



RESEARCH ARTICLE

Scholarly publications and data set evidence for the Human Reference Atlas

Yongxin Kong^{1,2} , Vicky Amar Daiya³ , and Katy Börner¹ 

¹Department of Intelligent Systems Engineering, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA

²School of Information Management, Sun Yat-sen University, Guangzhou, China

³Department of Computer Science, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA

an open access  journal



Citation: Kong, Y., Daiya, V. A., & Börner, K. (2024). Scholarly publications and data set evidence for the Human Reference Atlas. *Quantitative Science Studies*, 5(1), 246–260. https://doi.org/10.1162/qss_a_00299

DOI: https://doi.org/10.1162/qss_a_00299

Peer Review: https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss_a_00299

Received: 11 August 2023
Accepted: 28 January 2024

Corresponding Authors:
Yongxin Kong
yokong@iu.edu

Katy Börner
katy@indiana.edu

Handling Editor:
Vincent Larivière

Keywords: data visualization, Human Reference Atlas, scientometrics

ABSTRACT

Experts from 17 consortia are collaborating on the Human Reference Atlas (HRA) which aims to map the human body at single cell resolution. To bridge across scales—from the meter size human body to the micrometer size single-cell level—organ experts are constructing anatomical structures, cell types plus biomarkers (ASCT+B) tables, and associated spatial reference objects. The 3rd HRA (v1.2) release features 26 organ-specific ASCT+B tables that cite 456 scholarly papers and are linked to 61 spatial reference objects and Organ Mapping Antibody Panels (OMAPs); it is authored by more than 120 experts. This paper presents the first analyses and visualizations showcasing what data and scholarly evidence exist for which organs and how experts relate to the organs covered in the HRA. To identify potential HRA authors and reviewers, we queried the Web of Science database for authors who work on the 33 organs targeted for the next HRA release (v1.3). To provide scientific evidence for the HRA, we identified 620 high-quality, single-cell experimental data sets for 58 organs published in 561 unique papers. The results presented are critical for understanding and communicating the quality of the HRA, planning for future tissue data collection, and inviting leading experts to contribute to the evolving atlas.

1. INTRODUCTION

Constructing an atlas of the healthy human body at the single-cell level is a massive undertaking that requires close collaboration by researchers and practitioners with expertise in human anatomy, pathology, surgery, and single cell studies. Data sets at different levels of spatial scale—from computerized tomography and magnetic resonance imaging (MRI) scans at the whole-body level to single-cells data assays at the biomolecular level—need to be federated and combined to construct a multilevel atlas. Figure 1 illustrates a zoom from the whole body into the kidney organ, into the nephron, and to the single-cell level.

Supported by the National Institutes of Health and other funders, experts from 17 consortia are working on the HRA (Börner, Teichmann et al., 2021). The HRA captures ontology-aligned terms for naming anatomical structures (AS) and cell types (CT) plus biomarkers (B), in so-called ASCT+B tables. It links these terms to two-dimensional and three-dimensional spatial representations of major anatomical structures and the cell types commonly located in these, and the biomarkers (gene, protein, lipid, metabolites) used to characterize cell types. The HRA

Copyright: © 2024 Yongxin Kong, Vicky Amar Daiya, and Katy Börner.
Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



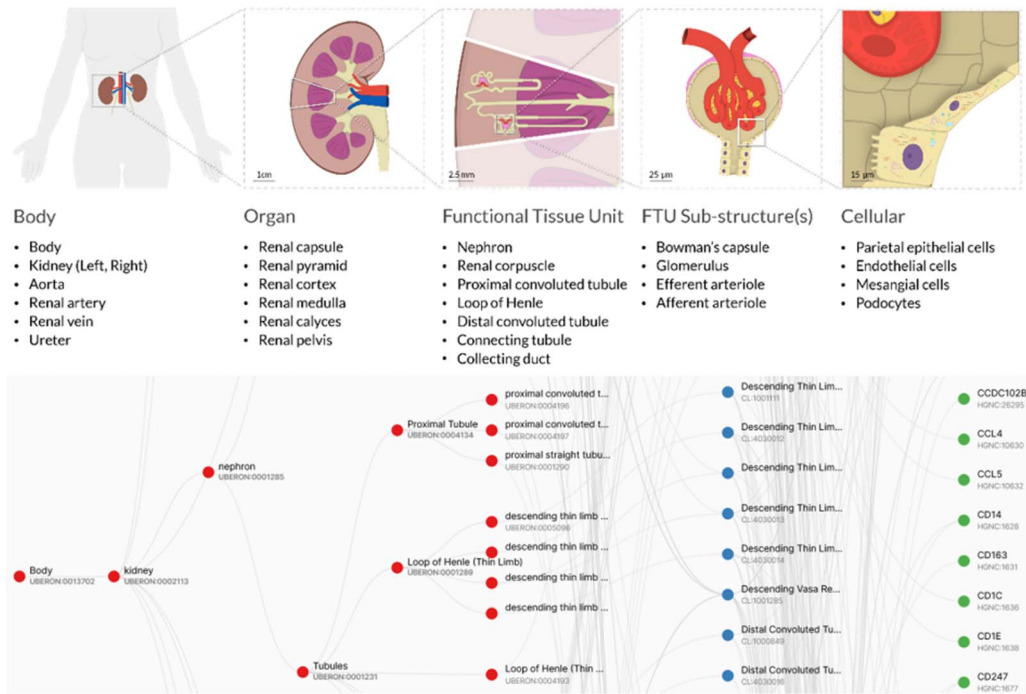


Figure 1. Multilevel Human Reference Atlas detail from body to single cell level anatomy (top) and terminology (bottom). Shown in top left is a human body with left and right kidney with zoom into kidney anatomical structures, nephron, and renal corpuscle functional tissue units, and further zoom into the single cell level. Below is a screenshot of the interactive ASCT+B Reporter visualization showing a subset of the kidney ASCT+B table with anatomical structures denoted by red nodes, cell types by blue nodes, and gene biomarkers by green nodes. Anatomical structures are linked to each other via *part_of* relationships. Cell types are linked to anatomical structures via *located_in* relationships. Biomarkers are linked to cell types via *characterizes* relationships.

also records the ORCID IDs of expert authors and reviewers as well as paper DOIs at the organ level and the cell type level. The HRA data can be explored via the ASCT+B Reporter and the Exploration User Interface (Börner, Bueckle et al., 2022). HRA data can be accessed programmatically using application programming interfaces (APIs) (Herr, Hardi et al., 2023).

Recently, high-quality, single-cell experimental data sets have been linked to the anatomical structures and cell types plus biomarkers in the ASCT+B tables. For example, cell by gene matrixes from single-cell studies now provide experimental evidence for which cells are located in which anatomical structures or which genes are highly expressed in which cell types. Azimuth references (Hao, Hao et al., 2021) make it easy to assign cell type names to clusters of cells that have similar gene expression values. Organ Mapping Antibody Panels (OMAPs) (Hickey, Neumann et al., 2022) save time and money by providing validated antibody panels for proteins commonly used to characterize cell types in different healthy human organs. There exist crosswalks from Azimuth and OMAPs to the ASCT+B tables; hence, experimental data sets that used them to identify cell types via gene or protein biomarkers can be easily compared with the HRA.

Keeping track of hundreds of experts working on the HRA, hundreds of experimental data sets, and thousands of papers that provide scholarly evidence for the HRA is nontrivial. This paper presents the first analyses and visualizations that showcase what data and scholarly evidence exist for which organs and how experts relate to the organs covered in the current and future HRA releases.

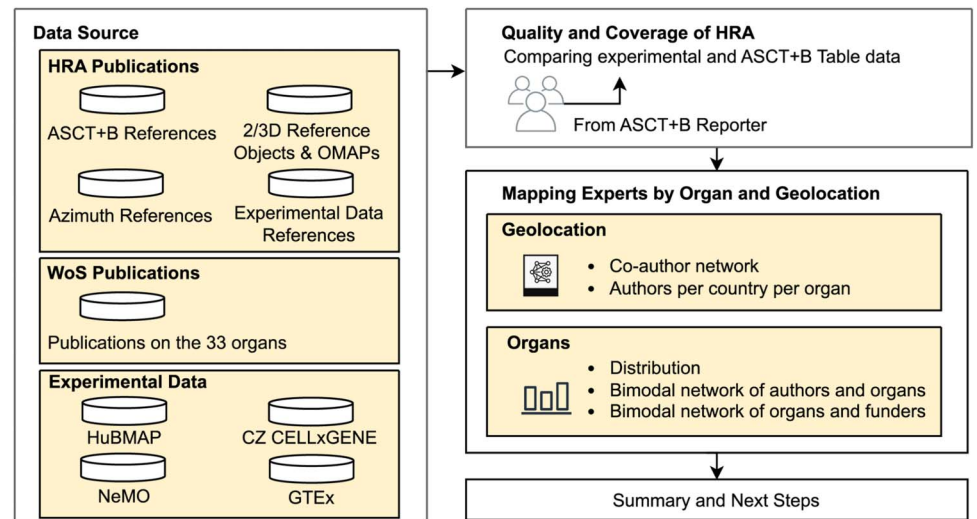


Figure 2. Overview of data compilation, HRA quality and coverage analyses, and data visualization designs.

The remainder of the paper is organized as follows: Section 2 introduces prior work. Next, we detail the data sources used in this paper and preprocessing performed on data. We then analyze and tabulate scholarly paper and experimental data evidence for the HRA. Next, we use Web of Science data to analyze and visualize experts and the organs they study, and to identify additional experts we plan to invite to review the HRA or to contribute in the future. We conclude with a summary of results and a discussion of next steps (for an overview of the data compilation, see Figure 2).

2. PRIOR WORK

Given recent advances in biomolecular experimental studies, it has become possible to study humans and other species at the single-cell level. A key goal of many studies is the development of a healthy reference atlas that can be compared to data about diverse diseases to understand associated structural and functional changes in tissues across scales. Data used for atlas design comes from many experimental studies conducted by teams around the globe. Harmonizing and interlinking this data is nontrivial. Most efforts focus on experimental data exclusively but some aim to capture links to scholarly publications and expertise. We discuss five exemplary efforts here:

- The human Ensemble Cell Atlas (hECA) effort (Chen, Luo et al., 2022) aims to build an atlas of human cells as a reference for future biological and medical studies of human health and disease. The hECA compiles data of cells across organs and studies into one data repository using a unified hierarchical annotation framework (uHAF) to harmonize data. In 2021, the hECA provided access to scRNA-seq data of more than one million human cells from diverse studies.
- CellMarker (Hu, Li et al., 2023; Zhang, Lan et al., 2019) (<https://bio-bigdata.hrbmu.edu.cn/CellMarker>) focuses on interlinking data on cell types, cell function, cell communication, etc. while keeping track of data provenance (e.g., what data were extracted from which scholarly paper).

- PanglaoDB (Franzén, Gan, & Björkegren, 2019) (<https://panglaodb.se>) interlinks cell types, genetic pathways and regulatory networks, covering preprocessed and precomputed analyses from more than 1,000 single-cell related experiments.
- SC2disease (Zhao, Lyu et al., 2021) (<https://easybioai.com/sc2disease>) is a comprehensive resource for differentially expressed gene profiles, which supports comparisons among cell types, tissues and disease-related health states. It contains thousands of entries on different cells, tissues, and diseases.
- SPEED database (Chen, Zhang et al., 2023) (<https://speedatlas.net>) is a single-cell pan-species atlas that covers more than 5 million cells across 127 species, aiming to advance our collective understanding of the heterogeneities among cells, tissues, organs, and species.

However, to the best of our knowledge, none of these five efforts has systematically studied or visualized the network of how experimental data and scholarly papers provide evidence for reference atlas construction.

Several teams within HuBMAP (HuBMAP Consortium, 2019) are working on the development of general methods and tools (Börner et al., 2021; Manz, Gold et al., 2022; Zhang, Srivastava et al., 2022), demonstration projects (Burnum-Johnson, Conrads et al., 2022), organ-specific atlases (Becker, Nevins et al., 2022; Kruse & Spraggins, 2022), and novel technologies that can be used to map human tissue at the single cell level (Deng, Bartosovic et al., 2022; Melani, Gerbasi et al., 2022; Schachner, Soye et al., 2022; Stockwell, 2022). This paper is unique in that it shows for the very first time how elements of the HRA are linked to scholarly papers and experimental data to understand and communicate atlas quality, to guide future tissue data collection, and to identify other leading experts that might be interested in serving as authors or reviewers of the evolving atlas.

3. DATA AND DATA PROCESSING

This section details all of the data used in this study: nearly 500 papers cited by the 33 reference organs covered in the Human Reference Atlas and Azimuth; approximately 250,000 papers on the 33 organs retrieved from the Web of Science; and roughly 300 experimental data sets that are associated with about 80 scholarly papers. Data details and code are available at <https://github.com/cns-iu/hra-evidence-issi-2023-supporting-information>.

3.1. Publication Evidence from HRA

The Human Reference Atlas captures data on ASCT+B tables (see Section 1 and Figure 1), associated two-dimensional (2D) and three-dimensional (3D) spatial representations of major anatomical structures and the cell types commonly located in these, and the biomarkers (gene, protein, lipid, metabolites) used to *characterize* cell types. In addition, Azimuth (Hao et al., 2021), a tool for reference-based single-cell analysis, publishes gene biomarkers used to characterize cell types in different organs; see existing organ references at <https://azimuth.hubmapconsortium.org>. This section details how paper evidence was retrieved from different websites and processed to get summary statistics.

3.1.1. ASCT+B references

The CCF ASCT+B Reporter (<https://hubmapconsortium.github.io/ccf-asct-reporter>) lets users explore ASCT+B table visualizations and download table reports of key statistics (e.g., number of cell types per organ). Table 1 shows the unique number of publication references listed in

Table 1. Count of references listed in the 26 ASCT+B tables at the organ and cell type level

Organ name and version	Organ level	Cell type level	Organ name and version	Organ level	Cell type level
Blood v1.2	4	9	Ovary v1.1	4	13
Blood Vasculature v1.2	2	18	Pancreas v1.1	1	0
Bone Marrow v1.2	4	12	Peripheral Nervous System v1.0	1	0
Brain v1.2	1	1	Placenta Full Term v1.0	0	48
Eye v1.1	9	30	Prostate v1.0	1	1
Fallopian Tube v1.1	3	2	Single Lobe Lung v1.2	8	42
Heart v1.1	1	0	Skin v1.2	1	85
Kidney v1.2	12	16	Small Intestine v1.0	1	1
Knee	0	0	Spleen v1.2	4	54
Large Intestine v1.2	3	15	Thymus v1.2	1	36
Liver v1.1	3	25	Ureter v1.0	1	0
Lymph Node v1.2	4	42	Urinary Bladder v1.0	1	0
Lymph Vasculature v1.1	1	0	Uterus v1.1	3	3
Total for all 26 organs*				70	425

* The counts shown are the number of unique references at organ level and cell type level.

the 26 ASCT+B Tables from the third HRA release v1.2 (<https://hubmapconsortium.github.io/ccf-releases/v1.2/docs>). Note that references are cited at the entire organ level but also for specific cell types in the organ. The number of all unique references cited in the 26 tables is 456, including 12 unique books, 439 unique papers (305 of them in WoS core collection) and five papers from PubMed other sources.

3.1.2. 2/3D reference objects and OMAPs

In the 3rd HRA release, there are 19 2D reference objects for functional tissue units in seven organs with 90 unique cell types; 53 3D reference organs with 1,542 named anatomical structures, and seven Organ Mapping Antibody Panels (OMAPs) for 187 anatomical structures, 179 cell types, and 197 protein biomarkers across the seven organs. Papers for the 2D and 3D References Library Objects and OMAPs were downloaded from the HuBMAP CCF Portal (<https://hubmapconsortium.github.io/ccf>). Table 2 shows the number of unique papers per organ per HRA object type and the total number of unique papers across all organs. There are a total of 16 publication references, comprising two unique books and 14 unique scientific papers with unique DOIs. None of these 16 scholarly works for 2/3D reference objects and OMAPs are cited in ASCT+B tables, likely due to the fact that ASCT+B table references focus on the cell type level.

3.1.3. Azimuth references

Azimuth references support cell type annotation for tissue data sets (Hao et al., 2021). They exist for 10 organs, and references to associated publications can be downloaded from <https://>

Table 2. Count of papers for 2/3D reference objects and OMAPs per organ

Organ	2D objects	3D objects	OMAPs	Total no. papers	No. papers with DOIs
Brain	NA	1**	NA	1	1
Kidney	3*			3	1
Large Intestine	2			2	1
Liver	3			3	1
Lung	3			3	1
Lymph Node	NA		2	2	2
Pancreas	3		2	5	3
Placenta	NA	2		2	1
Prostate	3			3	1
Spinal Cord	NA	1	NA	1	1
Thymus	3			3	1
Total across organs	9	4	4	16	14

* The kidney organ has two 2D objects (“Kidney, 2D Nephron FTU v.1.0” and “Kidney, 2D Renal Corpuscle FTU v.1.0”) with one book (ISBN 978-3-662-02676-2) each, and another two 2D objects (“Kidney, 2D Nephron FTU v.1.0” and “Kidney, 2D Renal Corpuscle FTU v.1.0”) with one paper (ISBN 978-3-642-08106-4) each.

** The brain organ has four 3D objects with one paper each.

azimuth.hubmapconsortium.org and are listed in Table 3. HuBMAP focuses on adults (excluding fetal development) with no ASCT+B tables for adipose and tonsils. Thirty-eight unique papers are associated with the 10 Azimuth references and two of the papers are preprints. Thirty-six unique papers have DOIs and eight of these are also cited in ASCT+B tables.

Papers listed in Azimuth single-cell references for which ASCT+B tables exist have been shared with table lead authors for possible inclusion in the ASCT+B tables.

3.1.4. Experimental data references

A total of 308 data sets from single-cell studies of healthy human adults were retrieved from HuBMAP Portal (Cao, Spielmann et al., 2019; Stuart, Butler et al., 2019), CZ CellxGene Portal (Domínguez Conde, Xu et al., 2022; Tabula Sapiens Consortium, 2022), NeMO (Orvis, Gottfried et al., 2021), and GTEEx (Eraslan, Drokhylyansky et al., 2022) in October 2022. These high-quality data sets cover 57 organs and the data sets are associated with 83 unique papers. Exactly 78 of these papers have DOIs and 67 of these DOIs are not cited in the existing 26 ASCT+B tables. A table showing the count of experimental data references per organ can be found on GitHub at <https://github.com/cns-iu/hra-evidence-issi-2023-supporting-information>. Note that there are 38 organs for which no ASCT+B table exists yet.

Papers associated with high-quality experimental data sets for organs that have ASCT+B tables were shared with table lead authors for possible inclusion in the ASCT+B tables.

3.1.5. Summary

In sum, there are 12 unique books, 439 unique papers (including 305 WoS core papers) and five papers from PubMed other sources listed in the 26 ASCT+B tables from the third HRA

Table 3. Count of papers per Azimuth cell by gene references per organ

Organ	No. papers	No. papers with DOIs	DOIs cited in ASCT+B tables
PBMC*	1	1	–
Adipose	1	1	NA****
Bone Marrow	3	3	–
Motor Cortex**	1	1	1
Fetal Development	4	4	NA****
Heart	4	4	–
Kidney	3	2	1
Lung	16	16	6
Pancreas	7	7	–
Tonsil	2	1	NA****
Total across organs	42***	40	8

* Peripheral blood mononuclear cells (PBMC) correspond to the Blood ASCT+B table.

** Motor cortex corresponds to the Brain ASCT+B table.

*** The count shown is the number of unique papers per organ; the total number of unique papers is 38. Pancreas and fetal development have the same reference (10.1016/j.cell.2017.09.004); PBMC and bone marrow have the same reference (10.1016/j.cell.2021.04.048); heart, kidney, and lung have the same reference (10.1038/s41586-021-03570-8).

**** No ASCT+B table exists for these three organs.

release; 16 papers (14 of them with DOIs) cited in the 2D, 3D reference objects, and OMAPs references; 26 unique papers associated with the 10 Azimuth single-cell annotation references; and in the set of 380 unique data sets, 195 have 49 unique papers associated.

3.2. WoS Papers for 33 Organs

To better understand which major papers were recently published on the 33 organs planned for the next HRA release (see the listing of all 33 organ names at <https://github.com/cns-iu/hra-evidence-issi-2023-supporting-information>), we ran a query over the Web of Science core collection provided via the Collaborative Archive & Data Research Environment (CADRE) (Mabry, Yan et al., 2020; Wittenberg, Mabry et al., 2020). The retrieval result comprises 250,620 papers that were published from 2018 to 2022 and have these organ words in titles or keywords and were cited at least 10 times. These papers cover all of the 33 organs except Bone Marrow-Pelvis.

The papers were tagged with HRA-specific organ tags based on the 33 organ names occurring in title or keywords. Next, we used the Web of Science (WoS) standard format to retrieve clean author names and affiliations. For these 250,620 papers, 672,892 unique authors and their 114,965 affiliations in 189 countries were identified, including 88 authors with more than 100 citations. To explore how this work is being funded, we identified all 177,987 papers from 89,924 organizations that are funded by 196,739 unique grants. The top five funding agencies listed most frequently in papers are the National Institutes of Health (NIH, listed on 106,311 papers), United States Department of Health & Human Services (HHS, 90,706 papers), National Natural Science Foundation of China (NSFC, 58,683 papers), National Cancer

Institute (NCI, 22,324 papers), and European Commission (EC, 18,137 papers). The top 50 funding agencies are from 12 countries and the European Union, including 18 funding agencies from the United States.

3.3. Experimental Data Evidence

To connect the HRA to experimental data we queried four major data portals that provide access to single cell data: HuBMAP Portal (<https://portal.hubmapconsortium.org>), CZ CellxGene Portal (CxG, <https://cellxgene.cziscience.com>), NeMO (<https://nemoarchive.org>), and GTEx (<https://gtexportal.org>). In addition, we retrieved all data used to compile the Azimuth references (<https://azimuth.hubmapconsortium.org>) as well as data for three key atlas papers: Tabula Sapiens Consortium (2022), Lake, Menon et al. (2023), and Ghose, Ju et al. (2023). The number of data sets, organs, and papers per data source is shown in Table 4.

The eight data sources cover 57 unique organs, 308 unique data sets, and 82 unique references (77 of them with DOIs). The top five organs with most data sets are blood, brain (motor cortex), kidney, lung, and skin with 20, 28 (14), 153, 21, and 14 data sets respectively. Kidney data sets come from four data sources: Azimuth reference, Lake et al. (2023), CxG, and HuBMAP. The top five organs with the most unique references are blood (18 papers), heart (seven papers), kidney (seven papers), lung (26 papers), and pancreas (12 papers).

A closer look at the 82 unique papers reveals that 77 of them have a DOI and 10 of these 77 are cited in the ASCT+B tables. A table of organ-specific papers that ASCT+B lead authors should review and consider for inclusion was compiled and published on GitHub at <https://cns-iu.github.io/hra-evidence-issi-2023-supporting-information>. This table was shared with table lead authors for possible inclusion in the ASCT+B tables.

4. QUALITY AND COVERAGE OF THE HRA

A comparison of experimental data to anatomical structures and cell types plus biomarkers covered in the ASCT+B tables helps individuals understand and communicate the coverage of the existing HRA and plan future tissue data collection (e.g., to collect a minimum amount of experimental data for major anatomical structures and cell types). The ASCT+B Reporter

Table 4. Number of data sets, organs, and reference for the eight data sources

Data Source	Source type	No. organs	No. data sets	No. papers
Azimuth	Reference	10	27	26
Lake et al.	Paper	1	21	1
The Tabula Sapiens Consortium	Paper	33	5	1
CxG	Portal	49	196	49
GTEx	Portal	8	25	NA
HuBMAP	Portal	1*	79	1*
NeMO	Portal	1	14	NA
Ghose et al.	Paper	1	10	1
Totals (Unique)		57	308	83

* Azimuth runs in HuBMAP portal in production mode for kidney. The paper listed is the Lake et al. (2023) paper; other data sets are unpublished.

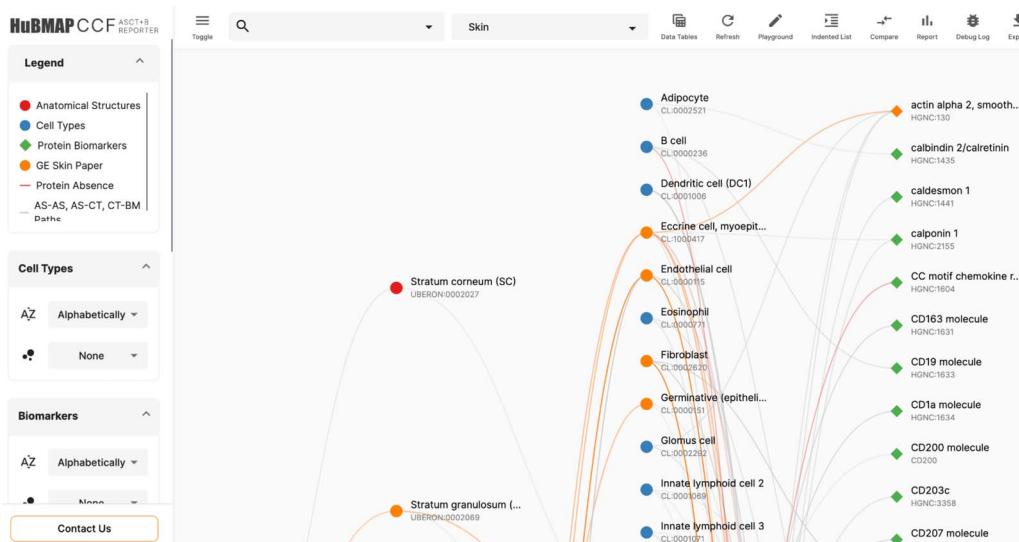


Figure 3. ASCT+B Reporter comparison visualization of an ASCT+B table and experimental skin data. Network of anatomical structures (red nodes), cell types (blue nodes), and protein biomarkers (green nodes) for the skin ASCT+B table is used as a basemap. Experimental data is overlaid in orange, making it easy to explore and communicate (non)matching anatomical structures, and cell types plus biomarkers covered in a study.

(<https://hubmapconsortium.github.io/ccf-asct-reporter>) was used to visualize the network of anatomical structures and cell types plus biomarkers in an ASCT+B Master table as a basemap and to overlay experimental data so that coverage can be compared and communicated. See workflow detailed in <https://hubmapconsortium.github.io/hra-previews/pilots/pilot1.html>.

As an example, we compare the 12 anatomical structures, 12 cell types located in these anatomical structures, and 18 protein biomarkers used to characterize these cell types from the skin data set published in Ghose et al. (2023). A partial screenshot of the interactive visualizations is shown in Figure 3. The interactive visualization can be explored at <https://hubmapconsortium.github.io/ccf-asct-reporter/vis?selectedOrgans=skin-v1.2&playground=false&comparisonCSVURL=https:%2F%2Fdocs.google.com%2Fspreadsheets%2Fd%2F1ebxX1VmZXRjfxZC8DdxtPjTGQLId9NBja71ii939c8%2Fedit%23gid%3D1990254927&comparisonName=Human%20Digital%20Twin&comparisonColor=%23ff8000&comparisonHasFile=false>.

The visualization shows what data and publication evidence (here 10 data sets published in one paper) exist for which anatomical structures, cell types, and protein biomarkers. The ASCT+B Reporter makes it possible to overlay data from multiple studies using different colors. Insights gained are valuable for planning future tissue data collection (e.g., to collect a minimum amount of experimental data that maximally improves HRA coverage and quality).

5. MAPPING EXPERTS BY ORGAN AND GEOLOCATION

The 26 ASCT+B tables list 88 directly involved experts who serve as authors or internal and external reviewers. For each expert, there exists an ORCID ID in the ASCT+B tables—52 unique authors, four unique project leaders, 47 unique reviewers. Some experts serve in multiple roles across organs. As for the 2D reference objects, there are 14 unique experts listed; for 3D reference objects, there are 32 experts, and for OMAPs there are 29 experts. Across the HRA, there are 116 unique experts, 113 of them with ORCID IDs.



Figure 4. Geospatial layout of the coauthor network. World map showing 672,892 experts in 189 countries and coauthor relationships from the United States to 170 countries. Node size and color represent the number of countries where coauthors' affiliations are located, including US affiliations and their collaborating affiliations from 170 countries. Edge weight and color refer to the number of times that authors from two countries appear on a paper together.

Using the WoS papers data comprising 250,620 papers that featured any of the 33 organ names in their title or keywords, we identified 672,892 indirectly involved expert authors. The authors have 114,965 unique affiliations in 189 unique countries. A map of the world with a country-level overlay of authors and their coauthor relationships is shown in Figure 4. The original network was almost fully connected and hence MST-Pathfinder Network (PFnet) (Sci2 Team, 2009) was applied to remove less important edges. In the resulting network, the United States has 84,287 papers and is the most highly connected node with these top five collaborators: China (CN; 9,541 papers), United Kingdom (UK; 6,338 papers), Canada (CA; 5,605 papers), Germany (DE; 5,131 papers), and Italy (IT; 4,271 papers).

To ascertain what organ expertise the paper authors bring to the table, we computed the distribution of the number of organs per expert and the number of papers per organ; see Figure 5.

At the author level, Figure 6 shows the bimodal network of highly cited experts (equal or more than 100 citations) and the organs they study. As can be seen, highly cited experts study liver (66 experts) and lung (40 experts). The paper with the most authors is entitled "Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer" and has 171 authors and 3,612 citations. In terms of geographical distribution of authors, Figure 7 presents the number of authors per country per organ for country-organ combinations with more than 1,000 authors. The number of authors from China and the United States is notably high, with over 10,000 experts specializing in liver, brain, and lung studies. Specifically, China has 15,526 liver experts and 11,119 lung experts, and the United States has 11,391 brain experts and 11,086 liver experts.

To understand the funding landscape for research efforts on the 33 organs, we constructed a bimodal network of organs and their respective funding agencies. Given the network's density,

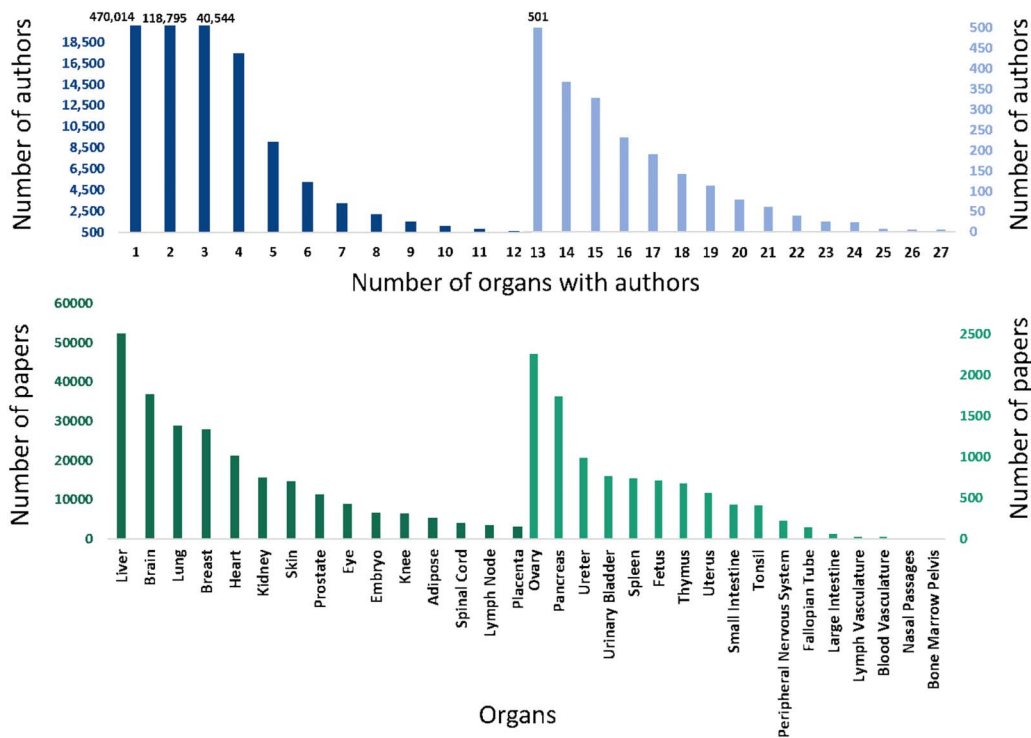


Figure 5. Number of organs experts work on (top) and number of papers per organ (bottom). Shown in the upper left is the range from 666 and 470,014 (dark blue bars) and the upper right shows the range between five and 501 (light blue bars). Many authors (470,014) work on one organ, five authors work on 27 organs. The lower left bars show the range between 3,102 and 52,304 (dark green) and in the lower right the range between one and 2,258 (light green). Many papers exist for the liver, followed by papers on the brain, lung, breast, heart, kidney, and skin.

we employed PFnet to emphasize the most significant connections, which are displayed in Figure 8. Note that funding agencies, such as the National Institutes of Health (NIH), United States Department of Health & Human Services (HHS), European Commission, and UK Research & Innovation (UKRI) support research on 32 of the organs. We excluded “bone

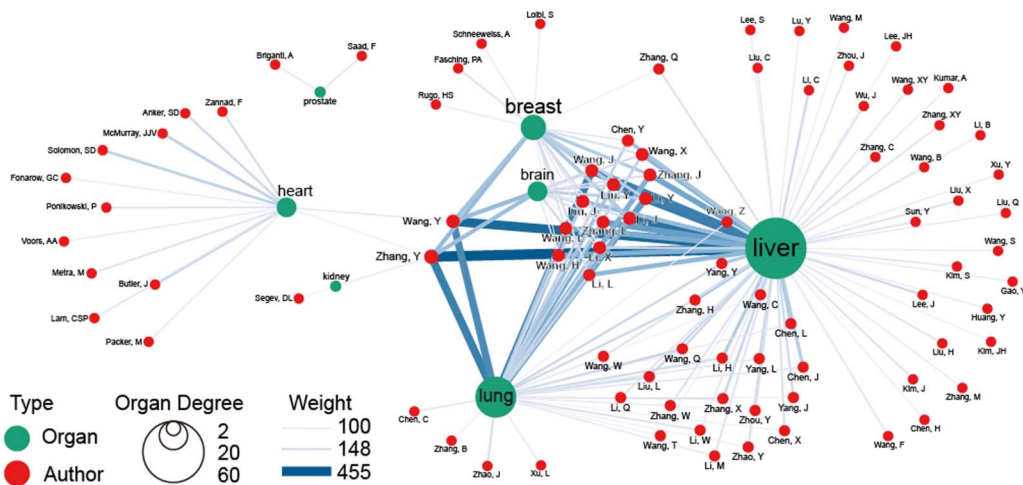


Figure 6. Bimodal network of highly cited authors and the organs they study. The seven organ nodes (green) and 88 author nodes (pink) are sized by the number of related authors and the number of related organs, respectively. Edge weights represent the number of papers published by an expert for one organ.

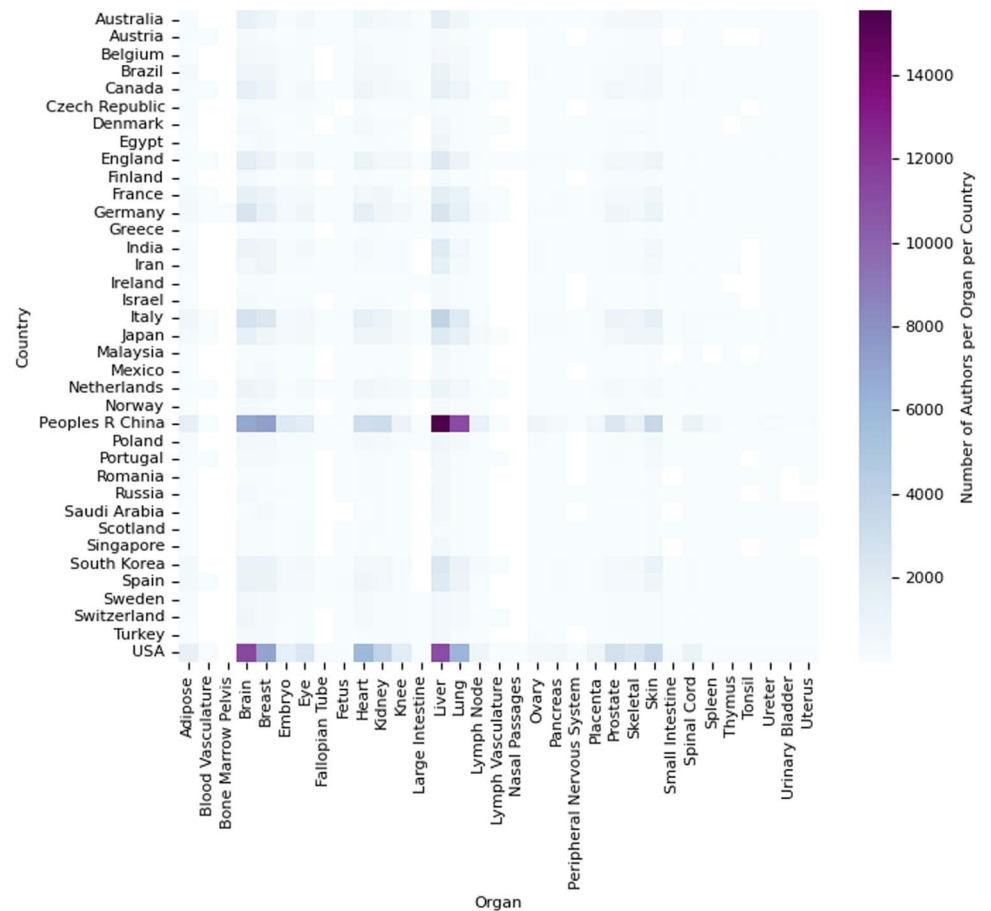


Figure 7. Heatmap of authors per country per organ. There are 37 countries that have contributed over 1,000 authors to the study of 33 organs and these are shown here. Dark purple indicates a high number of publications. There are 22,483 papers by 15,526 authors from the People's Republic of China for the liver organ.

marrow-pelvis" from our analysis due to the absence of relevant papers. Most papers on brain topics are funded by NIH, which is acknowledged in 24,352 papers.

6. SUMMARY AND NEXT STEPS

This paper presents initial analyses and visualizations of scholarly papers and experimental data set evidence for the Human Reference Atlas. We analyzed the number and type of scholarly evidence for subgraphs of the HRA and show that 96.15% of the 26 ASCT+B tables, all of the 2D reference objects, 12% of the 3D reference objects and 28.57% of the OMAPs and all Azimuth references have scholarly publications associated; all 26 organs have experimental cell type by biomarker data evidence, but coverage varies: see the example coverage for skin in Figure 3. We have been and will continue to share coverage results with the larger HRA community to highlight organ teams that have managed to provide extensive publication and experimental data evidence and to inspire other teams that have recently joined the HRA effort to do the same.

We analyzed the network of experts currently collaborating on the HRA and used WoS data to identify and visualize experts that work on the 33 existing and planned organs. The geospatial and bimodal networks showcase the number of experts and funders and their countries

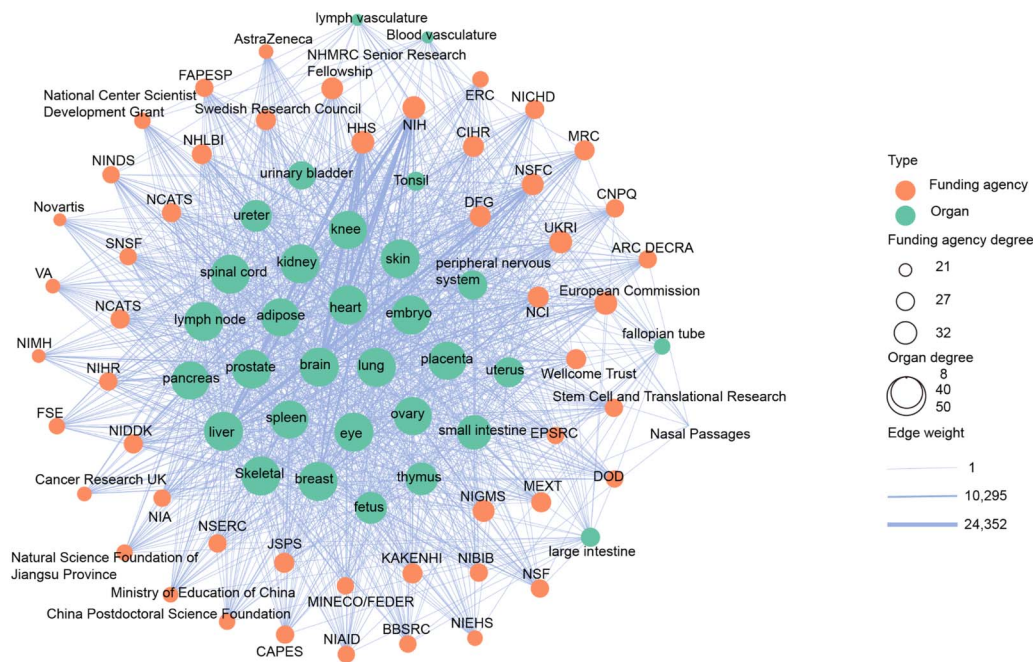


Figure 8. Bimodal network of 32 organs and top 50 funding agencies most often listed. Shown are 50 funding agencies (orange nodes) and 32 organs (green nodes; “bone marrow-pelvis” was omitted as there is no funding). Funding agencies are area sized by the number of organs they support (i.e., node degree). Organs are sized by the number of funding agencies that are acknowledged in papers that mention them. Edges interlink organs and funders and their weight indicates the number of papers that list the organ and acknowledge funding.

and we will use the results to invite other leading experts to serve as authors or reviewers of the evolving atlas. Connecting experts across projects and time zones will make it possible to benefit from international expertise, technologies, and data sets in support of highest quality HRA construction and usage of the HRA data in future scholarly publications.

Over the coming 5 years, we expect the number of active authors to grow from 200 to 1,000. The current set of organs will expand, and we expect the final HRA will cover ca. 5,000 cell types and 10,000 unique anatomical regions. Managing the systematic authoring, review, and validation of the Human Reference Atlas is nontrivial. Visualizations that show the coverage and quality of the evolving atlas, relevant expertise around the globe, and high-quality experimental data sets will be critically important for communicating progress to experts and funders engaged in constructing or using the atlas.

ACKNOWLEDGMENTS

We would like to thank Devin M. Wright and Ellen M. Quardokus for their support in the identification of HRA-relevant papers and experimental data sets; Abhay B. Rajde for sharing ASCT+B table data and expanding ASCT+B Reporter functionality with Bruce W. Herr II providing technical advice; and Leonard Cross for designing Figure 3. Nancy L. Ruschman served as a friendly reviewer and user of the expertise mapping and recommendation study.

AUTHOR CONTRIBUTIONS

Yongxin Kong: Data curation, Formal analysis, Visualization, Writing—original draft, Writing—review & editing. Vicky Amar Daiya: Data curation, Writing—original draft. Katy Börner: Conceptualization, Methodology, Writing—original draft, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

Mike Gallant and Jacob J. Shaw provided access to Web of Science data using the Collaborative Archive & Data Research Environment (CADRE) project (<https://doi.org/10.26313/rdy8-4w58>) developed with support from a National Leadership Grant from the Institute of Museum and Library Services (IMLS; grant number LG-70-18-0202-18), including cost-share from the Big Ten Academic Alliance Library Initiatives (BTAA), Microsoft Research, the Web of Science Group, and academic university libraries (Ohio State University, Michigan State University, Purdue University, University of Michigan, University of Minnesota, Penn State University, University of Iowa and Rutgers University).

Kong is funded by the China Scholarship Council. The Human Reference Atlas research is funded by the NIH Common Fund through the Office of Strategic Coordination/Office of the NIH Director under awards OT2OD033756 and OT2OD026671, by the Cellular Senescence Network (SenNet) Consortium through the Consortium Organization and Data Coordinating Center (CODCC) under award number U24CA268108, and by the NIDDK Kidney Precision Medicine Project grant U2CDK114886.

DATA AVAILABILITY

Data details and code are available at <https://github.com/cns-iu/hra-evidence-issi-2023-supporting-information>.

REFERENCES

- Becker, W. R., Nevins, S. A., Chen, D. C., Chiu, R., Horning, A. M., ... Greenleaf, W. J. (2022). Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. *Nature Genetics*, *54*(7), 985–995. <https://doi.org/10.1038/s41588-022-01088-x>, PubMed: 35726067
- Börner, K., Bueckle, A., Herr, B. W., Cross, L. E., Quardokus, E. M., ... Weber, G. M. (2022). Tissue registration and exploration user interfaces in support of a human reference atlas. *Communications Biology*, *5*, 1369. <https://doi.org/10.1038/s42003-022-03644-x>, PubMed: 36513738
- Börner, K., Teichmann, S. A., Quardokus, E. M., Gee, J. C., Browne, K., ... Weber, G. (2021). Anatomical structures, cell types and biomarkers of the Human Reference Atlas. *Nature Cell Biology*, *23*, 1117–1128. <https://doi.org/10.1038/s41556-021-00788-6>, PubMed: 34750582
- Burnum-Johnson, K. E., Conrads, T. P., Drake, R. R., Herr, A. E., Iyengar, R., ... Kelleher, N. L. (2022). New views of old proteins: Clarifying the enigmatic proteome. *Molecular & Cellular Proteomics*, *21*(7), 100254. <https://doi.org/10.1016/j.mcpro.2022.100254>, PubMed: 35654359
- Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., ... Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, *566*, 496–502. <https://doi.org/10.1038/s41586-019-0969-x>, PubMed: 30787437
- Chen, S., Luo, Y., Gao, H., Li, F., Chen, Y., ... Zhang, X. (2022). hECA: The cell-centric assembly of a cell atlas. *iScience*, *25*(5), 104318. <https://doi.org/10.1016/j.isci.2022.104318>, PubMed: 35602947
- Chen, Y., Zhang, X., Peng, X., Jin, Y., Ding, P., ... Chen, D. (2023). SPEED: Single-cell Pan-species atlas in the light of Ecology and Evolution for Development and Diseases. *Nucleic Acids Research*, *51*(D1), D1150–D1159. <https://doi.org/10.1093/nar/gkac930>, PubMed: 36305818
- Deng, Y., Bartosovic, M., Ma, S., Zhang, D., Kukanja, P., ... Fan, R. (2022). Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature*, *609*, 375–383. <https://doi.org/10.1038/s41586-022-05094-1>, PubMed: 35978191
- Domínguez Conde, C., Xu, C., Jarvis, L. B., Rainbow, D. B., Wells, S. B., ... Teichmann, S. A. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, *376*, eabl5197. <https://doi.org/10.1126/science.abl5197>, PubMed: 35549406
- Eraslan, G., Drokhyansky, E., Anand, S., Fiskin, E., Subramanian, A., ... Regev, A. (2022). Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science*, *376*, eabl4290. <https://doi.org/10.1126/science.abl4290>, PubMed: 35549429
- Franzén, O., Gan, L.-M., & Björkegren, J. L. M. (2019). PanglaoDB: A web server for exploration of mouse and human single-cell RNA sequencing data. *Database*, *2019*, baz046. <https://doi.org/10.1093/database/baz046>, PubMed: 30951143
- Ghose, S., Ju, Y., McDonough, E., Ho, J., Karunamurthy, A., ... Ginty, F. (2023). 3D reconstruction of skin and spatial mapping of immune cell density, vascular distance and effects of sun exposure and aging. *Communications Biology*, *6*(1), 718. <https://doi.org/10.1038/s42003-023-04991-z>, PubMed: 37468758

- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., ... Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573–3587. <https://doi.org/10.1016/j.cell.2021.04.048>, PubMed: 34062119
- Herr, B. W., Hardi, J., Quardokus, E. M., Bueckle, A., Chen, L., ... Börner, K. (2023). Specimen, biological structure, and spatial ontologies in support of a Human Reference Atlas. *Scientific Data*, *10*, 171. <https://doi.org/10.1038/s41597-023-01993-8>, PubMed: 36973309
- Hickey, J. W., Neumann, E. K., Radtke, A. J., Camarillo, J. M., Beuschel, R. T., ... Saka, S. K. (2022). Spatial mapping of protein composition and tissue organization: A primer for multiplexed antibody-based imaging. *Nature Methods*, *19*, 284–295. <https://doi.org/10.1038/s41592-021-01316-y>, PubMed: 34811556
- Hu, C., Li, T., Xu, Y., Zhang, X., Li, F., ... Zhang, Y. (2023). CellMarker 2.0: An updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Research*, *51*(D1), D870–D876. <https://doi.org/10.1093/nar/gkac947>, PubMed: 36300619
- HuBMAP Consortium. (2019). The human body at cellular resolution: The NIH Human Biomolecular Atlas Program. *Nature*, *574*, 187–192. <https://doi.org/10.1038/s41586-019-1629-x>, PubMed: 31597973
- Kruse, A. R. S., & Spraggins, J. M. (2022). Uncovering molecular heterogeneity in the kidney with spatially targeted mass spectrometry. *Frontiers in Physiology*, *13*, 837773. <https://doi.org/10.3389/fphys.2022.837773>, PubMed: 35222094
- Lake, B. B., Menon, R., Winfree, S., Hu, Q., Ferreira, R. M., ... Jain, S. (2023). An atlas of healthy and injured cell states and niches in the human kidney. *Nature*, *619*, 585–594. <https://doi.org/10.1038/s41586-023-05769-3>, PubMed: 37468583
- Mabry, P. L., Yan, X., Pentchev, V., Van Rennes, R., McGavin, S. H., & Wittenberg, J. V. (2020). CADRE: A collaborative, cloud-based solution for big bibliographic data research in academic libraries. *Frontiers in Big Data*, *3*, 556282. <https://doi.org/10.3389/fdata.2020.556282>, PubMed: 33693415
- Manz, T., Gold, I., Patterson, N. H., McCallum, C., Keller, M. S., ... Gehlenborg, N. (2022). Viv: Multiscale visualization of high-resolution multiplexed bioimaging data on the web. *Nature Methods*, *19*, 515–516. <https://doi.org/10.1038/s41592-022-01482-7>, PubMed: 35545714
- Melani, R. D., Gerbasi, V. R., Anderson, L. C., Sikora, J. W., Toby, T. K., ... Kelleher, N. L. (2022). The Blood Proteoform Atlas: A reference map of proteoforms in human hematopoietic cells. *Science*, *375*(6579), 411–418. <https://doi.org/10.1126/science.aaz5284>, PubMed: 35084980
- Orvis, J., Gottfried, B., Kancherla, J., Adkins, R. S., Song, Y., ... Hertzano, R. (2021). gEAR: Gene Expression Analysis Resource portal for community-driven, multi-omic data exploration. *Nature Methods*, *18*, 843–844. <https://doi.org/10.1038/s41592-021-01200-9>, PubMed: 34172972
- Schachner, L. F., Soye, B. D., Ro, S., Kenney, G. E., Ives, A. N., ... Kelleher, N. L. (2022). Revving an engine of human metabolism: Activity enhancement of triosephosphate isomerase via hemi-phosphorylation. *ACS Chemical Biology*, *17*(10), 2769–2780. <https://doi.org/10.1021/acscchembio.2c00324>, PubMed: 35951581
- Sci2 Team. (2009). *Science of Science (Sci2) Tool*. Indiana University and SciTech Strategies. Retrieved from <https://sci2.cns.iu.edu>.
- Stockwell, B. R. (2022). Ferroptosis turns 10: Emerging mechanisms, physiological functions, and therapeutic applications. *Cell*, *185*(14), 2401–2421. <https://doi.org/10.1016/j.cell.2022.06.003>, PubMed: 35803244
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., ... Satija, R. (2019). Comprehensive integration of single-cell data. *Cell*, *177*(7), 1888–1902. <https://doi.org/10.1016/j.cell.2019.05.031>, PubMed: 31178118
- Tabula Sapiens Consortium. (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science*, *376*(6594), eabl4896. <https://doi.org/10.1126/science.abl4896>, PubMed: 35549404
- Wittenberg, J., Mabry, P. L., Yan, X., Van Rennes, R., Pentchev, V., & Wilkinson, J. (2020). *Collaborative Archive & Data Research Environment (CADRE)*. IUScholarWorks. Retrieved November 13, 2022, from <https://hdl.handle.net/2022/25309>. <https://doi.org/10.26313/rdy8-4w58>
- Zhang, B., Srivastava, A., Mimitou, E., Stuart, T., Raimondi, I., ... Satija, R. (2022). Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro. *Nature Biotechnology*, *40*, 1220–1230. <https://doi.org/10.1038/s41587-022-01250-0>, PubMed: 35332340
- Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., ... Xiao, Y. (2019). CellMarker: A manually curated resource of cell markers in human and mouse. *Nucleic Acids Research*, *47*(D1), D721–D728. <https://doi.org/10.1093/nar/gky900>, PubMed: 30289549
- Zhao, T., Lyu, S., Lu, G., Juan, L., Zeng, X., ... Peng, J. (2021). SC2disease: A manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Research*, *49*(D1), D1413–D1419. <https://doi.org/10.1093/nar/gkaa838>, PubMed: 33010177