Scholarly publications and dataset evidence for the human reference atlas

Yongxin Kong[1,*], Vicky Amar Daiya[2], Katy Börner[3,*]

[1] *yokong@iu.edu*
Department of Intelligent Systems Engineering, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, (USA)
School of Information Management, Sun Yat-sen University, Guangzhou 510006, (China)

[2] *vdaiya@iu.edu*
Department of Computer Science, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, (USA)

[3] *katy@indiana.edu*
Department of Intelligent Systems Engineering, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408, (USA)

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 release features 26 organ-specific ASCT+B tables that cite 456 scholarly papers, are linked to 61 spatial reference objects and 7 Organ Mapping Antibody Panels (OMAPs), and are 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 that work on the 33 organs targeted for the next HRA release. To provide scientific evidence for the HRA, we identified 620 high quality, single-cell experimental datasets for 58 organs published in 561 unique papers. The presented results are critical for understanding and communicating the quality of HRA, planning future tissue data collection, and inviting leading experts to contribute to the evolving atlas.

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. Datasets 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 multi-level atlas. Fig. 1 illustrates detail 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), 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 also records 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 *et al.*, 2023).

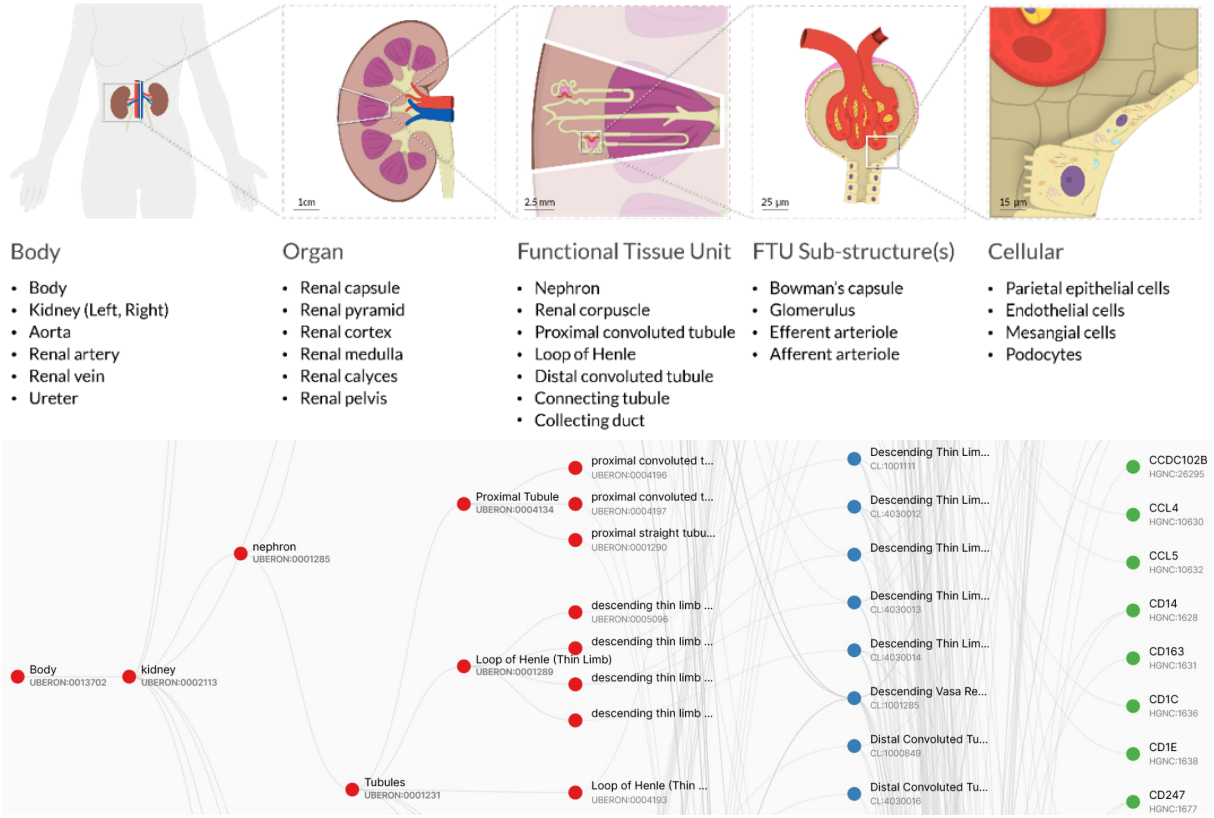

Figure 1. Multi-level Human Reference Atlas detail from body to single cell level anatomy (top) and terminology (bottom). Shown in top left is a human body with left and right kidney with zoom into kidney anatomical structures, nephron and renal corpuscle functional tissue units, and single cell level. 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.

Recently, high quality, single-cell experimental datasets have been linked to the anatomical structures, cell types plus biomarkers in the ASCT+B tables. For example, cell by gene matrices from single-cell studies now provide experimental evidence for what cells are located in which anatomical structures or what genes are highly expressed in what cell types. Azimuth references (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 *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 datasets that used them to identify cell types via gene or protein biomarkers can be easily compared with the HRA.

Keeping track of 100s of experts working on the HRA, 100s of experimental datasets, and 1000s of papers that provide scholarly evidence for the HRA is non-trivial. 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.

The remainder of the paper is organized as follows: The subsequent section 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 to analyze and visualize experts and the organs they study to identify additional experts that we plan to invite to review the HRA or contribute to it in the future. We conclude with a summary of results and a discussion of next steps.

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 for 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 non-trivial. Most efforts focus on experimental data exclusively while some aim to capture links to scholarly publications and expertise. We discuss five exemplary efforts here. (1) 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 HRA 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 HRA provides access to scRNA-seq data of more than 1 million human cells from diverse studies. (2) CellMarker (Zhang *et al.*, 2019; Hu *et al.*, 2022) (http://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 was extracted from which scholarly paper). (3) PanglaoDB (Franzén, Gan and Björkegren, 2019) (https://panglaodb.se) interlinks cell types, genetic pathways and regulatory networks, covering pre-processed and pre-computed analyses from more than one thousand single-cell related experiments. (4) SC2disease (Zhao *et al.*, 2021) (http://easybioai.com/sc2disease) is a comprehensive resource for differentially expressed genes profiles, which supports comparisons between cell types, tissues and disease-related health states. It contains thousands of entries about different cells, tissues and diseases. (5) SPEED database (Chen, Zhang *et al.*, 2022) (http://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 *et al.*, 2019) are working on the development of general methods and computational (Börner, Teichmann, *et al.*, 2021; Manz *et al.*, 2022; Zhang *et al.*, 2022), demonstration projects (Burnum-Johnson *et al.*, 2022), organ-specific atlases (Becker *et al.*, 2022; Kruse and Spraggins, 2022), and novel technologies that can be used to map human tissue at the single cell level (Deng *et al.*, 2022; Melani *et al.*, 2022; Schachner *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 to serve as authors or reviewers of the evolving atlas.

Data and data processing

This section details all 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 datasets that are associated with about 80 scholarly papers. Data details and code are available at https://github.com/cns-iu/hra-evidence-supporting-information.

Publication evidence from HRA

The Human Reference Atlas captures data on ASCT+B tables (see Introduction and Fig. 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. Gene biomarkers used to characterize cell types in different organs are published via Azimuth (Hao *et al.,* 2021), 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.

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 references listed in the 26 ASCT+B Tables from the 3rd HRA release v1.2. 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 5 papers from PubMed other sources.

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**

*Note that the counts shown are the number of unique references corresponding to at organ level and at cell type level. And there are 456 unique references in 26 ASCT+B Tables.

2/3D reference objects & OMAPs

In the 3rd HRA release, there are 19 2D reference objects for functional tissue units in 7 organs with 90 unique cell types; 53 3D reference organs with 1,542 named anatomical structures, and 7 Organ Mapping Antibody Panels (OMAPs) for 187 anatomical structures, 179 cell types and 197 protein biomarkers across the 7 organs. Papers for the 2D and 3D References Library Objects and OMAPs were downloaded from the HuBMAP CCF Portal (https://hubmapconsortium.github.io/ccf) and are listed and summed up in Table 2. The publication references include 2 unique books and 14 unique scientific papers (with unique

DOIs). None of the 16 scholarly works are cited in ASCT+B tables, likely due to the fact that ASCT+B table references focus on the cell type level.

Table 2. Count of papers for 2/3D reference objects and OMAPs per organ. Counts equal the number of unique papers per organ per HRA object type. Total count equals the number of unique papers across all organs.

Organ	*2D Objects*	*3D Objects*	*OMAPs*	*Total #Papers*	*#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**

*Note that 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.
** Note that the brain organ has four 3D objects with one paper each.

Azimuth references

Azimuth references support cell type annotation for tissue datasets (Hao et al., 2021). They exist for 10 organs and references to associated publications can be downloaded from https://azimuth.hubmapconsortium.org and are listed in Table 3. HuBMAP focuses on adults (excluding fetal development) and there do not yet exist ASCT+B tables for adipose and tonsils. In total, there are 38 unique papers associated with the 10 Azimuth references and 2 of the papers are preprints. There are 36 unique papers with DOIs and 8 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.

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

Organ	*#Papers*	*#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.
*** Note that 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.

Experimental data references
A total of 308 datasets from single-cell studies of healthy human adults were retrieved from HuBMAP Portal (Cao *et al.*, 2019; Stuart *et al.*, 2019), CxG Portal (Domínguez Conde *et al.,* 2022; The Tabula Sapiens Consortium* *et al.,* 2022), NeMO (Orvis *et al.,* 2021), and GTEx (Eraslan *et al.,* 2022) in October 2022. These high quality datasets cover 57 organs and the datasets 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 datasets for organs that have ASCT+B tables were shared with table lead authors for possible inclusion in the ASCT+B tables.

Summary
In sum, there are 12 unique books, 439 unique papers (including 305 WoS core papers) and 5 papers from PubMed other sources listed in the 26 ASCT+B tables from the 3rd HRA 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 datasets, 195 have 49 unique papers associated.

WoS papers for 33 organs
To better understand which major papers were recently published on the 33 organs planned for the next HRA release, we ran a query over the Web of Science core collection provided via the Collaborative Archive & Data Research Environment (CADRE) (Mabry *et al.,* 2020; Wittenberg *et al.,* 2020). The retrieval result comprises 250,620 papers that were published in 2018 to 2022 and have these organ words in titles or keywords and were cited at least 10 times. These papers cover all the 33 organs except Blood 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-5 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), 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 USA.

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), CxG Portal (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: (The Tabula Sapiens Consortium* *et al.,* 2022), Lake et al. (2021), and (Ghose *et al.,* 2022). The number of datasets, organs, and papers per data source is shown in Table 4.

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

Data Source	Source Type	#Organs	#Datasets	#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. paper; other datasets are unpublished.

In total, the 8 data sources cover 57 unique organs, 308 unique datasets, and 82 unique references (77 of them with DOIs). The top-5 organs with most datasets are blood, brain (motor cortex), kidney, lung, and skin with 20, 28 (14), 153, 21, 14 datasets respectively. Kidney datasets come from four data sources: Azimuth reference, Lake et al., CxG, HuBMAP. The top-5 organs with the most unique references are blood (18 papers), heart (7 papers), kidney (7 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.

Quality and coverage of the HRA

A comparison of experimental data to anatomical structures, 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 (https://hubmapconsortium.github.io/ccf-asct-reporter) was used to visualize the network of anatomical structures, cell types, plus biomarkers in a 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.

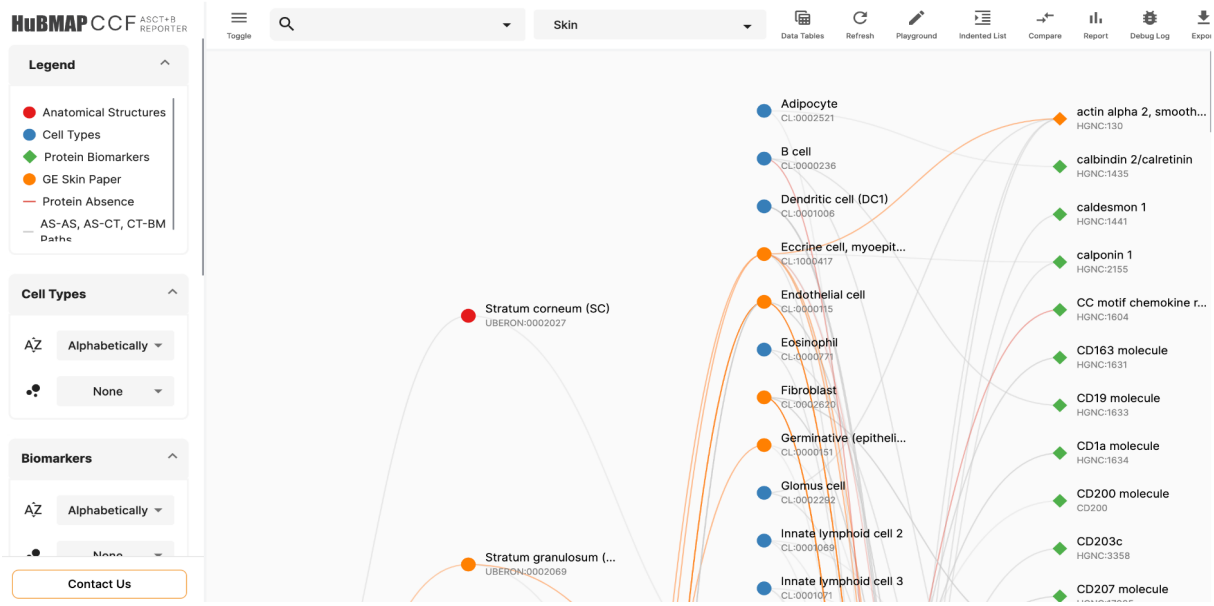

Figure 2. ASCT+B Reporter comparison visualization of an ASCT+B table and experimental data. Network of anatomical structures (red nodes), cell types (blue nodes), and protein biomarkers (green nodes) for the skin ASCT+B table is used as a basemap. Experimental data is overlaid in orange, making it easy to explore and communicate (non)matching anatomical structures, cell types plus biomarkers covered in a study.

Exemplarily, 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 dataset published in (Ghose et al., 2022). A partial screenshot of the interactive visualizations is shown in Fig. 2. 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%2F1ebxX1VmZXrxjfxZC8DdxtPjTGQLId9NBja71ii939c8%2Fedit%23gid%3D1990254927&comparisonName=Human%20Digital%20Twin&comparisonColor=%23ff8000&comparisonHasFile=false.

The visualization shows what data and publication evidence (here 10 datasets published in 1 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.

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—a total of 52 unique authors, 4 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, 32 experts, and for OMAPs 29 experts. Across the HRA, there are 116 unique experts and 113 of them with ORCID IDs.

Using the WoS papers data comprising 250,620 papers that featured any of the 33 organ names in their title and 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 co-author relationships is shown in Fig. 3. 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 US has 84,287 papers and is the most highly connected node with these top-5 collaborators: CN (9,541 papers), UK (6,338 papers), CA (5,605 papers), DE (5,131 papers), and IT (4,271 papers).

Figure 3. Geospatial layout of the co-author network. World map showing 672,892 experts in 189 countries and co-author relationships from the US to 170 countries. Node size and color represent the number of countries where co-authors' 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.

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 Fig. 4.

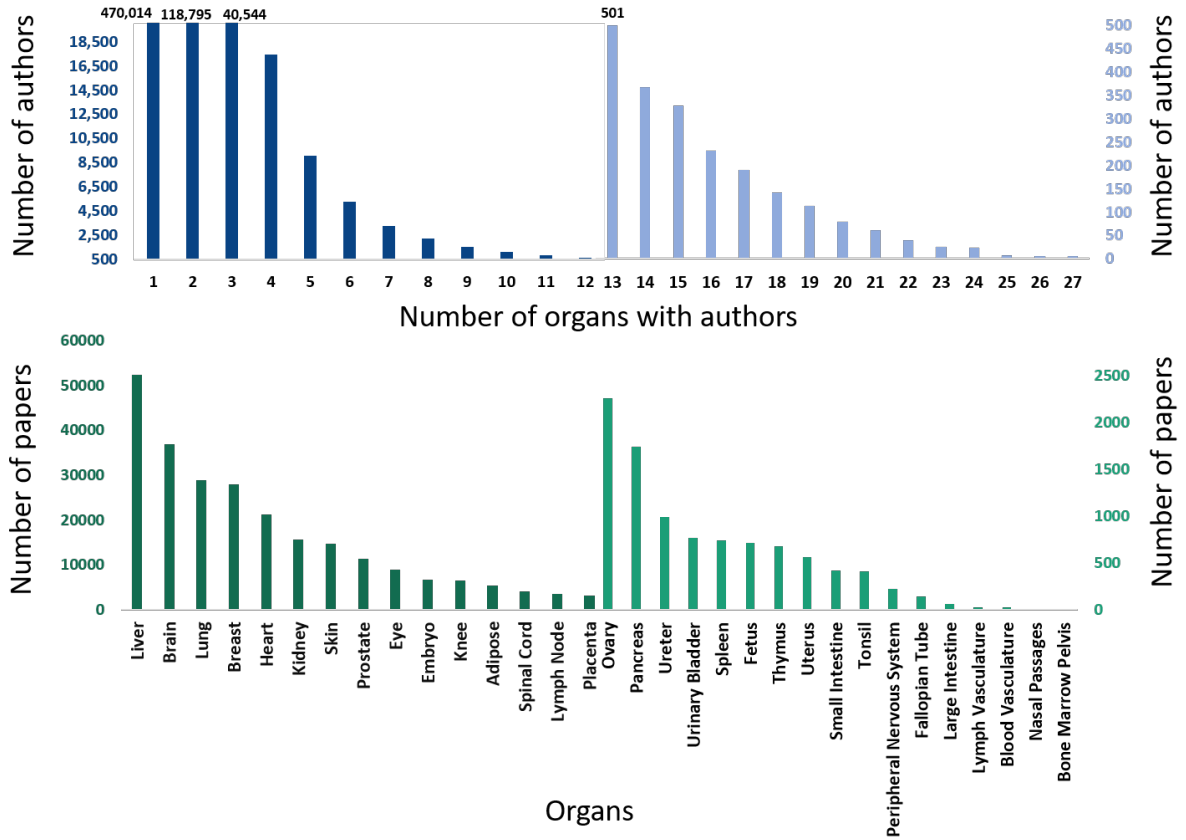

Figure 4. Number of organs experts work on (top) and number of papers per organ (bottom). Shown in the upper left is the range from 666 and 470,014 (deep blue bars) while the upper right shows the range between between 5 and 501 (light blue bar). Many authors (470,014) work on one organ, five authors 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 1 and 2,258 (light green). Many papers exist for the liver, followed by papers on the brain, lung, breast, heart, kidney, and skin.

To discern who funds work on the 33 organs, we extracted the bimodal network of organs and funding agencies. As the network was rather dense, we applied PFnet to retain the strongest linkages, see result in Fig. 5. The National Institutes of Health (NIH), United States Department of Health & Human Services (HHS), European Commission and UK Research & Innovation (UKRI) fund 32 organs (note that 'bone marrow pelvis' was excluded from this study as this combination did not retrieve any papers). Most papers on brain topics are funded by NIH, which is acknowledged in 24,352 papers.

At the author level, Fig. 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, Fig. 7. presents the number of authors per country per organ for country-organ combinations with more than 1000 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, while the United States has 11,391 brain experts and 11,086 liver experts.

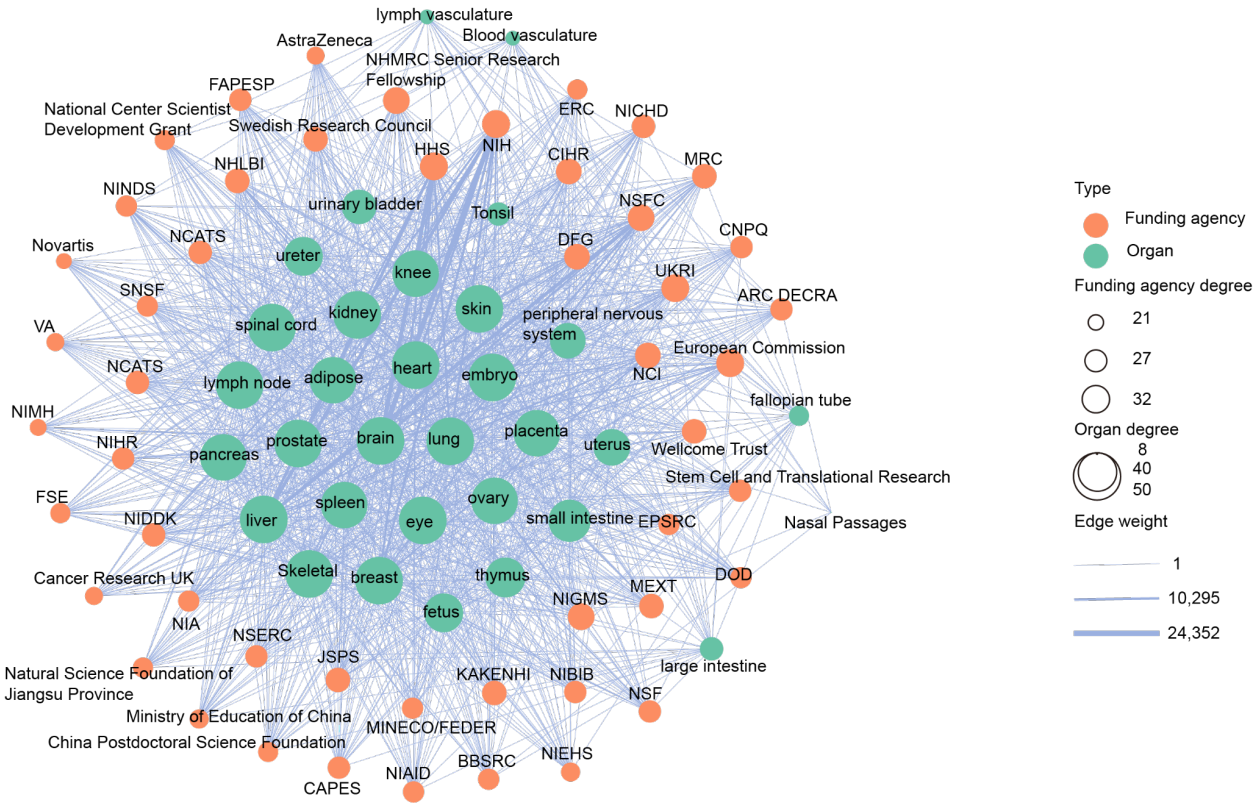

Figure 5. Bimodal network of 32 organs and top 50 funding agencies most often listed. Shown are 50 funding agencies (orange nodes) and 32 organs (green nodes). Funding agencies are area sized by the number of organs they support (i.e., node degree). Organs are sized by the number of funding agencies 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.

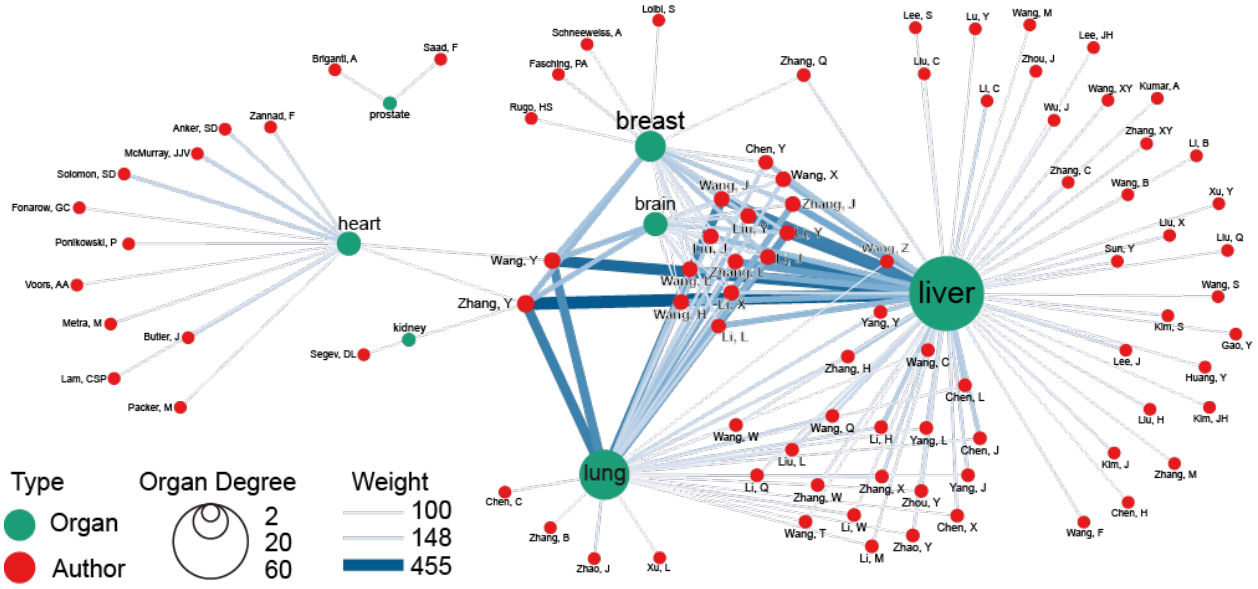

Figure 6. Bimodal network of highly cited authors and the organs they study. 7 organ nodes (green) and 88 author nodes (pink) are sized by the number of related authors and the number of related organs, respectively. Edge weights represent the number of papers published by an expert for one organ.

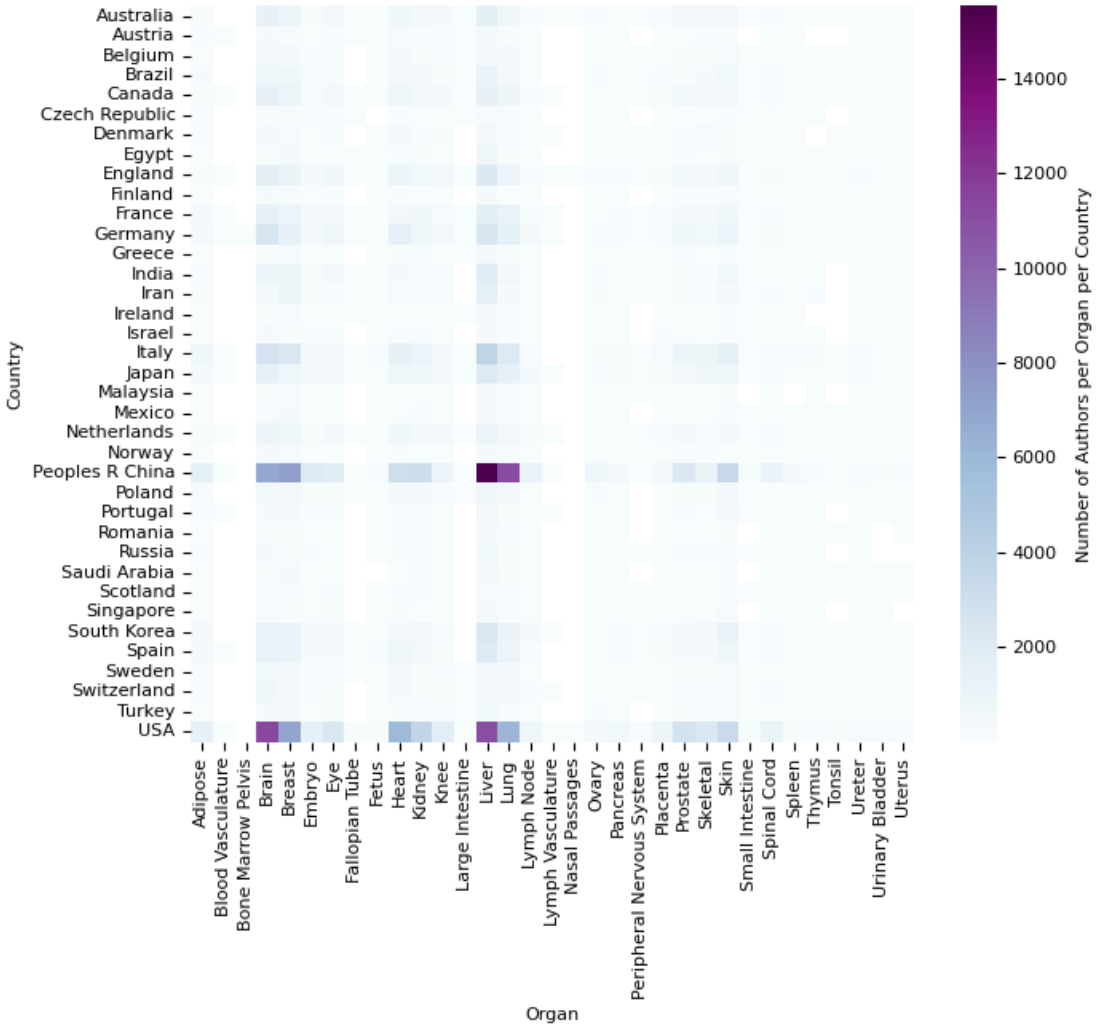

Figure 7. Heatmap of authors per country per organ. There are 37 countries which have contributed over 1,000 authors to the study of 33 organs and these are shown here. Dark purple color indicates a high number of publications. There are 22,483 papers by 15,526 authors from the People's Republic of China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

Summary and next steps

The paper presents the initial analyses and visualizations of scholarly papers and experimental dataset evidence for the Human Reference Atlas. We analyze the number and type of scholarly evidence for subgraphs of the HRA and show that 96.15 percent 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 exemplary coverage for skin in Fig. 2. We have been and will continue to share 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 just 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 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 datasets in support of highest quality HRA construction and usage of the HRA data in future scholarly publications.

Over the coming five years, we expect the number of active authors to grow from 200 to 1,000. The current set of organs will double to about 60 organs and organ parts, 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 non-trivial. Visualizations that show the coverage and quality of the evolving atlas, relevant expertise around the globe, and high quality experimental datasets will be critically important for communicating progress to experts and funders engaged in constructing or using the atlas.

Acknowledgements
We would like to thank Devin M. Wright and Ellen M. Quardokus for their support in the identification of HRA-relevant papers and experimental datasets; Abhay B. Rajde for sharing ASCT+B table data and expanding ASCT+B Reporter functionality with Bruce W. Herr II providing technical advice; Leonard Cross for designing Fig. 1. Nancy L. Ruschman served as a friendly reviewer and user of the expertise mapping and recommendation study.

Mike Gallant and Jacob J. Shaw provided access to Web of Science data using the Collaborative Archive & Data Research Environment (CADRE) project (http://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.

References

Becker, W. R., Nevins, S. A., Chen, D. C., Chiu, R., Horning, A. M., Guha, T. K., Laquindanum, R., et al. (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.

Börner, K., Bueckle, A., Herr, B. W., Cross, L. E., Quardokus, E. M., Record, E. G., Ju, Y., et al. (2022). Tissue registration and exploration user interfaces in support of a human reference atlas. *Communications Biology*, 5(1), 1–9. Nature Publishing Group.

Börner, K., Teichmann, S. A., Quardokus, E. M., Gee, J. C., Browne, K., Osumi-Sutherland, D., Herr, B. W., et al. (2021). Anatomical structures, cell types and biomarkers of the Human Reference Atlas. *Nature Cell Biology*, 23(11), 1117–1128.

Burnum-Johnson, K. E., Conrads, T. P., Drake, R. R., Herr, A. E., Iyengar, R., Kelly, R. T., Lundberg, E., et al. (2022). New Views of Old Proteins: Clarifying the Enigmatic Proteome. *Molecular & Cellular Proteomics*, 21(7), 100254.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566(7745), 496–502.

Chen, S., Luo, Y., Gao, H., Li, F., Chen, Y., Li, J., You, R., et al. (2022). hECA: The cell-centric assembly of a cell atlas. *IScience*, 25(5), 104318.

Chen, Y., Zhang, X., Peng, X., Jin, Y., Ding, P., Xiao, J., Li, C., et al. (2022). SPEED: Single-cell Pan-species atlas in the light of Ecology and Evolution for Development and Diseases. *Nucleic Acids Research*, gkac930.

Deng, Y., Bartosovic, M., Ma, S., Zhang, D., Kukanja, P., Xiao, Y., Su, G., et al. (2022). Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature*, 609(7926), 375–383.

Domínguez Conde, C., Xu, C., Jarvis, L. B., Rainbow, D. B., Wells, S. B., Gomes, T., Howlett, S. K., et al. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*, 376(6594), eabl5197.

Eraslan, G., Drokhlyansky, E., Anand, S., Fiskin, E., Subramanian, A., Slyper, M., Wang, J., et al. (2022). Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science*, 376(6594), eabl4290.

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. Retrieved November 23, 2022, from https://academic.oup.com/database/article/doi/10.1093/database/baz046/5427041

Ghose, S., Ju, Y., McDonough, E., Ho, J., Karunamurthy, A., Chadwick, C., Cho, S., et al. (2022). *Human Digital Twin: Automated Cell Type Distance Computation and 3D Atlas Construction in Multiplexed Skin Biopsies* (preprint). Cell Biology. Retrieved November 20, 2022, from http://biorxiv.org/lookup/doi/10.1101/2022.03.30.486438

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell*, 184(13), 3573-3587.e29.

Herr, B. W., Hardi, J., Quardokus, E. M., Bueckle, A., Chen, L., Wang, F., Caron, A. R., et al. (2023). Specimen, biological structure, and spatial ontologies in support of a Human Reference Atlas. *Scientific Data*, 10(1), 171.

Hickey, J.W., Neumann, E.K., Radtke, A.J. et al. (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.

Hu, C., Li, T., Xu, Y., Zhang, X., Li, F., Bai, J., Chen, J., et al. (2022). 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*, gkac947.

HuBMAP Consortium, Writing Group, Snyder, M. P., Lin, S., Posgai, A., Atkinson, M., Regev, A., et al. (2019). The human body at cellular resolution: The NIH Human Biomolecular Atlas Program. *Nature*, 574(7777), 187–192.

Kruse, A. R. S., & Spraggins, J. M. (2022). Uncovering Molecular Heterogeneity in the Kidney With Spatially Targeted Mass Spectrometry. *Frontiers in Physiology*, 13, 837773.

Lake, B. B., Menon, R., Winfree, S., Hu, Q., Ferreira, R. M., Kalhor, K., Barwinska, D., et al. (2021). An atlas of healthy and injured cell states and niches in the human kidney. *BioRxiv*, 2021.07.28.454201.

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.

Manz, T., Gold, I., Patterson, N. H., McCallum, C., Keller, M. S., Herr, B. W., Börner, K., et al. (2022). Viv: Multiscale visualization of high-resolution multiplexed bioimaging data on the web. *Nature Methods*, 19(5), 515–516.

Melani, R. D., Gerbasi, V. R., Anderson, L. C., Sikora, J. W., Toby, T. K., Hutton, J. E., Butcher, D. S., et al. (2022). The Blood Proteoform Atlas: A reference map of proteoforms in human hematopoietic cells. *Science*, 375(6579), 411–418.

Orvis, J., Gottfried, B., Kancherla, J., Adkins, R. S., Song, Y., Dror, A. A., Olley, D., et al. (2021). gEAR: Gene Expression Analysis Resource portal for community-driven, multi-omic data exploration. *Nature Methods*, 18(8), 843–844.

Schachner, L. F., Soye, B. D., Ro, S., Kenney, G. E., Ives, A. N., Su, T., Goo, Y. A., et al. (2022). Revving an Engine of Human Metabolism: Activity Enhancement of Triosephosphate Isomerase via Hemi-Phosphorylation. *ACS Chemical Biology*, 17(10), 2769–2780.

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.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., et al. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, 177(7), 1888-1902.e21.

The Tabula Sapiens Consortium*, Jones, R. C., Karkanias, J., Krasnow, M. A., Pisco, A. O., Quake, S. R., Salzman, J., et al. (2022a). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science*, 376(6594), eabl4896.

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 http://hdl.handle.net/2022/25309.

Zhang, B., Srivastava, A., Mimitou, E., Stuart, T., Raimondi, I., Hao, Y., Smibert, P., et al. (2022). Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro. *Nature Biotechnology*, 40(8), 1220–1230.

Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., et al. (2019). CellMarker: A manually curated resource of cell markers in human and mouse. *Nucleic Acids Research*, *47*(D1), D721–D728.

Zhao, T., Lyu, S., Lu, G., Juan, L., Zeng, X., Wei, Z., Hao, J., et al. (2021). SC2disease: A manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Research*, *49*(D1), D1413–D1419.

* Corresponding authors: Yongxin Kong, yokong@iu.edu; Katy Börner, katy@indiana.edu.