, N. Ledru, Q. Xuanyuan, H. Wu, A. Asthana, L.N. Byers, S.G. Tullius, G. Orlando, S.S. Waikar, B.D. Humphreys, Transcriptomic, epigenomic, and spatial metabolomic cell profiling redefines regional human kidney anatomy, Cell Metab. (2024) Online ahead of print, doi:10.1016/J.CMET.2024.02.015.

[2] A. Palmer, P. Phapale, I. Chernyavsky, R. Lavigne, D. Fay, A. Tarasov, V. Kovalev, J. Fuchser, S. Nikolenko, C. Pineau, M. Becker, T. Alexandrov, FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry, Nat. Methods (2016) 57–60 2016 141 14, doi:10.1038/nmeth.4072.

[3] P.A. Gallardo, C.P. Vio, Functional anatomy of the kidney, Ren. Physiol. Hydrosaline Metab. (2022), doi:10.1007/978-3-031-10256-1_2.

[4] S. Kalim, E.P. Rhee, An overview of renal metabolomics, Kidney Int. 91 (2017) 61, doi:10.1016/J.KINT.2016.08.021.
[5] H. Li, B.D. Humphreys, Single cell technologies: beyond microfluidics, Kidney 360 (2) (2021) 1196–1204, doi:10.34067/kid.0001822021.
[6] L.R. Conroy, H.A. Clarke, D.B. Allison, S.S. Valenca, Q. Sun, T.R. Hawkinson, L.E.A. Young, J.E. Ferreira, A.V Hammonds, J.B. Dunne, R.J. McDonald, K.J. Absher, B.E. Dong, R.C. Bruntz, K.H. Markussen, J.A. Juras, W.J. Alilain, J. Liu, M.S. Gentry, P.M. Angel, C.M. Waters, R.C. Sun, Spatial metabolomics reveals glycogen as an actionable target for pulmonary fibrosis, Nat. Commun. 14 (2023) 2759, doi:10.1038/s41467-023-38437-1.
[7] P. Zheng, N. Zhang, D. Ren, C. Yu, B. Zhao, Y. Zhang, Integrated spatial transcriptome and metabolism study reveals metabolic heterogeneity in human injured brain, Cell Rep. Med. 4 (2023) 101057, doi:10.1016/J.XCRM.2023.101057.
[8] T. Alexandrov, Spatial metabolomics and imaging mass spectrometry in the age of artificial intelligence, 3 (2020) 61–87. https://doi.org/10.1146/ANNUREV-BIODATASCI-011420-031537.
[9] T. Alexandrov, Spatial metabolomics: from a niche field towards a driver of innovation, Nat. Metab. 2023 (2023) 1–3, doi:10.1038/s42255-023-00881-0.
[10] S. Ma, Y. Leng, X. Li, Y. Meng, Z. Yin, W. Hang, High spatial resolution mass spectrometry imaging for spatial metabolomics: Advances, challenges, and future perspectives, TrAC Trends Anal. Chem. 159 (2023) 116902, doi:10.1016/J.TRAC.2022.116902.
[11] R.G.J. Rietjens, G. Wang, A.I.M. van der Velden, A. Koudijs, M.C. Avramut, S. Kooijman, P.C.N. Rensen, J. van der Vlag, T.J. Rabelink, B. Heijs, B.M. van den Berg, Phosphatidylinositol metabolism of the renal proximal tubule S3 segment is disturbed in response to diabetes, Sci. Rep. (2023) 1–12 2023 131 13, doi:10.1038/s41598-023-33442-2.
[12] G. Wang, B. Heijs, S. Kostidis, A. Mahfouz, R.G.J. Rietjens, R. Bijkerk, A. Koudijs, L.A.K. van der Pluijm, C.W. van den Berg, S.J. Dumas, P. Carmeliet, M. Giera, B.M. van den Berg, T.J. Rabelink, Analyzing cell-type-specific dynamics of metabolism in kidney repair, Nat. Metab. 2022 (2022) 1–10, doi:10.1038/s42255-022-00615-8.
[13] G. Wang, B. Heijs, S. Kostidis, R.G.J. Rietjens, M. Koning, L. Yuan, G.L. Tiemeier, A. Mahfouz, S.J. Dumas, M. Giera, J. Kers, S.M. Chuva de Sousa Lopes, C.W. van den Berg, B.M. van den Berg, T.J. Rabelink, Spatial dynamic metabolomics identifies metabolic cell fate trajectories in human kidney differentiation, Cell Stem Cell 29 (2022) 1580–1593 e7, doi:10.1016/j.stem.2022.10.008.
[14] 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.Todd Valerius, S. Jain, L.M. Satlin, O.G. Troyanskaya, M. Kretzler, R. Iyengar, E.U. Azeloglu, A reference tissue atlas for the human kidney, Sci. Adv. 8 (2022) 4965, doi:10.1126/sciadv.abn4965.
[15] F.A. Wolf, P. Angerer, F.J. Theis, SCANPY: large-scale single-cell gene expression data analysis, Genome Biol. 19 (2018) 1–5 2018 191, doi:10.1186/S13059-017-1382-0.
[16] I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P. ru Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony, Nat. Methods 16 (2019) 1289–1296 2019 1612, doi:10.1038/s41592-019-0619-0.
[17] V.A. Traag, L. Waltman, N.J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities, Sci. Rep. 9 (2019) 1–12 2019 91, doi:10.1038/s41598-019-41695-z.