

Human BioMolecular Atlas Program (HuBMAP): 3D Human Reference Atlas construction and usage

Received: 27 March 2024

Accepted: 11 November 2024

Published online: 13 March 2025

 Check for updates

Katy Börner^{1,2}✉, Philip D. Blood³, Jonathan C. Silverstein⁴, Matthew Ruffalo⁵, Rahul Satija⁶, Sarah A. Teichmann^{2,7,8}, Gloria J. Pryhuber⁹, Ravi S. Misra⁹, Jeffrey M. Purkerson⁹, Jean Fan¹⁰, John W. Hickey¹¹, Gesmira Molla⁶, Chuan Xu^{10,7}, Yun Zhang¹², Griffin M. Weber¹³, Yashvardhan Jain¹, Danial Qaurooni¹, Yongxin Kong¹, HRA Team*, Andreas Bueckle¹✉ & Bruce W. Herr II¹✉

The Human BioMolecular Atlas Program (HuBMAP) aims to construct a 3D Human Reference Atlas (HRA) of the healthy adult body. Experts from 20+ consortia collaborate to develop a Common Coordinate Framework (CCF), knowledge graphs and tools that describe the multiscale structure of the human body (from organs and tissues down to cells, genes and biomarkers) and to use the HRA to characterize changes that occur with aging, disease and other perturbations. HRA v.2.0 covers 4,499 unique anatomical structures, 1,195 cell types and 2,089 biomarkers (such as genes, proteins and lipids) from 33 ASCT+B tables and 65 3D Reference Objects linked to ontologies. New experimental data can be mapped into the HRA using (1) cell type annotation tools (for example, Azimuth), (2) validated antibody panels or (3) by registering tissue data spatially. This paper describes HRA user stories, terminology, data formats, ontology validation, unified analysis workflows, user interfaces, instructional materials, application programming interfaces, flexible hybrid cloud infrastructure and previews atlas usage applications.

Inaugurated in 2018, the Human BioMolecular Atlas Program (HuBMAP) aims to construct a comprehensive reference model of the healthy ('non-diseased') human body across all levels, from organs and tissues down to cells and canonical biomarkers^{1,2}. The HuBMAP Portal (<https://hubmapconsortium.org>) introduces goals and links to experimental and atlas data, tools and training materials. The Data Portal (<https://portal.hubmapconsortium.org>) serves experimental datasets and supports data processing, search, filtering and visualization. The Human Reference Atlas Portal (<https://humanatlas.io>) provides open access to atlas data, code, procedures and instructional materials. The Human Reference Atlas (HRA)³ includes a Common Coordinate Framework (CCF; see Box 1), which helps harmonize multimodal data,

including three-dimensional (3D) organ models, histology images and omics data from profiling of single cells. HRA data comprise human-expert-generated information (for example, anatomical systems; anatomical structures, cell types, biomarker (ASCT+B) tables and two-dimensional (2D) and 3D reference objects), experimental data mapped to the HRA, as well as enriched atlas data in support of different atlas applications. The origin and evolution of HRA and ASCT+B tables are detailed in previous work³. The CCF provides quantitative workflows for integrating new experimental data into the growing atlas, such as histology images, vascular pathways and single-cell analyses. The resulting HRA provides data evidence for common states of cells and anatomical structures in the human body at specific 3D locations

A full list of affiliations appears at the end of the paper. ✉ e-mail: katy@iu.edu; abueckle@iu.edu; bherr@iu.edu

BOX 1

Key HRA terminology

- 3D collision: the intersection of 2D or 3D bounding-box volumes or surface polygon meshes.
- 3D model: DO with a shape and size defined by polygon meshes (vertices and edges) that can be used to represent the real-world form of anatomical structures, cells or proteins in 3D.
- 3D Reference Object: polygon mesh of 3D spatial objects (for example, anatomical structures), their object node hierarchy, materials and surface color and texture. They are created by medical illustrators with the involvement of subject matter experts following standard operating procedures.
- Anatomical structure: a distinct biological entity with a 3D volume and shape, for example, an organ, FTU or cell.
- AVR: document providing details on the characterization of individual antibodies for multiplexed antibody-based imaging assays. These details include target protein information (for example, target name, UniProt accession number) and antibody information (for example, RRID, host organism, vendor, catalog number, lot number). AVRs also provide details on controls used for antibody characterization and validation (positive and negative tissues, cell lines, isotype controls, etc.), exemplar imaging data and information on other antibodies tested.
- ASCT+B tables: these are authored by multiple experts across many consortia. Tables capture the partonomy of anatomical structures, cell types and major biomarkers (for example, gene, protein, lipid or metabolic markers) defining cellular identity supported by scientific evidence and are linked to ontologies.
- Atlas-enriched dataset graph: a graph for highest quality datasets used for HRA construction (it has an extraction site, cell type population and publication or provenance on a major atlas portal), enriched with additional metadata, computed by HRApop.
- Biomarkers (B): HRA biomarkers are used to characterize or identify cell types. They include genes (BG), proteins (BP), metabolites (BM), proteoforms (BF) and lipids (BL).
- Cell suspension: single cells or nuclei isolated from a tissue (for example, using enzymes or mechanical means) for single-cell assays, for example, before sc/snRNA-seq assay is run.
- Cell types: tissue is composed of different (resident and transitory) cell types that are characterized or identified via biomarkers.
- CTann: Azimuth and other cell type annotation tools are used to assign cell types to cells from sc/snRNA-seq studies. Manually compiled crosswalks are used to assign ontology IDs to CTann cell types.
- Cell type population: a listing of unique cell types, the number of cells per cell type and mean biomarker expression values per cell type computed for anatomical structures, extraction sites and datasets.
- CCF: the HRA CCF consists of ontologies and reference object libraries, computer software (for example, user interfaces) and training materials that (1) enable biomedical experts to semantically annotate tissue samples and to precisely describe their locations in the human body ('registration'); (2) align multimodal tissue data extracted from different individuals to a 3D reference coordinate system ('mapping') and; (3) provide tools for searching and browsing HuBMAP data at multiple levels from the whole body down to single cells ('exploration'). Alternative CCF definitions do exist⁴².
- Cosine similarity: measures the cosine of the angle between two vectors, with 1 indicating identical vector directions and 0 indicating no similarity.
- Crosswalk: an ontological mapping of terms in HRA DOs (for example, 2D/3D reference objects, OMAPs and Azimuth references) to ontology terms in the ASCT+B tables.
- Dataset graph: a JSON-LD file containing a graph of RUI registration, donor, experimental (for example, links to cell by biomarker (C×B) H5AD files or cell type population), literature and provenance data.
- Digital object (DO): unit of information that includes properties (attributes or characteristics of the object) and may also include methods (means of performing operations on the object).
- Digital object identifier (DOI): centrally registered identifier composed of a string of numbers, letters and symbols used to uniquely identify an article or document with a permanent web address or uniform resource locator (URL).
- Extraction site: digital, 3D representation of a tissue block. If the RUI is used to register tissue, then each site has a unique ID; data on size, location and rotation in 3D in relation to a HRA reference organ; a listing of all anatomical structures that the cuboid intersects with (bounding-box collision by default); and metadata on who registered it.
- FAIR principles: acronym for findable, accessible, interoperable and reusable, which is a way of sharing data to maximize its utility¹⁷.
- FTU: a small tissue organization (that is, set of cells) that performs a unique physiologic function and is replicated multiple times in an organ. Examples are liver lobule, alveolus of lung or pancreatic acinus.
- H&E stain: histology stain that is widely used as a gold standard for pathological evaluation of tissue sections.
- HRA: the HRA is a comprehensive, high-resolution, three-dimensional atlas of major cells in the healthy human body. The HRA provides standard terminologies and data structures for describing specimens, biological structures and spatial positions linked to existing ontologies.
- HRAlit: scholarly publication linked to HRA DOs to provide scholarly evidence.
- HRApop: experimental data linked to HRA DOs to provide data evidence and number of cells per cell type per 3D anatomical structure.
- LOD: a method for sharing data in standard, non-proprietary formats and deeply interlinked with other data resources.
- Millitome: a device used to hold and slice organs into a grid of equally sized tissue blocks plus a process of generating HRA-aligned digital tissue blocks.
- Ontology: a set of subject area concepts (here, anatomical structures, cell types and biomarkers), their properties and the relationships between them. Ontologies used in the HRA include Uberon and Cell Ontology (CL).

(continued from previous page)

- OMAP: a comprehensive panel of curated antibodies that identifies the major anatomical structures and cell types in a specific organ. The selected antibodies are optimized for a tissue preservation method (fixed or frozen) and multiplexed imaging modality (for example, CODEX and Cell DIVE) through published protocols (protocols.io) and AVR.
- Partonomy: a classification hierarchy that represents part-whole relationships.
- Polygon mesh: a collection of vertices, edges and faces defining the shape of a polyhedral object, for example, tissue block cuboids or reference objects denoting anatomical structures.
- Reference objects: a 3D model of anatomical structures created by medical illustrators with the involvement of subject-matter experts following SOPs.
- Registration set (see also dataset graph): grouping of tissue blocks by the study/paper in which they were published. Each set has a human-readable registration set name, a machine-readable internationalized resource identifier and it links to a paper DOI and tissue block ID.
- RUI registered: tissue spatially and semantically registered to the HRA using the RUI.
- Segmentation: image processing that predicts boundaries of objects, for example, anatomical structures such as nuclei, cellular membranes, cells or FTUs in 2D or 3D.
- Single-cell transcriptomic data (sc/snRNA-seq): data from single-cell (sc) or single-nucleus (sn), high-throughput suspension-based studies that measure polyadenylated RNA molecules in an individual cell.
- Single-cell proteomic data: data from single-cell studies using CODEX, Cell DIVE, Ibex, CycIF or other assays that detect proteins in situ in a tissue, consequently enabling protein expression quantification.
- Tissue block: a sample or specimen derived from an organ or tissue including subsections thereof obtained from a donor that has a unique ID and links to donor organ extraction site, processing, preservation and other metadata. The locations of tissue blocks are registered using the RUI.
- Tissue section: a thin (several- μm) section of a tissue block usually obtained using a cryotome or microtome. Tissue sections inherit the location and rotation of their parent tissue block. The thickness and number of tissue sections is captured in an input field inside the RUI.
- Typology: a classification that represents general types, for example, cell types or biomarker types.
- United file: a GLB 3D object that contains all the modeled organs in the HRA.
- VCCF: a proposed approach to use vasculature as a coordinate system to map all the cells in the human body.
- Web Ontology Language (.owl) file format: the World Wide Web Consortium (W3C) Web Ontology Language is a Semantic Web language (documented at <https://www.w3.org/OWL>).

and this can be used as a canonical reference to describe the changes that occur across biological variables (for example, age, sex, race and body mass) and acute or chronic diseases. It can benefit applications such as drug development by providing a better understanding of perturbations of cell types and states in diseased conditions, which could reveal relevant targets for precision medicine through comparisons of diseased to non-diseased tissue, and a CCF-matched reference.

When HuBMAP was launched, the first unifying concepts that map major organs in the human body across scales were emerging^{4,5}. Existing atlases used organ-specific references (for example, Waxholm Space for the brain⁶ or the Helmsley one-dimensional distance reference system for the colon), but most of these references do not map to a shared human body CCF⁷. To advance CCF development, in March 2020, the National Institutes of Health (NIH) and the Human Cell Atlas (HCA)⁸ Consortium organized a joint virtual meeting with a CCF breakout session. This resulted in the formation of the HRA Working Group (WG). Over the last 55 months, WG members jointly developed a definition and key properties for the HRA. These properties are:

1. **The HRA defines a reference 3D multiscale space and shape of anatomical structures and cell types and the biomarkers used to characterize cell types.** Anatomical structures, cell types and biomarkers are validated against, or are added to, existing ontologies (for example, the uber-anatomy ontology (Uberon)⁹, the Foundational Model of Anatomy Ontology (FMA)^{10,11}, Cell Ontology (CL)¹², Provisional Cell Ontology (PCL)¹³ and the Human Gene Ontology Nomenclature Committee (HGNC; <https://www.genenames.org>)). As more data are collected, the HRA will increasingly be able to show how body shape and size, plus cell type populations differ across individuals and change over a person's lifespan.
2. **The HRA enables adding new experimental datasets and mapping these to existing data through a variety of mechanisms.** For example, the location of tissue specimens can be specified relative to virtual 3D reference organ models in the

HRA; single-cell genomic data can be mapped to the HRA using annotation tools like Azimuth¹⁴; and, single-cell-resolution spatial proteomics data can be mapped to the HRA using validated organ mapping antibody panels (OMAPs)¹⁵. With the development of new technologies and computational methods in the future, additional mappings and linkages, such as integration of multi-omics data, will be possible.

3. **The HRA follows best practices and standards for sharing scientific data.** To do this, the HRA needs to be authoritative (it should be supported by peer-reviewed scholarly publications, experimental data evidence or expert consensus); meet the Transparency, Responsibility, User focus, Sustainability and Technology (TRUST) principles for digital repositories¹⁶; be representative (covering all major human demographics and welcoming everyone to contribute to and use the HRA data); be open and adhere to the Findable, Accessible, Interoperable, Reusable (FAIR) principles¹⁷ (where anyone can use the HRA data and code and these are provided in community standard formats with linked ontologies; be published as linked open data (LOD) connected to ontologies and other LOD; and application programming interface (API) queries and user interfaces are supported); detailed protocols and standard operating procedures (SOPs) are provided; and be continuously evolving (for example, as new technologies, data and methods become available).

Experts also agreed on key SOPs for atlas construction and usage plus HRA terminology (Box 1), adopted from the HRA SOPs Glossary¹⁸. Plus, the HRA WG brought together technical leads from the HuBMAP Integration, Visualization & Engagement (HIVE) Collaboration with experimental teams in HuBMAP and experts from the Genotype-Tissue Expression (GTEx)¹⁹, GUDMAP: GenitoUrinary Developmental Molecular Anatomy Project²⁰, Kidney Precision Medicine Project (KPMP)^{21,22}, LungMAP^{23,24}, BRAIN Initiative Cell Census Network Initiative (BICCN)^{25,26}, Cellular Senescence Network (SenNet)²⁷

Table 1 | User stories. Feature summary, target user roles, user activities and added value for seven user stories that drive HRA development

Feature	User role	User activities	Added value
Facilitate atlas construction by aligning new tissue blocks with existing data			
US 1. Predict cell-type populations	Programmers that support researchers, clinicians, pathologists	Predict and explore the likely cell type populations for a RUI-registered tissue block.	Improve cell type annotation through information on what cell type populations exist in what anatomical structures.
US 2. Predict spatial origin of tissue samples	Programmers that support Researchers, clinicians	Predict and explore the likely 3D location in the human body for a given tissue block with known cell type population.	Compensate for the absence of spatial origin information in many single-cell datasets.
Use the atlas to gain insights into changes that occur at all levels in the body with aging or disease			
US 3. Compare reference tissue with aging/diseased tissue	Researchers, clinicians	Compare tissue blocks, cell types and biomarker expression levels between healthy reference tissue and aging/diseased tissue.	Understand and communicate changes in tissue structure and function with age or disease.
US 4. Compare reference FTU with aging/diseased FTUs	Researchers, clinicians	Compare FTUs in terms of cell types and mean biomarker expression levels for healthy reference tissue and aging/diseased tissue.	Understand and communicate changes in FTU structure and function with age or disease
US 5. Provide cell distance distribution visualizations	Researchers, pathologists	Compute, visualize and explore distance distributions between different cells, cell types and anatomical structures (for example, FTUs) and cell types and morphological features (for example, the edge of an organ).	Add granularity to our understanding of how disease develops (for example, how tumor cells grow or metastasize) in support of targeted therapies.
Ensure atlas sustainability with processes that encourage collaboration and guide future development			
US 6. Develop lightweight atlas components	Programmers that support researchers and clinicians	Implement usable and useful HRA components (interfaces and APIs) into other portals in the growing ecosystem of human atlases.	Facilitate collaboration and data/code reuse between the HRA and other portals in support of FAIR data principles.
US 7. Implement dashboard for HRA	Researchers, clinicians, funders	Track the evolution and usage of the HRA using data, code and portal usage statistics in aggregate and divided by portal (for example, HuBMAP or SenNet) or user survey results.	Enable evidence-based decision-making by providing insights into the atlas' construction and usage (for example, gaps in data, application areas, user demographics, equitable access).

and other NIH-funded consortia and with strong support from the HCA effort^{8,28} to develop the HRA data, code and portal infrastructure together.

An important next step in the HRA effort was collecting user stories in support of atlas construction and usage to encourage dialog, deliberation and iteration by designers and users around three key questions: Who are the users involved in a particular user story? What are the outcomes they hope to achieve? What value do they stand to gain? Agreement on answers to these questions helped prioritize user needs, provided context for proposed user stories and reduced ambiguity.

More than 30 one-on-one interviews were conducted with atlas architects (experts who serve as principal investigators or are otherwise intimately involved in the construction of the latest generation of human atlases, including BICCN, GTeX, GUDMAP, HCA, HuBMAP, Human Tumor Atlas Network (HTAN)²⁹, KPMP, LungMAP, (Re)building the Kidney (RBK)³⁰ and SenNet). Given the interdisciplinary nature of this effort, the atlas architects who were interviewed comprise a diverse group of physicians, laboratory and computational biologists, engineers and computer and data scientists. In addition, six programmers from different human atlas projects were surveyed.

The interview and survey results helped identify three key objectives and seven concrete user stories (US 1–7) for the construction, usage and sustainability of the HRA (Table 1). The three objectives are:

1. The HRA should facilitate atlas construction by aligning new tissue blocks with existing data. For example, developers of the HRA want to predict cell type populations for new tissue blocks (US 1 in Table 1) and predict the spatial origin of tissue samples with known cell type populations (US 2).
2. The HRA should contain functionality that provides insights into changes (for example, with aging, disease or other perturbations) that occur at all levels in the body. To do this, researchers

and clinicians need to be able to search for and explore cell types and biomarker expression values for tissues and functional tissue units (FTUs) (US 3 and US 4) and determine the location and distances between cells (US 5).

3. The HRA should use processes that encourage collaboration and guide future development to ensure long-term sustainability. This includes leveraging architectures with modular, lightweight components that can be easily shared (US 6) with metrics of success provided via an HRA dashboard to researchers, clinicians and funders to gain feedback and support (US 7).

The three key objectives and associated user stories help focus presentations and discussions in the monthly HRA WG; they are driving HRA development and iterative optimization. Every 6 months, a new HRA release is published. With every release, existing ontologies are expanded plus HRA data structures and algorithms are improved to better serve the needs of the international human atlas community. Figure 1 details major components of the sixth release of the HRA and their interlinkages.

Results

The HIVE Infrastructure and Engagement Component (IEC) developed HuBMAP's flexible hybrid cloud microservices architecture (Supplementary Fig. 1 and Methods) to support data curation, ingestion, integration, access, analysis, exploration and download via the HuBMAP Data Portal (<https://portal.hubmapconsortium.org>). HIVE Tools Components focused on the HuBMAP Data Portal User Interface, visualization, workflow integration and tool development. HIVE Mapping Components developed Azimuth¹⁴ references and the HRA Portal (<https://humanatlas.io>) in close collaboration with external experts.

The HuBMAP Consortium website (Supplementary Fig. 2) provides easy access to HuBMAP resources, publications, news, internship

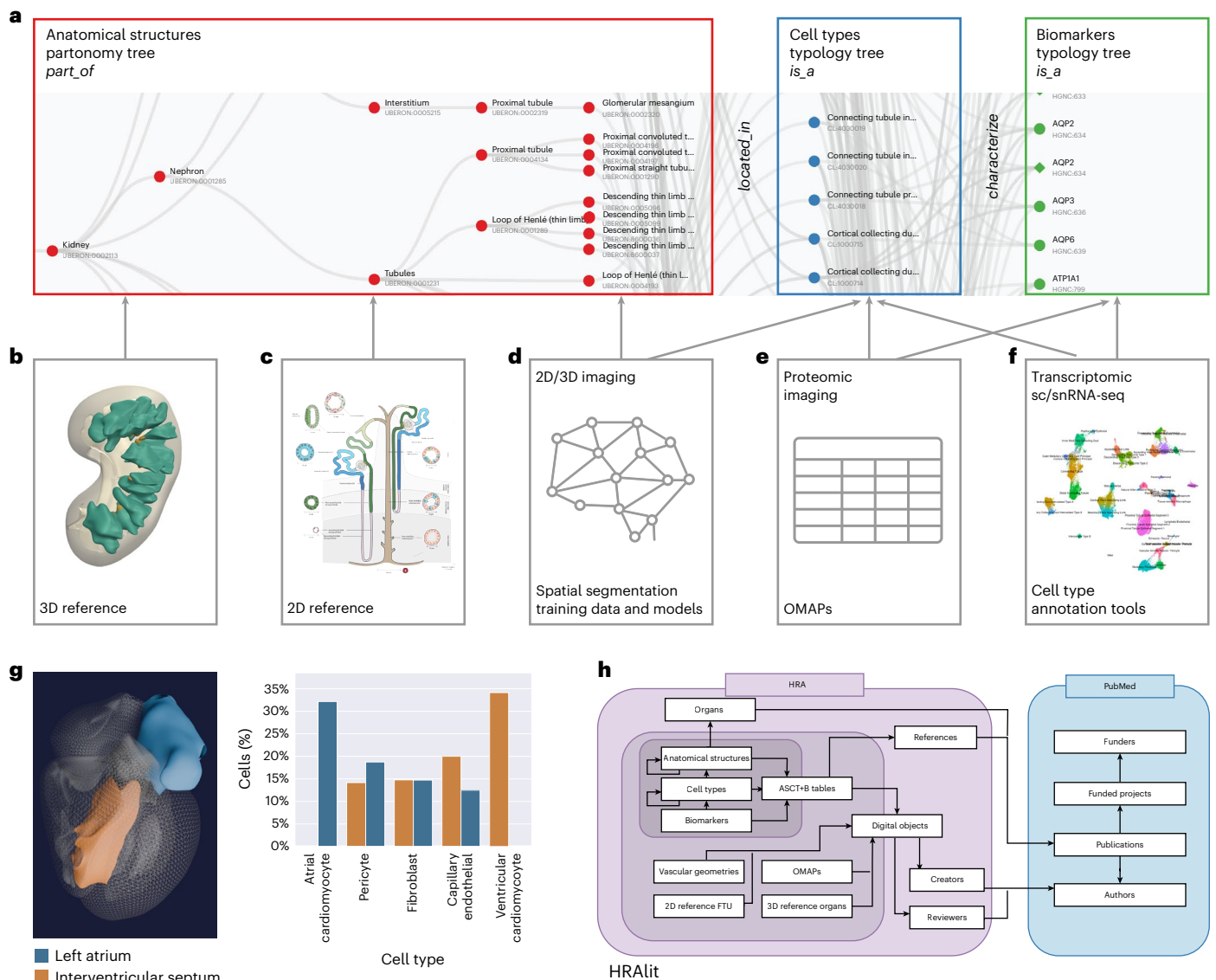


Fig. 1 | Human Reference Atlas components and linkages. **a**, The ASCT+B tables document the nested ‘part_of’ structure of organs (for example, cells that make up FTUs, successively larger anatomical structures, an entire organ such as the kidney, which is part_of the body). The cells that make up (are ‘located_in’) each of the anatomical structures are organized in a multilevel cell-type typology with ‘cell’ at the root and more and more specialized child nodes and ‘is_a’ relationships between nodes. The biomarkers used to characterize cell types might have one of five types: genes, proteins, metabolites, proteoforms and lipids organized in a biomarker typology. Gray arrows indicate crosswalks that connect other HRA DOs to ASCT+B tables. **b**, HRA 3D reference objects represent the shape, size, location and rotation of 1,192 3D anatomical structures with 516 unique Uberon IDs for 65 organs with crosswalks to ASCT+B tables. Shown are ‘renal papilla’ and ‘renal pyramid’ in the kidney. **c**, 2D reference illustrations document the shape, size and spatial layout of 3,742 2D cells of 116 types for

22 FTUs in ten organs with crosswalks to ASCT+B tables. Shown is the kidney nephron. **d**, Labeled training data exist for FTUs in five organs (gray arrows) to anatomical structures and cell types in the ASCT+B tables. **e**, 13 OMAPs are linked to 197 AVRs and there exist crosswalks to cell types and biomarkers in ASCT+B tables. **f**, Ten Azimuth references for healthy adult organs plus crosswalks to cell types and biomarkers in ASCT+B tables. **g**, HRApop reports cell type populations for anatomical structures compiled from experimental data. Exemplarily shown is the left atrium (blue) and the interventricular septum (orange) of the female heart plus a bar graph with the cell types that have the highest percentage across these two anatomical structures (annotated with Azimuth). Note that some cell types appear only in one anatomical structure. **h**, The HRAlit database links HRA DOs to existing ontologies (for example, Uberon and CL), expert ORCID, publication evidence, funding and experimental data used for HRApop computation.

programs, member services, etc. It links to the HuBMAP Data Portal and the HRA Portal. The HuBMAP Data Portal provides access to HuBMAP data, APIs and user interfaces with continuous data and code releases. The HRA Portal serves atlas-level data and code created by 18 atlas projects and new HRA releases are published every 6 months. Both portals use knowledge graphs (KGs) to store data and the HRA KG is regularly ingested into the Unified Biomedical Knowledge Graph (UBKG; <https://ubkg.docs.xconsortia.org>) to link HuBMAP experimental data to existing ontologies and the HRA. The HRA uses HuBMAP and other

experimental data to compute cell type populations for anatomical structures (see HRA cell type populations (HRApop) in Methods and Supplementary Table 1). Several HRA user interfaces (see section User interfaces) are deployed in the HuBMAP Data Portal and other portals to support HRA construction and usage.

Atlas construction is complex and requires community agreement on data formats, APIs and user interfaces. Previews are used to showcase and optimize new functionality before it is integrated into the HuBMAP or HRA Portal. Primary data repositories

are listed in Supplementary Table 2 and HRA code repositories in Supplementary Table 3.

Flexible hybrid cloud infrastructure for HRA and HuBMAP

Systematic integration of more than 50 open-source algorithms developed by more than 30 teams is non-trivial. Agreement on meta-data and API calls is required to make the output of one algorithm compatible with the input expected by the next (set of) algorithms. Several algorithms crucial to tissue segmentation and annotation were developed by biologists with deep subject-matter domain expertise but limited knowledge on how to build production pipelines. The HIVE production development team worked closely with the original algorithm authors to package their algorithms in a way so that they can be run reliably at scale in a hybrid cloud infrastructure that is flexible and extendable to meet evolving needs.

Specifically, the HIVE IEC, composed of members from the Pittsburgh Supercomputing Center (PSC), the University of Pittsburgh (Pitt) and Stanford University, implemented a flexible hybrid cloud infrastructure and community engagement platform supporting delivery of HuBMAP's vision in the following key areas: (1) curation and ingestion: semi-automated data ingestion (<https://software.docs.hubmapconsortium.org>) currently from HuBMAP data providers and (in the future) from community partners and the general research community, to maximize efficiency and usefulness for building the HRA; (2) integration: automated analysis and annotation of ingested data and alignment of these annotations to the HRA via the UBKG; (3) findability and accessibility: manifestation of backend resources in the modular architecture of APIs and containers, services and documentation (<https://software.docs.hubmapconsortium.org>) that minimizes user friction in integrated searching, querying, analyzing and viewing of HuBMAP data and in the future of tissue maps at multiple spatial scales and among multiple layers of information; (4) interoperability: use of the HuBMAP deployment of the UBKG with extensions to create the HuBMAP Ontology API (<https://smart-api.info/ui/d10ff85265d8b749fbc3ad7b51d0bf0a>) to translate HuBMAP data, HRA assets and community data among one another via ontologies; the HuBMAP Ontology API contains end points for querying a UBKG instance with content from the HuBMAP context (<https://ubkg.docs.xconsortia.org/contexts/#hubmapsennet-context>); (5) analysis: infrastructure support to currently enable users with interactive analyses of HuBMAP data via Jupyter notebooks and in the future, batch workflows among both HuBMAP and user-contributed data and tools, including integration and mapping against the HRA; and (6) sustainability: HuBMAP's flexible hybrid cloud infrastructure (efficiently leveraging on-premises resources at PSC for services that would incur much higher public cloud charges compared to on-premises, such as data storage, processing, analysis and download (Supplementary Fig. 1 and Methods)) will facilitate sustainability of open tools, data and infrastructure beyond the end of the HuBMAP program.

Atlas construction and publication

HRA data comprises human-expert-generated data (for example, ASCT+B tables, OMAPs, antibody validation reports (AVRs) and 2D/3D reference objects), experimental data mapped to the HRA (via registration user interface (RUI) location, HRA-aligned cell type annotation (CTann) or OMAP/AVR) and enriched atlas data (for example, HRApop and HRA literature (HRAlit)); see Fig. 1 for an overview of HRA digital object (DO) types and their crosswalks (see Box 1 for terminology and Methods for details). HRA data, usage and extension of ontologies, unified data processing workflows, user interfaces, documentation and instructional materials are detailed here.

Data types and status. The sixth release of the HRA v.2.0 (December 2023) includes an anatomical structure systems graph which groups major organs into organ systems (for example, digestive system and

reproductive system); three ASCT+B tables that represent the branching structures for the blood and lymph and the peripheral nervous system; and 29 ASCT+B tables that document the nested 'part of' structure of other organs (for example, kidney with cells that compose smaller and the subsequently large FTUs and organ parts) for a total of 33 ASCT+B tables. The cells that make up each of the anatomical structures are organized in a multilevel cell type typology, with 'cell' at the root and successively more specialized child nodes. Cells are mapped to five biomarker types: genes, proteins, metabolites, proteoforms and lipids organized in a biomarker typology.

Anatomically based 3D reference objects (Fig. 1b) in the HRA include the shape, size, location and rotation of 1,192 3D anatomical structures with 516 unique ontology IDs in 65 organs. A SPARQL query (<https://apps.humanatlas.io/api/grlc/ccf.html#get-/as-3d-counts>) returns all anatomical structures with an Uberon ID (it retrieves the 1,192 anatomical structures plus the 65 organs for a total of 1,257 items). 2D references (Fig. 1c) describe the spatial layout of 3,742 rendered 2D cells of 116 unique cell types for 22 FTUs in 10 organs. Labeled training data for spatial segmentation and machine-learning models (Fig. 1d) exist for five FTUs in five organs. A total of 13 OMAPs (Fig. 1e) are linked to 197 AVRs and aligned with ASCT+B tables. Cell-type annotation tools (Fig. 1f) include Azimuth and other references for healthy adult organs with crosswalks to cell types and biomarkers in the ASCT+B tables.

An important part of HRA processing is data enrichment. One example is HRApop (Fig. 1g), which covers 553 tissue datasets that are used to compute cell type populations for 40 anatomical structures for which 3D reference objects exist, across 23 organs with 13 unique Uberon IDs. The code to reproduce the bar graph with HRApop data (seven datasets) is available³¹. HRAlit³² (Fig. 1h) links HRA DOs to 7,103,180 publications, 583,117 authors, 896,680 funded projects and 1,816 experimental datasets.

Data enrichment. This HRA processing step ensures that HRA DOs are high quality, usable and useful for the user stories listed in Table 1 and other applications. Normalization ensures that the raw data are well structured and presented in a format that can be readily translated into a knowledge graph via LinkML (<https://linkml.io>). During enrichment, certain implicit relationships are made explicit using OWL reasoning (for example, transitive relationships like subclass and 'part of' are made explicit); external metadata are added from ontologies via APIs to enhance the graph's usefulness (for example, via queries to the scicrunch API to look up antibody information for OMAPs); queries are used to add data from related graphs (for example, extracting additional metadata and hierarchies related to anatomical structures, cell types and biomarkers from popular biomedical ontologies like Uberon and Cell Ontology); and finalizes conversion from LinkML to knowledge graph (for example, converting and combining all into an RDF-formatted graph in Turtle format).

Data publication. A new revised and extended version of the HRA DOs together with updated user interfaces and APIs are published every 6 months via the HRA Portal (<https://humanatlas.io>). The three HRA core ontologies (specimen, biological structure and spatial ontologies)⁷ are shared as FAIR, versioned LOD at <https://lod.humanatlas.io>. Select data are also provided in a relational database and as comma-separated value (CSV) files. RUI data are published via the HuBMAP, SenNet, GUDMAP, GTeX and other portals. For instance, the HuBMAP Search API is queried by the HRA API to generate dataset graphs from HuBMAP data. The public graph with all donor, tissue block, tissue section, RUI data and experimental dataset information can be accessed via the HRA dataset graph at <https://lod.humanatlas.io/ds-graph>.

The HRA DO processor (<https://github.com/hubmapconsortium/hra-do-processor>) supports automated processing of HRA data, including data normalization, validation, graph transformation, enrichment

Table 2 | Ontologies used and extended

DO type	Ontology name and version	No. terms added/updated	No. links added/updated
Specimen			
	LOINC v.2022-07-11 (v2022AB)		
Biological structure			
Anatomical structures	Uberon v.2024-09-01	125 new terms added to Uberon through Oct 2024	609 relationships added to Uberon through Oct 2024 294 Uberon terms assigned 3D reference objects ^a
Cell types	CL v2024-09-26 PCL 2024-07-12	141 ^b new terms added to CL through Oct 2024 468 new terms added to PCL through Oct 2024	302 relationships added to CL through Oct 2024 468 relationships added to PCL through Oct 2024
Biomarkers	HGNC v2023-09-18 Ensembl Release 111 UniProt Release 2024_1 GeneCards v.5.19: 15 January 2024	None	None
Human genome	HGNC v.2023-09-18	None	None
Spatial			
Terms	DCTERMS v.2020-01-20	None	None
Relations	Relation Ontology used by Open Biological and Biomedical Ontologies	None	None
BioAssay and Experimental Factor types			
BioAssay Experimental Factor	BAO v.2023-01-31 EFO v.2023-02-15	None	None

^aThe foaf:depiction annotation is used to link 294 unique Uberon terms to 586 GL transmission binary (GLB) files. ^b25 out of 129 terms and 37 out of 249 links were added to CL supported by projects other than HuBMAP.

and publishing. The end product is the HRA KG (<https://github.com/hubmapconsortium/hra-kg>) and a set of flat files suitable for hosting all data as LOD. HRA infrastructure is optimized for deployment to Amazon S3, Amazon Web Services (AWS) AppRunner and AWS CloudFront, but could be adapted to other file hosting platforms.

The HRA provenance graph keeps track of all HRA DOs (described using standard terminology from DCAT (<https://www.w3.org/TR/vocab-dcat>) for organizing catalogs of data and W3C-Prov (<https://www.w3.org/TR/prov-overview>) for describing the provenance of any particular piece of data) and code versions (via GitHub) so HRA KG provenance can be accessed and the HRA KG can be recomputed every 6 months.

Supplementary Table 2 lists all data used in HuBMAP Data Portal (H), HRA Portal (A) and demonstration Previews (P). Note that HRA data are mirrored by the European Bioinformatics Institute's (EBI's) Ontology Lookup Service (OLS), Stanford University's NCBO BioPortal and University of Michigan Medical School Ontobee. Publishing the HRA via widely used repositories for biomedical ontologies makes the HRA FAIR; users can browse the HRA data online or access it programmatically via APIs.

Usage and extension of ontologies. Data and workflows are linked to existing ontologies whenever possible (Table 2). The sixth release of the HRA v.2.0 uses biological structure ontologies Uberon 2023-10-27 (ref. 9) and FMA v.5.0.0 (refs. 10,11) for anatomical structures; Cell Ontology (CL) v.2023-10-19 (ref. 12) and PCL 2023-02-27 (ref. 13) (<https://www.ebi.ac.uk/ols4/ontologies/pcl>) for cell types; HGNC v.2023-09-18 (ref. 33), Ensembl Release 111 (ref. 34), GeneCards v.5.19: 15 January 2024 (refs. 35,36) and UniProt Release 2024_1 (ref. 37) for biomarkers. The Human Genome HGNC v.2023-09-18 is used for the FTU Explorer. Spatial data are annotated using Dublin core terms (DCTERMS) v.2020-01-20 (<https://www.dublincore.org>). Specimen data use LOINC v.2022-07-11 (v.2022AB)³⁸ for standardized representation of sex, race and ethnicity data. Meta-ontologies such as DCTERMS and Relation Ontology³⁹ (RO) are used to capture relationships among

concepts within the HRA data. Assay type names come from BioAssay Ontology (BAO) v.2023-01-31 (ref. 40) and Experimental Factor Ontology (EFO) v.2023-02-15 (ref. 41). The use of these ontologies is strongly encouraged to maintain consistency among ASCT+B tables, Azimuth and other CTann tools and OMAP data in support of atlas construction and usage.

A major contribution of the cross-consortium HRA effort is the extension of cross-species ontologies such as Uberon and CL to cover healthy human terms. Between 2021 and October 2024, 125 anatomical structure terms have been added to Uberon, 141 cell types were added to Cell Ontology. By October 2024, 468 cell types were added to the PCL, 461 for the human brain⁴² (in support of HRA construction and usage). PCL uses computationally derived marker genes from NS-Forest⁴³ to define sc/snRNA-seq-derived cell types in the brain. The 461 human brain cell types were added to the ASCT+B tables. All PCL cell type terms are associated with biomarker genes using a has_characterizing_marker relation in the ontology. In the sixth release of the HRA, there are 962 anatomical structure terms that are either missing from Uberon or not yet crosswalked to Uberon terms in the ASCT+B table. The majority of missing terms are for blood and lymph vasculature, skeleton or skeletal muscle systems and are typically more specific than currently represented in Uberon (for example, 'dorsal branch of lateral proper palmar digital artery of fifth digit of hand'). Work is ongoing to improve mappings (~100 mappings were recently added and will be published in seventh HRA release). A total of 119 cell types are either unmapped or not yet in CL or PCL (an initial assessment suggests 60) and 70% of these are genuinely new terms for CL. These 387 biomarkers have Ensembl IDs or GeneCards IDs or have not been mapped rather than HGNC IDs (all of these terms have ASCTB-TEMP IDs). There are GitHub issues to add new terms to existing ontologies to properly represent data in the ASCT+B tables, including requests for 128 anatomical structures in Uberon. There now exists a formal operating procedure to include new cell types into CL via Minimal Information Reporting About a Cell (MIRACL) sheets⁴⁴. The number of ontology relationships added to Uberon, CL and PCL is listed in the last column in Table 2.

Unified processing workflows. The HRA SOPs⁴⁵ detail the human expert and algorithmic steps needed to construct the HRA and to use it properly. Protocols published on protocols.io and other places are used to compile experimental data in a reproducible manner. In January 2024, there existed 235 HuBMAP protocols⁴⁶ (many of these document the reproducible workflows required to generate data used in HRA construction). Figures 1 and 2 provide an overview of the numerous steps required to construct the HRA and to map new experimental data to it.

The HuBMAP Consortium has developed uniform computational processing pipelines for multiple data types: single-cell (sc)/single-nucleus (sn) RNA-seq, sc/snATAC-seq, multiplexed antibody-based spatial proteomics (CODEX (recently renamed to PhenoCycler) and Cell DIVE), multiplexed ion beam imaging (MIBI), Slide-seq and Visium sequencing spatial transcriptomics and fluorescence in situ hybridization spatial transcriptomics, among others. HuBMAP computational pipelines are all open source, published on GitHub as CWL workflows wrapping tools in Docker images (also executable via Singularity), with supplementary data (genome indexes/annotations and deep-learning models) built into the published Docker images for portability and reproducibility.

The HuBMAP sc/snRNA-seq pipeline (<https://github.com/hubmapconsortium/salmon-rnaseq>, also used for sequencing spatial transcriptomics such as Slide-seq and Visium), is built on the Salmon quasi-mapping method⁴⁷ and performs gene expression quantification for intronic and exonic sequences, with downstream analysis using Scanpy⁴⁸ and RNA velocity computation via scVelo⁴⁹. Outputs of the sc/snRNA-seq pipeline are annotated with an automated version of the Azimuth cell type annotation tool for supported tissues; these currently include heart, lung and kidney, with additional annotations computed as new Azimuth references are integrated into HuBMAP processing infrastructure.

HuBMAP imaging pipelines (Methods) are end-to-end analysis methods that accept raw images, perform illumination correction, background subtraction and tile stitching if necessary, then perform cell and nucleus segmentation, writing expression and segmentation mask images as multichannel OME-TIFF files. The expression and mask images are further processed via spatial process and relationship modeling (SPRM)⁵⁰, which computes image and segmentation quality metrics using the CellSegmentationEvaluator tool^{51,52}, creates cell adjacency maps, computes features for each cell and nucleus, performs unsupervised clustering of cells, nuclei and image pixels, computes biomarkers differentiating one cluster versus the rest for each clustering type and writes results to CSV and HDF5 format for use by end users and in the HuBMAP Data Portal.

For HRApop (Fig. 1g), 445 public datasets from HuBMAP^{2,53}, two datasets from SenNet⁵⁴, 91 healthy datasets from two collections from CZ CELLxGENE^{55,56} ('Cells of the adult human heart' and 'LungMAP – human data from a broad age healthy donor group') and 15 single-cell datasets from GTEx^{57,58} were mapped to the HRA (Methods). As a result, cell-type population data exist for 40 anatomical structures in 23

organs with 13 unique Uberon IDs, separated by single-cell transcriptomics (for example, sc/snRNA-seq) and OMAP-aligned spatial proteomics (for example, CODEX and Cell DIVE). Three organs (large intestine, small intestine and skin) have cell type populations computed from transcriptomics and proteomics data.

For HRAlit³² (Fig. 1h), 583,117 experts, 7,103,180 publications, 896,680 funded projects and 1,816 experimental datasets were mapped to the DOs in the HRA (Methods).

User interfaces. The HuBMAP Portal (<https://hubmapconsortium.org>; Supplementary Fig. 2) introduces HuBMAP goals and links to experimental and atlas data, tools and training materials. The HuBMAP Data Portal (<https://portal.hubmapconsortium.org>) supports ingest, search, exploration and download of experimental data. The HRA Portal (<https://humanatlas.io>; Supplementary Fig. 3) supports the construction, access, exploration, usage and download of HRA data.

The ASCT+B Reporter³ (<https://humanatlas.io/asctb-reporter>; Supplementary Fig. 4) supports the authoring and review of ASCT+B Tables and OMAPs by human organ experts. Detailed SOPs⁴⁵ and video tutorials^{59,60} exist and more than 170 unique experts have contributed to the HRA as authors and/or reviewers using this tool as measured by the number of unique ORCID IDs listed in relevant DOs of the sixth release HRA.

Azimuth¹⁴ (<https://azimuth.hubmapconsortium.org>; Supplementary Fig. 5) was developed by HuBMAP to automate the processing, analysis and interpretation of sc/snRNA-seq and ATAC-seq data. Its reference-based mapping pipeline reads a cell-by-gene matrix and performs normalization, visualization, cell annotation and differential expression (biomarker discovery) analyses (Figs. 1f and 2b). Results can be explored within the app or downloaded for additional analysis. In HuBMAP, Azimuth is used in production mode to automatically annotate sc/snRNA-seq datasets. Crosswalks exist to associate Azimuth cell types to ASCT+B table terms and ontology IDs.

The RUI⁶⁰ (<https://apps.humanatlas.io/rui>; Supplementary Fig. 6 and SOP⁶¹) supports the registration of human tissue blocks into the 3D CCF with automatic assignment of anatomical structure annotations that are linked to the Uberon and FMA ontologies based on surface mesh-level collision events. The anatomical structure annotations in combination with ASCT+B table and experimental data make it possible to predict cell types that are commonly found in anatomical structures and colliding tissue blocks. RUI output in JSON format records registration data (for example, tissue block universal unique identifier (UUID) and 3D size, location and rotation plus anatomical structure annotation based on bounding box) together with provenance data (for example, operator name and date). The RUI is available as a stand-alone tool for anyone to use to contribute HRA-aligned spatial data. It is fully integrated in the HuBMAP, SenNet and GUDMAP data ingest portals but requires authentication.

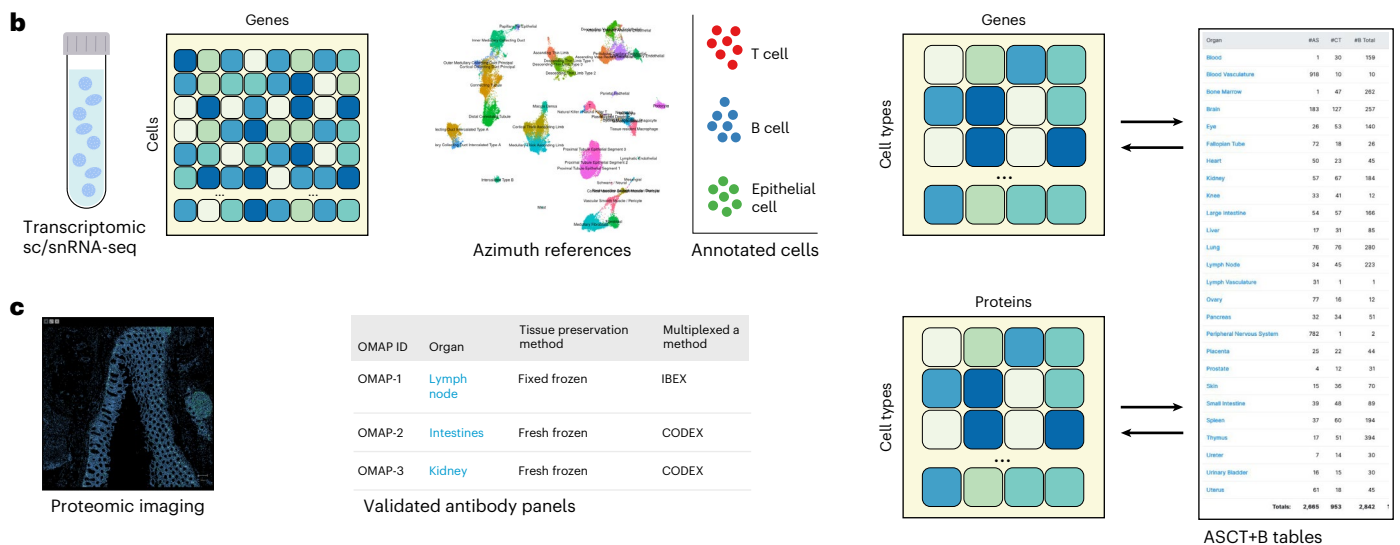
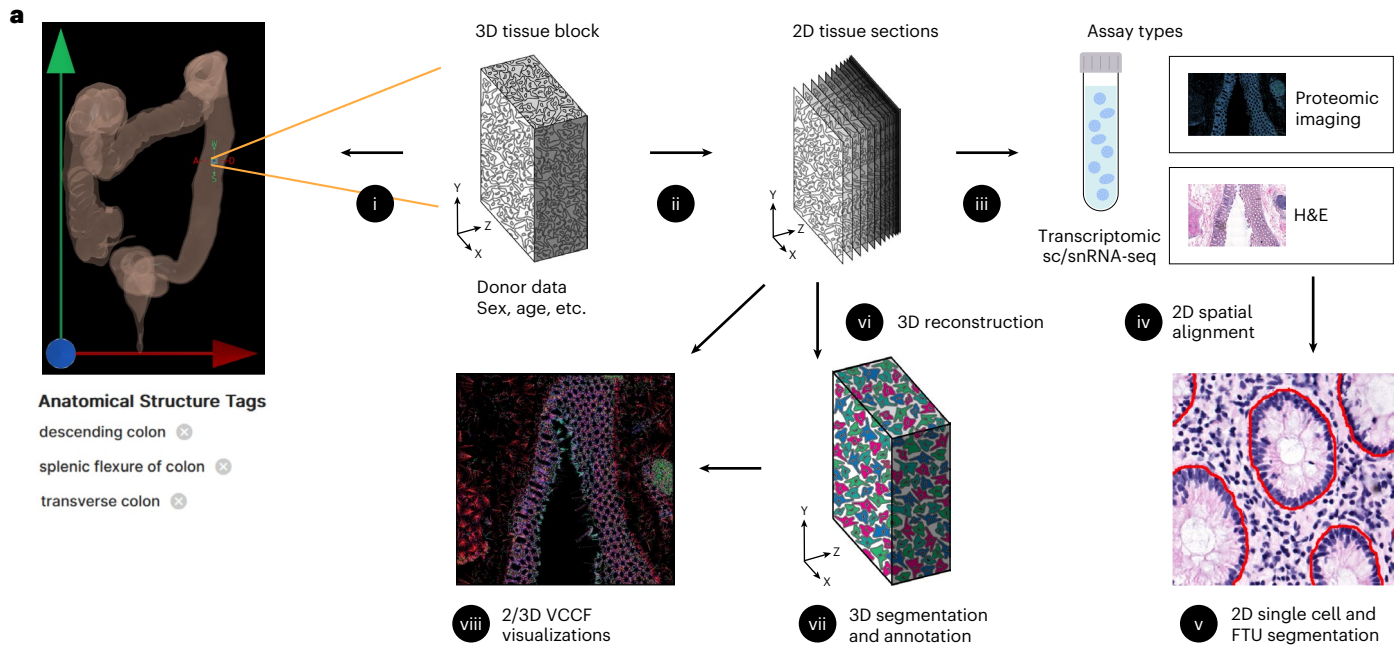
The Exploration User Interface (EUI) (<https://apps.humanatlas.io/eui>; Supplementary Fig. 7) supports visual browsing of tissue samples

Fig. 2 | Mapping experimental data to the HRA. **a**, A tissue block is 3D spatially registered and semantically annotated using the RUI or millitome (i). A smaller part of the tissue block might be used for sc/snRNA-seq analysis (not shown) or cut into tissue sections (ii). Tissue sections are analyzed using one or more assay types (iii). Shown are single-cell transcriptomics (for example, sc/snRNA-seq), OMAP-aligned spatial proteomics (for example, CODEX and Cell DIVE) and high-resolution hematoxylin and eosin (H&E)-stained histology images. Spatial alignment of different assay types for the very same or different tissue sections is non-trivial (iv). H&E data are used to segment FTUs using trained machine-learning models (v). A 3D reconstruction of tissue volumes is accomplished by aligning data from multiple serial tissue sections computationally (vi) followed by 3D segmentation and annotation (vii). The 2D or 3D data are analyzed to identify the distance of different cell types to the vasculature (VCCF visualizations) as a multiscale CCF from which no other cell is very distant (viii).

b, Single-cell/nucleus data (sc/snRNA-seq) is stored as a cell-by-gene matrix; cell types are annotated using Azimuth or other cell type annotation tools; results are aggregated to cell-type-by-gene biomarker expression value matrices that are aligned with the ASCT+B tables; and are used in diverse HRA user interfaces (for example, EUI and FTU Explorer). **c**, OMAP-aligned spatial data generated using validated antibody panels linked to AVRs are analyzed to compute cell type by protein biomarker expression value matrices that are aligned with the ASCT+B tables using semi-automated workflows. **d**, The EUI provides full provenance for donors (sex, age and body mass index), data providers (upload date, contact name and affiliation), tissue blocks and sections (size, number, date and contact info for RUI registration) and datasets (assay type) with links to raw data in the HuBMAP Data Portal, other data portals or publications. **e**, CWL workflows detail which tools (yellow) are run on which input/output data (blue). Shown is the Azimuth cell type annotation workflow.

and metadata at the whole-body organ, tissue and cell levels (Table 1, US 3). In January 2024, 901 human tissue blocks with 4,221 datasets from 351 donors and 19 consortia/studies were RUI-registered into the HRA 3D CCF. Users can filter by donor demographics (for example, sex and age) or data source (for example, consortium/study and technology). They can search for specific anatomical structures, cell

types or biomarkers to explore the number of tissue blocks that collide with an anatomical structure but also the cell types located in these anatomical structures or their characterizing biomarkers (according to the ASCT+B tables). Users can also run a 3D spatial search using an adjustable probing sphere, explore details on demand on the right with links to Vitessce^{62,63} visualizations in the HuBMAP Data Portal



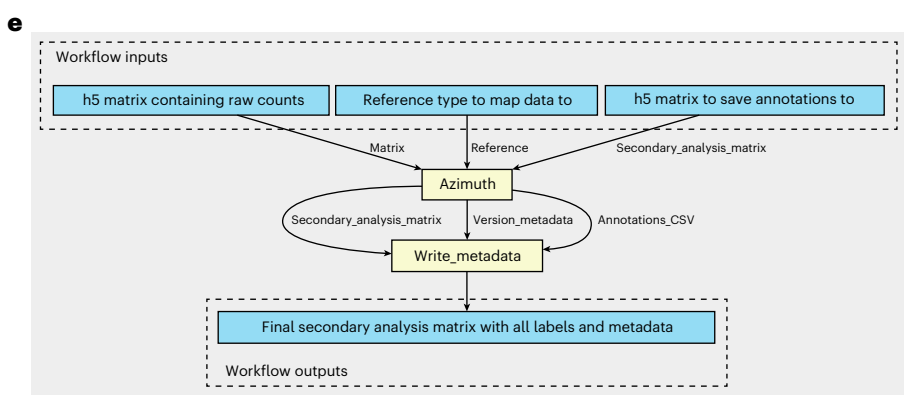
Male, Age 78, BMI 35.1
Entered 2/19/2020, Yiing Lin, TMC-Stanford

Registered 2/17/2022, Yiing Lin, TMC-Stanford
3 × 30 × 15 millimeter, 7.5 millimeter, 2 Sections

Registered 2/16/2022, daniel cotter, TMC-...
3 × 30 × 7.5 millimeter, 7.5 millimeter, block

Registered 9/7/2021, daniel cotter, TMC-St...
3 × 30 × 7.5 millimeter, 7.5 millimeter, block

Proteomic imaging



and links to data and tools in other data portals. The EUI with all HRA data is available as a stand-alone tool that supports exploration of all experimental data that has been mapped to the HRA. The EUI was customized, branded and fully integrated in the HuBMAP, SenNet and GTEx data portals to support exploration of consortia specific data (Supplementary Fig. 8).

Vitesse^{62,63} (<http://vitesse.io>) is a tool used to visually explore experimental data, Azimuth references (Supplementary Fig. 5), HRA segmentations and annotations or cell–cell distance distribution visualizations (Supplementary Fig. 9), see previews in the Atlas usage section.

The Interactive FTU Explorer⁶⁴ (<https://apps.humanatlas.io/ftu-explorer>; Supplementary Fig. 10) supports the exploration of cell types in their 2D spatial context together with mean biomarker expression matrices (Table 1, US 4). For example, tissue data (cell type populations with gene or protein expression levels, as available) can be compared against healthy HRA reference data to determine differences in the number of cells, cell types or mean biomarker expression values to inform clinical decision-making.

The HRA Organ Gallery^{65,66} (<https://github.com/cns-iu/hra-organ-gallery-in-vr>; Supplementary Fig. 11) supports the multiscale exploration of 1,192 anatomical structures in the 65 3D Reference Objects of the HRA 2.0. Using a Meta Quest VR device, users select the male or female reference body; they can then select a specific organ and explore it with both hands. To achieve view update rates of 60 frames per second, lower level-of-detail models are used that were derived from the original HRA 3D Reference Objects.

The HRA API (<https://humanatlas.io/api/>; Supplementary Figs. 12–14) supports programmatic access to all HRA DOs and the experimental HRApop data mapped into it. Users first select an API server and route, input query parameters, then view the query response, see Methods for details.

The HRA dashboard (<https://apps.humanatlas.io/dashboard>) compares HRA, publication and experimental data to world population data. Supplementary Fig. 15a shows population pyramids by age group of HRA survey respondents and tissue data donors in comparison to world population plus population pyramids by career age for HRA experts and publication authors. Supplementary Fig. 15b features the ethnic composition of survey respondents, HRA tissue donors, HRA experts, paper authors and world population in percentages. The choropleth map in Supplementary Fig. 15c shows the number of paper authors overlaid on a world map. CCF–HRA data dashboards help understand what HuBMAP data have been RUI registered (<https://hubmapconsortium.github.io/hra-data-dashboard>).

Documentation and instructional material. In January 2024, the HuBMAP Data Portal provides access to 8 publications and associated datasets, 50+ technical documents (<https://software.docs.hubmap-consortium.org/technical>) and links to 235 experimental protocols on protocols.io; the HRA Portal links to 20 SOPs (<https://zenodo.org/communities/hra>) and to the Visible Human Massive Open Online Course (VHMOOC; <https://expand.iu.edu/browse/sice/cns/courses/hubmap-visible-human-mooc>) with 39 videos, 4 self-tests and 3 quizzes, 2 hands-on tutorials, plus entrance and exit surveys (Supplementary Fig. 16).

Previews of Atlas usage

Two exemplary previews demonstrate the usage of atlas data and code developed in HuBMAP for gaining insights into pathology, see user stories that drive HRA construction and usage (Table 1). All data and code are publicly available on GitHub^{67,68} and Dryad⁶⁹. The cell distance distribution code is available via the HRA Portal⁷⁰ in support of Table 1, US 5. Cell type annotations for the CODEX multiplexed imaging dataset of the human intestine are published via Dryad. Full data and code integration into the HuBMAP Data Portal workflows are planned for future releases.

Perivascular immune cells in lung. Normal lung function depends on careful matching of airflow to blood flow to achieve normal gas exchange. The abnormal presence and activity of immune cells results in leaky vascular membranes and edema that thickens the gas exchanging membrane and accumulation of mucus and cellular debris in the airspace can cause a mismatch between flow of air and blood. Persistent inflammation results ultimately in fibrosis. Previous work using single-cell RNA sequencing data and the CellTypist common reference dataset discovered previously under-appreciated organ-specific features and aggregates of T cells and B cells⁷¹. Recent publications in the field of mucosal immunology illustrate the segregation of immune cells within aggregates in human lung tissue and their role in abnormal regulation of vascular function^{72–75}. Molecular and cellular changes, including fibrotic and immune-cell-rich regions were recently imaged in the lungs of children with bronchopulmonary dysplasia (BPD), a chronic lung disease following premature birth⁷⁶. The Vitesse tool is used in Fig. 3a (left) to visualize tissue data. Cell distance analyses and visualizations are used to comparatively visualize and quantify cellularity of specific regions of healthy adult and BPD lung to demonstrate an assessment of multiple cell types relative to nearest vascular endothelial cell nuclei using single-cell spatial protein biomarkers.

Demonstrated are whole-slide photomicroscopy images of PhenoCycler^R multiplexed immunofluorescence assays (WSI-MxI) applied to examples of healthy lung tissue (top row, 28 antibody panel) and BPD tissue (bottom row, 25 antibody panel). Digital zoom is used to highlight similar regions of interest (red box, MxI-ROI) focused on a bronchiole (br) and an accompanying small pulmonary artery (pa). Immune-cell aggregates, primarily CD3⁺ lymphocytes, are noted around both structures in the BPD lung. To assess a vascular CCF (VCCF) for the localization of the immune and other lung cells, the cell types are masked in cell specific colors (see key) and distances are measured to nearest endothelial cell nucleus (red circles). The cell to nearest endothelial cell measurements are seen as spokes colored by the cell type. In brief, the graphical representations quantitatively demonstrate increased cellularity with a predominance of CD4⁺, but not CD8⁺, lymphocytes, as well as myeloid immune cells, positioned in close proximity to the lung vasculature in the diseased lung. The VCCF visualizations suggest endothelial cells embedded within the lymphocytic aggregates that congregate around the pulmonary artery (pa) in the diseased as compared to the healthy lung. Analyses of cell populations (number of cells per cell type and their mean biomarker expression values), as well as cell to cell and cell to FTU spatial distribution patterns are valuable to understand tissue and cellular disruptions that account for organ failure in lung disease. In this example, the diseased tissue has gas exchange membranes thickened in part by extravascular immune cell aggregates such that distribution of cell distance to nearest endothelial cells is compressed and exaggerated (see graphs in right of Fig. 3a). Measurements on 2D images, as demonstrated, provide novel insights; however, cell segmentation and determination of relative locations in the lung is particularly challenging given the complex airway and vascular branching system and the very thin, highly multicellular gas exchange membranes of the alveoli. It is anticipated that application of similar segmentation, cell to cell and cell to FTU measures to 3D lung tissue volumes will identify currently under-appreciated relationships in health and disease (Table 1, US 5). HuBMAP code can be used on human lung tissue to understand how the spatial organization of specific immune cell infiltrates relate to disease pathophysiology revealing potentials for targeted therapy to ameliorate human disease. Code and data are available at ref. 68.

Hierarchical cell type populations within FTUs. FTU segmentation algorithms for histology data^{77,78} and hierarchical cell community analysis⁶⁹ for paired spatial data (see Methods for both) can be combined to analyze and communicate the structure and function of FTUs across scales. FTU segmentation algorithms are run as part of the standard

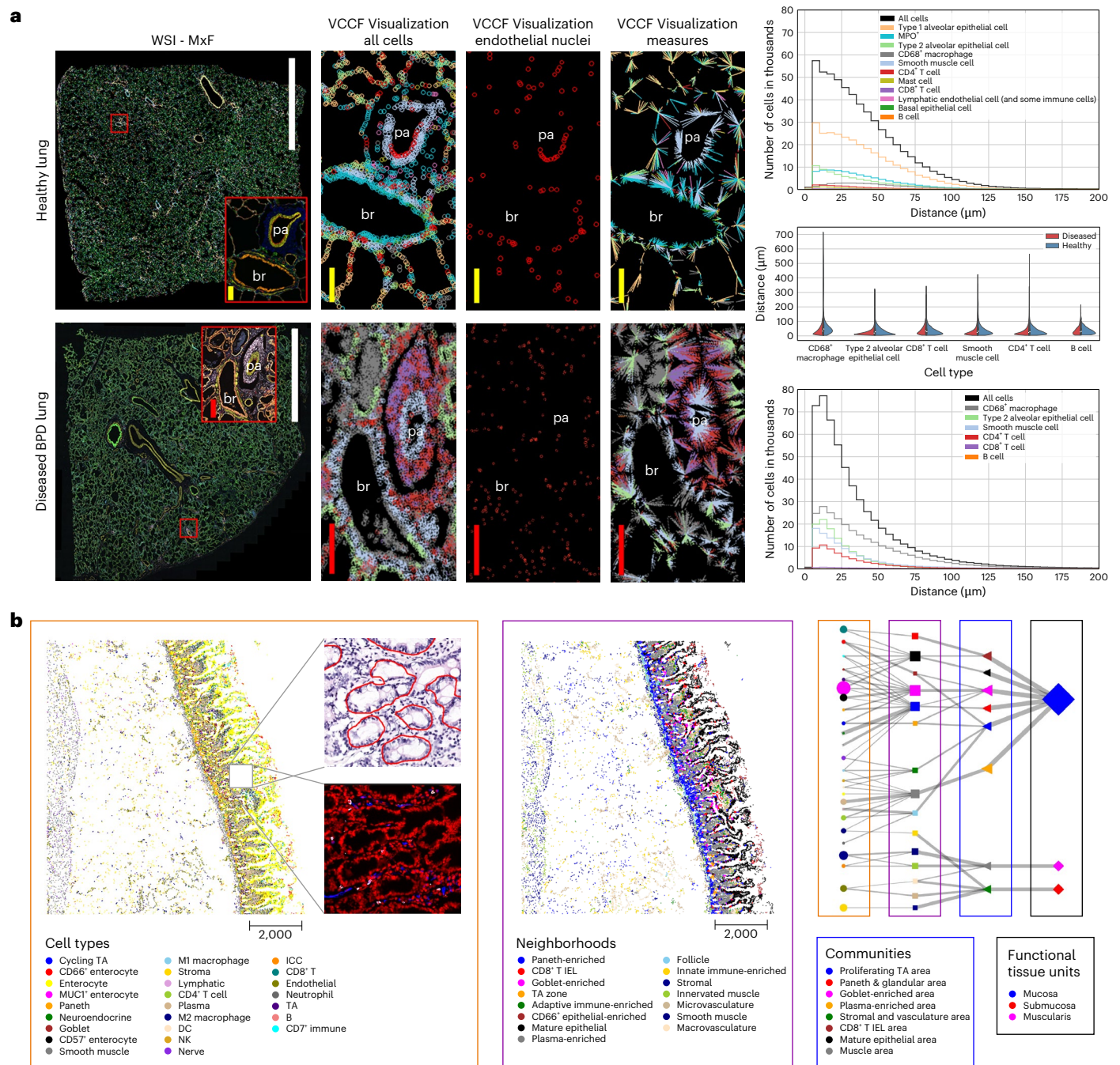


Fig. 3 | Human Reference Atlas usage. a, HRA can be used to compare the distribution of parenchymal cells including endothelial, epithelial and muscle that compose the blood vessels, airways and gas exchanging functional lung structures and resident immune cells including macrophages, to local vasculature (VCCF visualization) in healthy (top) and diseased (bottom) lung using multiplexed immunofluorescence microscopy images with bronchiole (br) and an accompanying small pulmonary artery (pa). Scale bars, white 5 mm; red 200 μ m; and yellow 100 μ m. The graphs on the right show distance distributions for cell types present in the healthy lung (top) and diseased BPD lung (bottom);

the violin plot (middle) shows a comparison between distance distributions for cell types common in both datasets. Datasets are on GitHub⁵¹. **b**, Multilevel cell neighborhoods can be computed to analyze and communicate the structure and function of FTUs; tissue image with cell type annotations and zoom into H&E with FTU segmentations (red outlines) and zoom into the multiplexed image (CODEX) is shown in left, neighborhoods are given in the middle; hierarchy of FTUs, neighborhoods, communities and cell types are shown on the right. Datasets are on GitHub⁵¹. ICC, Interstitial cells of Cajal; TA, transit-amplifying; NK, natural killer; DC, dendritic cell; IEL, intraepithelial lymphocytes.

HuBMAP workflows (currently limited to glomeruli in kidneys, soon to be expanded to crypts in the large intestine and white pulp in spleen).

Exemplarily, we feature an example hierarchical cell neighborhood analysis previously developed for analyzing cell type neighborhoods across scales and applied within the healthy human intestine⁶⁹ (see Jupyter Notebook on Github⁶⁷). We have named some of these scales: cellular ‘neighborhoods’, ‘communities’ and ‘functional tissue units.’

The calculation of similar cellular neighborhoods, communities and tissue units across different scales is analogous to how we might think that people form neighborhoods, cities and states.

Currently, the HuBMAP Data Portal supports cell segmentation of antibody-based multiplexed imaging data but lacks the ability to annotate the cell types for such datasets. This functionality is under active development (Methods). Consequently, to demonstrate this user

story, a separately processed version of the intestine data⁷⁹ (containing cell type annotations) was used. The cell type predictions for the same dataset, using the current development version of the cell type model by the Van Valen laboratory (Methods) are also made available via <https://cns-iu.github.io/hra-construction-usage-supporting-information>. This version of the model was trained on several multiplexed datasets spanning tissue types and multiplexed imaging modalities. Cell segmentation masks, generated using Mesmer, are also included with the prediction data.

The Jupyter Notebook on GitHub⁶⁷ demonstrates how to read a previously published CODEX multiplexed imaging dataset of the human intestine⁶⁹, identify how cell types correspond to larger multicellular structures and support exploration of the relationships between these higher order cellular neighborhoods. By visualizing the data in tissue coordinates, we can observe potential layering or consistent FTU structures, such as the repeat structure of the intestinal crypt in the proximal jejunum in the small intestine (see Fig. 3b, left). Furthermore, we can quantify these relationships across different scales of cellular neighborhoods and represent them as a network graph (Fig. 3b, right) in which line thickness indicates the percentage of cells in the next level. Note that the tissue samples and the graph use the very same node color coding and naming.

Usage statistics

Between July 2023 and October 2024, over 33,500 unique users visited the HuBMAP Consortium website (<https://hubmapconsortium.org>). These users visited 480 distinct pages; the top six most-frequent referrers were Google, pathwaystoscience.org, nature.com, Bing, X/Twitter and psc.edu. Between January 2021 and December 2023, 87,310 unique users visited 382,384 pages in the HuBMAP Data Portal (<https://portal.hubmapconsortium.org>); the top five most-frequent referrers were nature.com, hubmapconsortium.org, humanatlas.org, azimuth.hubmapconsortium.org and humanatlas.io. Azimuth supported the upload and cell type annotation of 27,000 datasets with more than 366,000,000 cells. Between June 2023 and October 2024, 1,194,130 HRA Portal requests and 524,358 HRA API requests were fulfilled; the top five referrers were the HuBMAP Entity API, the GTEx Portal (<https://gtexportal.org>), the HuBMAP Data Portal, the SenNet Data Portal (<https://data.sennetconsortium.org/search>) and EMBL-EBI (<https://www.ebi.ac.uk>). The 3D reference objects were accessed 3,065 times via the NCBI BioPortal Ontology Browser (<https://biportal.bioontology.org/ontologies/CCF>) and 11,531 times via the EBI OLS Ontology Browser (<https://www.ebi.ac.uk/ols/ontologies/ccf>). A total of 310 students registered for the VHMOOC and spent 5,652 h reviewing materials, taking self-tests and engaging in a community of practice.

Discussion

This Resource paper describes data, code and tools that are of broad utility, interest and significance to the construction and usage of a multiscale HRA. The HRA effort and evolving data and code infrastructure is novel and unique in several ways: (1) The HRA integrates many assay types across scales, from whole human body to single-cell level. (2) It provides SOPs and tools to spatially and semantically register human tissue from 65 organs into one CCF. (3) It links anatomical structures, cell types and biomarkers to ontologies and extends existing ontologies when needed. (4) The HRA comes with diverse interfaces that allow users to explore and inspect diverse HRA DOs (3D reference objects, ASCT+B tables, OMAPs, etc.), experimental data and documentation from the participating consortia, as well as HuBMAP data in particular (HRA Portal, ASCT+B reporter, RUI, EUI, Cell-Cell Distance Distribution Visualizations, Interactive FTU Explorer and HRA Organ Gallery in VR). For each user interface, we provide Supplementary Figs. 3, 4, 6, 7 and 9–11) with high-resolution screenshots and detailed annotations. (5) HRA development is community driven and collaborative; monthly

WG meetings inform strategic decision-making; 50+ open-source algorithms developed by 30+ teams have been systematically integrated into a flexible and adaptive system architecture that adds value to many atlasing projects; new HRA data and code are publicly released every six months via the HRA Portal and ontology services. The resulting HRA is a multiscale, multimodal, 3D digital product that unifies biomedical knowledge across organs, anatomical structures scales, demographic markers, assay types and links them to ontologies and makes human reference data computable.

The sixth release HRA has several known limitations that will be addressed in future iterations. Starting with the eighth release (published in mid-December 2024), all HRA DOs and their complete provenance are covered in the HRA KG. Cell states are not currently captured in CL nor are ever specific cell types emerging from single-cell technologies; however, the HRA started to use cell type annotation that have the format 'CL ontology term:cell state or specificity terminology' (for example, 'pancreatic stellate cell:quiescent' mapped to 'pancreatic stellate cell CL:0002410', which is in CL or 'enterocyte: MUC1 positive' mapped to 'enterocyte CL:0000584') along with a confidence term to the CL match (for example, skos:narrowMatch (for cell states or new cell types) or skos:exactMatch (those that exactly match a cell type in CL)) to allow it to be represented in the HRA KG and UBKG and updated at a later date when the computational community has settled on a method to ontologically represent such cells. When there is no exact match, terms in the HRA KG will be given an ASCTB-TEMP ID and individual cells annotated by a cell type annotation model will be given a cell type ID to facilitate future updates when they become available.

In addition, existing workflows for mapping new experimental data to the HRA will be expanded in three main ways: (1) HuBMAP plans to add several new Azimuth references (for example, for large and small intestine) and update existing references (for example, for kidney and lung) to capture new data with additional/revised cell type annotations and CL terms plus CL IDs via crosswalks; (2) eight new OMAPs were published in the seventh HRA release and several more are in progress for the eighth HRA release, substantially increasing the number of spatial datasets that can be mapped to the HRA; and (3) starting with the eighth HRA release, new 3D organs will be added to the RUI: quadriceps femoris and triceps surae skeletal muscles, esophagus and lymphatic vasculature.

Currently, the HRA knowledge graph and API drive different 2D and 3D user interfaces in the HuBMAP, SenNet, GUDMAP, GTEx data portals and the CZ CellGuide. In line with US 6 (see Table 1), we started to develop additional lightweight web components that make it easy to access HRA data and feature HRA functionality in other websites (<https://apps.humanatlas.io/us6>). Plus, we are implementing diverse HRA dashboards (US 7; <https://apps.humanatlas.io/dashboard>) to communicate what HRA DOs exist; what experimental data is used for HRA construction (full provenance); how existing ontologies are expanded to capture healthy human terms and linkages; who is using HRA data, tools and APIs; and how representative the atlas is.

Last but not least, we will expand the interlinking of the HuBMAP Data Portal and the HRA Portal. Specifically, we will ingest new releases of the HRA into the HuBMAP UBKG so that anatomical structures, cell types and biomarkers supported in the HuBMAP Data Portal are aligned with existing ontologies and the 3D spatial reference framework. As HuBMAP teams start to compile 3D datasets, there is a need to compare existing algorithms for spatial alignment of multiple subsequent tissue sections in support of 3D tissue block reconstruction, as was done for 2D cell-segmentation methods^{51,52}. 3D data are expected to considerably improve HRA quality and predictions.

Community input to HRA user stories, data, code, user interfaces and training materials is welcome and experts interested to learn more about or contribute to the HRA effort are encouraged to register for the monthly WG events online⁸⁰.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-024-02563-5>.

References

- HuBMAP Consortium et al. The human body at cellular resolution: the NIH Human BioMolecular Atlas Program. *Nature* **574**, 187–192 (2019).
- Jain, S. et al. Advances and prospects for the Human BioMolecular Atlas Program (HuBMAP). *Nat. Cell Biol.* **25**, 1089–1100 (2023).
- Börner, K. et al. Anatomical structures, cell types and biomarkers of the Human Reference Atlas. *Nat. Cell Biol.* **23**, 1117–1128 (2021).
- Hunter, P. et al. A vision and strategy for the virtual physiological human: 2012 update. *Interface Focus* **3**, 20130004 (2013).
- Rood, J. E. et al. Toward a common coordinate framework for the human body. *Cell* **179**, 1455–1467 (2019).
- Kleven, H. et al. Waxholm Space atlas of the rat brain: a 3D atlas supporting data analysis and integration. *Nat. Methods* **20**, 1822–1829 (2023).
- Herr II, B. W. et al. Specimen, biological structure, and spatial ontologies in support of a Human Reference Atlas. *Sci. Data* **10**, 171 (2023).
- Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041 (2017).
- Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E. & Haendel, M. A. Uberon, an integrative multi-species anatomy ontology. *Genome Biol* **13**, R5 (2012).
- Golbreich, C., Grosjean, J. & Darmoni, S. J. The foundational model of anatomy in OWL 2 and its use. *Artif. Intell. Med.* **57**, 119–132 (2013).
- Rosse, C. & Mejino, J. L. V. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J. Biomed. Inform.* **36**, 478–500 (2003).
- Meehan, T. F. et al. Logical development of the cell ontology. *BMC Bioinform.* **12**, 6 (2011).
- Tan, S. Z. K. et al. Brain data standards: a method for building data-driven cell-type ontologies. *Sci. Data* **10**, 50 (2023).
- Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
- Quardokus, E. M. et al. Organ mapping antibody panels: a community resource for standardized multiplexed tissue imaging. *Nat. Methods* **20**, 1174–1178 (2023).
- Lin, D. et al. The TRUST principles for digital repositories. *Sci. Data* **7**, 144 (2020).
- Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
- Cyberinfrastructure for Network Science Center. Human Reference Atlas SOP Glossary. *Zenodo* <https://doi.org/10.5281/zenodo.14653033> (2025).
- Lonsdale, J. et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
- McMahon, A. P. et al. GUDMAP: the genitourinary developmental molecular anatomy project. *J. Am. Soc. Nephrol.* **19**, 667 (2008).
- Himmelstein, D. S. et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife* **6**, e26726 (2017).
- El-Achkar, T. M. et al. A multimodal and integrated approach to interrogate human kidney biopsies with rigor and reproducibility: guidelines from the Kidney Precision Medicine Project. *Physiol. Genomics* **53**, 1–11 (2021).
- Ardini-Poleske, M. E. et al. LungMAP: The molecular atlas of lung development program. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **313**, L733–L740 (2017).
- Gaddis, N. et al. LungMAP Portal Ecosystem: systems-level exploration of the lung. *Am. J. Respir. Cell Mol. Biol.* **70**, 129–139 (2024).
- Devor, A. et al. The challenge of connecting the dots in the B.R.A.I.N. *Neuron* **80**, 270–274 (2013).
- Hawrylycz, M. et al. A guide to the BRAIN Initiative Cell Census Network data ecosystem. *PLoS Biol.* **21**, e3002133 (2023).
- SenNet Consortium et al. NIH SenNet Consortium to map senescent cells throughout the human lifespan to understand physiological health. *Nat. Aging* **2**, 1090–1100 (2022).
- Rozenblatt-Rosen, O., Stubbington, M. J. T., Regev, A. & Teichmann, S. A. The Human Cell Atlas: from vision to reality. *Nature* **550**, 451–453 (2017).
- Srivastava, S. et al. The making of a precancer atlas: promises, challenges, and opportunities. *Trends Cancer* **4**, 523–536 (2018).
- Oxburgh, L. et al. (Re)Building a kidney. *J. Am. Soc. Nephrol.* **28**, 1370–1378 (2017).
- Cyberinfrastructure for Network Science Center. HRA-construction-usage-supporting-information/hra_pop. *GitHub* https://github.com/cns-iu/hra-construction-usage-supporting-information/blob/main/hra_pop/figure_1_g_hra_pop.ipynb (2024).
- Kong, Y. & Börner, K. Publication, funding, and experimental data in support of Human Reference Atlas construction and usage. *Sci. Data* **11**, 574 (2024).
- Seal, R. L. et al. Genenames.org: the HGNC resources in 2023. *Nucleic Acids Res.* **51**, D1003–D1009 (2023).
- Martin, F. J. et al. Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).
- Stelzer, G. et al. The genecards suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinform.* **54**, 1.30.1–1.30.33 (2016).
- Barshir, R. et al. GeneCaRNA: a comprehensive gene-centric database of human non-coding RNAs in the GeneCards Suite. *J. Mol. Biol.* **433**, 166913 (2021).
- The UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
- Regenstrief Institute. LOINC: the international standard for identifying health measurements, observations, and documents. *LOINC* <https://loinc.org/> (2024).
- Huntley, R. P. et al. A method for increasing expressivity of Gene Ontology annotations using a compositional approach. *BMC Bioinform.* **15**, 155 (2014).
- Visser, U. et al. BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results. *BMC Bioinform.* **12**, 257 (2011).
- Malone, J. et al. Modeling sample variables with an experimental factor ontology. *Bioinformatics* **26**, 1112–1118 (2010).
- Siletti, K. et al. Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).
- Aevermann, B. et al. A machine learning method for the discovery of minimum marker gene combinations for cell type identification from single-cell RNA sequencing. *Cold Spring Harb. Lab. Press* <https://doi.org/10.1101/gr.275569.121> (2021).
- Lubiana, T. et al. Guidelines for reporting cell types: the MIRACL standard. Preprint at <https://arxiv.org/abs/2204.09673> (2022).
- Cyberinfrastructure for Network Science Center. Human reference atlas standard operating procedures <https://humanatlas.io/standard-operating-procedures> (2023).
- Human BioMolecular Atlas Program (HuBMAP) method development community <https://www.protocols.io/workspaces/human-biomolecular-atlas-program-hubmap-method-development> (2019).
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

48. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
49. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
50. HuBMAP Consortium. sprm. *GitHub* <https://github.com/hubmapconsortium/sprm> (2023).
51. Chen, H. & Murphy, R. F. Evaluation of cell segmentation methods without reference segmentations. *Mol. Biol. Cell* **34**, ar50 (2023).
52. Murphy Group. CellSegmentationEvaluator. *GitHub* <https://github.com/murphygroup/CellSegmentationEvaluator> (2024).
53. Human BioMolecular Atlas Program. *HuBMAP Data Portal* <https://portal.hubmapconsortium.org/> (2022).
54. SenNet Consortium. Home - SenNet <https://sennetconsortium.org/> (2021).
55. Chan Zuckerberg Initiative. Chan Zuckerberg CELLxGENE Discover. *Cellxgene Data Portal* <https://cellxgene.cziscience.com/> (2022).
56. CZI Single-Cell Biology Program et al. CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. *Nucleic Acids Res.* **53**, D886–D900 (2025).
57. Genotype-Tissue Expression project. GTEx Portal <https://gtexportal.org/home/> (2022).
58. Eraslan, G. et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science* **376**, eabl4290 (2022).
59. Cyberinfrastructure for Network Science Center. HuBMAP Visible Human MOOC (VHMOOC) <https://expand.iu.edu/browse/sice/cns/courses/hubmap-visible-human-mooc> (2023).
60. Börner, K. et al. Tissue registration and exploration user interfaces in support of a human reference atlas. *Commun. Biol.* **5**, 1369 (2022).
61. Bueckle, A. & Qaurooni, D. Using the Standalone Registration User Interface. *Zenodo* <https://doi.org/10.5281/zenodo.14346542> (2024).
62. Keller, M. S. et al. Vitesse: integrative visualization of multimodal and spatially resolved single-cell data. *Nat. Methods* **22**, 63–67 (2024).
63. Manz, T. et al. Viv: multiscale visualization of high-resolution multiplexed bioimaging data on the web. *Nat. Methods* **19**, 515–516 (2022).
64. Bidanta, S. et al. Functional tissue units in the Human Reference Atlas. *Nat. Commun.* **16**, 1526 (2025).
65. Bueckle, A. et al. The HRA organ gallery affords immersive superpowers for building and exploring the Human Reference Atlas with virtual reality. *Front. Bioinform.* **3**, 1162723 (2023).
66. Bueckle, A. & Cyberinfrastructure for Network Science Center. HRA Organ Gallery in VR <https://humanatlas.io/hra-organ-gallery> (2024).
67. Hickey Lab. Hierarchical-Tissue-Unit-Annotation. *GitHub* <https://github.com/HickeyLab/Hierarchical-Tissue-Unit-Annotation> (2024).
68. Cyberinfrastructure for Network Science Center. HRA-construction-usage-supporting-information. *GitHub* <https://github.com/cns-iu/hra-construction-usage-supporting-information> (2024).
69. Hickey, J. W. et al. Organization of the human intestine at single-cell resolution. *Nature* **619**, 572–584 (2023).
70. Cyberinfrastructure for Network Science Center. HRA cell distance explorer. <https://apps.humanatlas.io/cde/> (2024).
71. Domínguez Conde, C. et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).
72. Poon, M. M. L. et al. Tissue adaptation and clonal segregation of human memory T cells in barrier sites. *Nat. Immunol.* **24**, 309–319 (2023).
73. Matsumoto, R. et al. Induction of bronchus-associated lymphoid tissue is an early life adaptation for promoting human B cell immunity. *Nat. Immunol.* **24**, 1370–1381 (2023).
74. Lao, J. C. et al. Type 2 immune polarization is associated with cardiopulmonary disease in preterm infants. *Sci. Transl. Med.* **14**, eaaz8454 (2022).
75. Wang, R. et al. Immunity and inflammation in pulmonary arterial hypertension: From pathophysiology mechanisms to treatment perspective. *Pharmacol. Res.* **180**, 106238 (2022).
76. Dylag, A. M. et al. New insights into the natural history of bronchopulmonary dysplasia from proteomics and multiplexed immunohistochemistry. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **325**, L419–L433 (2023).
77. Jain, Y. et al. Segmenting functional tissue units across human organs using community-driven development of generalizable machine learning algorithms. *Nat. Commun.* **14**, 4656 (2023).
78. Jain, Y. et al. Segmentation of human functional tissue units in support of a Human Reference Atlas. *Commun. Biol.* **6**, 717 (2023).
79. Hickey, J. Processed single cell data from CODEX multiplexed imaging of the human intestine. 2,910,510,864 bytes. *Dryad* <https://doi.org/10.5061/dryad.pk0p2ngrf> (2022).
80. HRA Working Group registration https://iu.co1.qualtrics.com/jfe/form/SV_bpaBhlr8XfdiNRH (2024).
81. Cyberinfrastructure for Network Science Center. hra-construction-usage-supporting-information/source-data-for-figures at main · cns-iu/hra-construction-usage-supporting-information. *GitHub* <https://github.com/cns-iu/hra-construction-usage-supporting-information/tree/main/source-data-for-figures> (2024).
82. NIH. Common Coordinate Framework Meeting. <https://commonfund.nih.gov/sites/default/files/CCF%20summary%20final%20.pdf> (2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

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

²CIFAR MacMillan Multiscale Human program, CIFAR, Toronto, Ontario, Canada. ³Pittsburgh Supercomputing Center, Carnegie Mellon University, Pittsburgh, PA, USA. ⁴Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ⁵Ray and Stephanie Lane Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA. ⁶New York Genome Center, New York, NY, USA. ⁷Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, UK. ⁸Department of Medicine, University of Cambridge, Cambridge, UK. ⁹University of Rochester Medical Center, Rochester, NY, USA. ¹⁰Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA. ¹¹Department of Biomedical Engineering, Duke University, Durham, NC, USA. ¹²J. Craig Venter Institute, La Jolla, CA, USA. ¹³Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: katy@iu.edu; abuেকে@iu.edu; bherr@iu.edu

HRA Team

Jakub Abramson¹⁴, David Anderson¹⁵, Kristin Ardlie¹⁶, Mark J. Arends¹⁷, Bruce J. Aronow¹⁸, Rachel Bajema¹⁹, Richard A. Baldock¹⁷, Ross Barnowski²⁰, Daria Barwinska²¹, Amy Bernard²², David Betancur³, Supriya Bidanta¹⁹, Frida Björklund²³, Axel Bolin¹⁹, Avinash Boppana¹⁹, Luke Boulter¹⁷, Kristen Browne²⁴, Maigan A. Brusko²⁵, Albert Burger²⁶, Martha Campbell-Thompson²⁵, Ivan Cao-Berg³, Anita R. Caron²⁷, Megan Carroll³, Chrystal Chadwick²⁸, Haoran Chen²⁹, Lu Chen³⁰, Bernard de Bono²¹, Gail Deutsch³¹, Song-Lin Ding³², Sean Donahue²⁹, Tarek M. El-Achkar³³, Adel Eskaros³⁴, Louis Falo Jr.³⁵, Melissa Farrow³⁶, Michael J. Ferkowicz²¹, Stephen A. Fisher³⁷, James C. Gee³⁷, Ronald N. Germain³⁸, Michael Ginda¹⁹, Fiona Ginty²⁸, Sarah A. Gitomer³⁹, Melanie B. Goldstone¹⁹, Katherine S. Gustilo^{19,39}, James S. Hagood⁴⁰, Marc K. Halushka⁴¹, Muzlifah A. Haniffa⁴², Peter Hanna⁴³, Josef Hardi²³, Yongqun Oliver He⁴⁴, Brendan John Honick⁴⁵, Derek Houghton²⁶, Maxim Itkin⁴⁶, Sanjay Jain⁴⁷, Laura Jardine⁴⁸, Z. Gordon Jiang⁴⁹, Yingnan Ju¹⁹, Arivarasan Karunamurthy³⁵, Neil L. Kelleher⁵⁰, Timothy J. Kendall¹⁷, Angela R. S. Kruse³⁶, Monica M. Laronda⁵¹, Louise C. Laurent⁵², Elisa Laurenti⁵³, Sujin Lee⁵⁴, Ed Lein⁵⁵, Chenran Li³⁰, Zhuoyan Li⁶, Shin Lin³¹, Yiing Lin⁴⁷, Scott A. Lindsay⁵², Teri A. Longacre²³, Emma Lundberg²³, Libby Maier¹⁹, Rajeev Malhotra⁵⁴, Anna Martinez Casals²³, Anna Maria Masci⁵⁶, Clayton E. Mathews²⁵, Elizabeth McDonough²⁸, James A. McLaughlin⁵⁷, Rajasree Menon⁵⁸, Vilas Menon⁵⁹, Jeremy A. Miller³², Richard Morgan⁴, Werner Müller⁶⁰, Robert F. Murphy²⁹, Mark A. Musen²³, Harikrishna Nakshatri²¹, Martijn C. Nawijn⁶¹, Elizabeth K. Neumann⁶², Debra J. Nigra³, Kathleen O'Neill³⁷, Mana M. Parast⁶³, Ushma Patel¹⁹, Liming Pei⁶⁴, Hemali Phatnani⁶⁵, Gesina A. Phillips³, Alison M. Pouch³⁷, Alvin C. Powers⁶⁶, Juan F. Puerto³, Aleix Puig-Barbe²⁷, Ellen M. Quardokus¹⁹, Andrea J. Radtke^{38,67}, Presha Rajbhandari⁶⁸, Elizabeth G. Record¹⁹, Drucilla J. Roberts⁶⁹, Alexander J. Ropelewski³, David Rowe⁷⁰, Nancy L. Ruschman¹⁹, Diane C. Saunders¹⁵, Richard H. Scheuermann⁷¹, Kevin L. Schey³⁶, Birgit Schilling⁷², Heidi Schlehlein¹⁹, Melissa Schwenk⁴, Robin Scibek³, Robert P. Seifert⁷³, Bill Shirey⁴, Kalyanam Shivkumar⁴³, Kimberly Siletti⁷⁴, J. Alan Simmons⁴, Dhruv Singhal⁷⁵, Michael Snyder²³, Jeffrey M. Spraggins¹⁵, Valentina Stanley⁵², Douglas W. Strand⁷⁶, Joel C. Sunshine⁷⁷, Christine Surette²⁸, Ayako Suzuki⁷⁸, Purushothama Rao Tata⁷⁸, Deanne M. Taylor⁷⁹, Todd Theriault¹⁹, Tracey Theriault¹⁹, Jerin Easo Thomas¹⁹, Elizabeth L. Tsui⁸⁰, Jackie Uranic³, M. Todd Valerius⁸¹, David Van Valen²⁰, Chad M. Vezina⁸², Ioannis S. Vlachos⁴⁹, Fusheng Wang³⁰, Xuefei 'Julie' Wang²⁰, Clive H. Wasserfall²⁵, Joel S. Welling³, Christopher Werlein⁸³, Seth Winfree⁸⁴, Devin M. Wright¹⁹, Li Yao⁸⁵, Zhou Yuan⁴ & Ted Zhang²⁹

¹⁴Weizmann Institute of Science, Rehovot, Israel. ¹⁵Vanderbilt University, Nashville, TN, USA. ¹⁶Broad Institute, Cambridge, MA, USA. ¹⁷University of Edinburgh, Edinburgh, UK. ¹⁸Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. ¹⁹Indiana University, Bloomington, IN, USA. ²⁰California Institute of Technology, Pasadena, CA, USA. ²¹Indiana University School of Medicine, Indianapolis, IN, USA. ²²The Kavli Foundation, Los Angeles, CA, USA. ²³Stanford University, Stanford, CA, USA. ²⁴National Institute of Allergy and Infectious Diseases, National Institutes of Health, Buffalo, NY, USA. ²⁵University of Florida, Gainesville, FL, USA. ²⁶Heriot-Watt University, Edinburgh, UK. ²⁷EMBL-EBI, Hinxton, Cambridgeshire, UK. ²⁸GE HealthCare Technology & Innovation Center, Niskayuna, NY, USA. ²⁹Carnegie Mellon University, Pittsburgh, PA, USA. ³⁰Stony Brook University, Stony Brook, NY, USA. ³¹University of Washington, Seattle, WA, USA. ³²Allen Institute for Brain Science, Seattle, WA, USA. ³³Indiana University School of Medicine & Indianapolis VA Medical Center, Indianapolis, IN, USA. ³⁴Vanderbilt University Medical Center, Nashville, TN, USA. ³⁵University of Pittsburgh, Pittsburgh, PA, USA. ³⁶Vanderbilt University School of Medicine, Nashville, TN, USA. ³⁷University of Pennsylvania, Philadelphia, PA, USA. ³⁸National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA. ³⁹University of Colorado School of Medicine, Aurora, CO, USA. ⁴⁰University of North Carolina-Chapel Hill, Chapel Hill, NC, USA. ⁴¹Cleveland Clinic Foundation, Cleveland, OH, USA. ⁴²Wellcome Sanger Institute, Cambridge, UK. ⁴³University of California, Los Angeles, Los Angeles, CA, USA. ⁴⁴University of Michigan, Ann Arbor, MI, USA. ⁴⁵Pittsburgh Supercomputing Center & Carnegie Mellon University, Pittsburgh, PA, USA. ⁴⁶Penn Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁴⁷Washington University in St. Louis, St. Louis, MO, USA. ⁴⁸Newcastle University, Newcastle upon Tyne, UK. ⁴⁹Beth Israel Deaconess Medical Center & Harvard Medical School, Boston, MA, USA. ⁵⁰Northwestern University, Evanston, IL, USA. ⁵¹Ann and Robert H. Lurie Children's Hospital of Chicago, Northwestern University, Chicago, IL, USA. ⁵²University of California, San Diego, San Diego, CA, USA. ⁵³University of Cambridge, Cambridge, UK. ⁵⁴Massachusetts General Hospital & Harvard Medical School, Boston, MA, USA. ⁵⁵Allen Institute, Seattle, WA, USA. ⁵⁶University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁵⁷EMBL-EBI, Cambridge, UK. ⁵⁸University of Michigan Medical School, Ann Arbor, MI, USA. ⁵⁹Columbia University Irving Medical Center, New York, NY, USA. ⁶⁰University of Manchester, UK and Miltenyi Biotec, Germany, Bergisch-Gladbach, Germany. ⁶¹University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. ⁶²UC Davis, Davis, CA, USA. ⁶³University of California San Diego, La Jolla, CA, USA. ⁶⁴Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA, USA. ⁶⁵Columbia University & New York Genome Center, New York City, NY, USA. ⁶⁶Vanderbilt University, Vanderbilt University Medical Center, VA Tennessee Valley Healthcare System, Nashville, TN, USA. ⁶⁷Leica Microsystems Inc., Wetzlar,

Germany. ⁶⁸Columbia University, New York, NY, USA. ⁶⁹Harvard Medical School-Mass General Brigham, Boston, MA, USA. ⁷⁰University of Connecticut Health, Farmington, CT, USA. ⁷¹National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. ⁷²Buck Institute for Research on Aging, Novato, CA, USA. ⁷³University of Florida College of Medicine, Gainesville, FL, USA. ⁷⁴University Medical Center Utrecht, Utrecht, The Netherlands. ⁷⁵Harvard-Beth Israel Deaconess Medical Center, Boston, MA, USA. ⁷⁶University of Texas Southwestern Medical Center, Dallas, TX, USA. ⁷⁷Johns Hopkins University School of Medicine, Baltimore, MD, USA. ⁷⁸Duke University School of Medicine, Durham, NC, USA. ⁷⁹The Children's Hospital of Philadelphia & University of Pennsylvania Perelman Medical School, Philadelphia, PA, USA. ⁸⁰Northwestern University, Chicago, IL, USA. ⁸¹Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ⁸²University of Wisconsin-Madison, Madison, WI, USA. ⁸³Hannover Medical School, Hannover, Germany. ⁸⁴National Institute of Allergy and Infectious Disease, National Institutes of Health, Rockville, MD, USA. ⁸⁵Liyaovisuals Design Studio, Dewood, MD, USA.

Methods

Human-expert-generated data and experimental tissue data are used to construct the HRA (Fig. 1). New experimental data is mapped to the HRA via (1) 3D spatial registration; (2) using suspension-based (for example, sc/snRNA-seq); or (3) spatial (for example, CODEX⁸³, Cell DIVE⁸⁴, IBEX^{85,86}, Imaging Mass Cytometry⁸⁷ and other multiplexed, antibody-based protein imaging platforms) assay types that are aligned with the HRA (Fig. 2).

Expert-generated data

ASCT+B tables. ASCT+B tables (<https://humanatlas.io/asctb-tables>; Fig. 1a) are compiled by experts using the ASCT+B Reporter (Supplementary Fig. 4) following SOPs⁸⁸. Note that the brain ASCT+B table is unique in that it was computationally derived using the common cell type nomenclature approach⁸⁹, which chains together critical cell type features (for example, brain region and cortical layer), broad cell-type class and gene biomarker information into the annotation.

Starting with the sixth release of the HRA, new and revised tables list cell type parents present in CL for the about 600 cell types that currently have ASCTB-TEMP IDs (temporary ontology terms and IDs) as they do not yet exist but are systematically added into CL via the HRA effort. This makes it possible to show the complete cell typology in CellGuide (<https://cellxgene.cziscience.com/cellguide>) and other tools. For example, Supplementary Fig. 17 depicts a CZ CellGuide visualization for neurons (CL:0000540) showing the CL ontology typology with the ‘neuron’ cell type highlighted in green together with its parent (‘neural cell’, which is a ‘somatic cell’, which is an ‘animal cell’) and children nodes (for example, ‘GABAergic neuron’ and ‘glutamatergic neuron’). The interactive visualization is at https://cellxgene.cziscience.com/cellguide/CL_0000540.

A special focus within HuBMAP has been the development of a detailed blood vasculature ASCT+B table in support of a VCCF^{90–92} (<https://humanatlas.io/vccf>). Relevant data captured in the VCCF include blood vessels and their branching relationships, as well as associated cell types and biomarkers, the vessel type, anastomoses, portal systems, microvasculature, FTUs, links to 3D reference objects, vessel geometries (length and diameter) and mappings to anatomical structures the vessels supply or drain.

2D and 3D reference objects. Professional medical illustrators follow SOPs^{93,94} to generate 2D reference FTU illustrations and 3D reference anatomical structures (Fig. 1b,c). Most 3D reference organs were modeled using the male and female datasets from the Visible Human Project provided by the National Library of Medicine⁹⁵.

The ASCT+B tables in the sixth HRA release feature ontology-aligned terminology for 4,499 unique anatomical structures and 1,195 cell types. For some of these anatomical structures and cell types there exist anatomically aligned, spatially explicit reference objects. Specifically, there are 2D illustrations of 22 FTUs in 10 organs with 3,742 cells of 116 cell types plus 3D reference objects for 1,192 anatomical structures with 516 unique Uberon IDs in 65 3D reference objects (male and female, left and right organs) with 37 unique Uberon IDs. A crosswalk associates each of the 2D/3D anatomical structures and cell types with their corresponding terms in the ASCT+B tables (see SOP section)⁹⁶.

Segmentation masks. Different tools are used to support manual segmentation of images by human experts (to assign each pixel in an image to an object such as a single cell, FTU or anatomical structure). Within the HRA effort, the QuPath⁹⁷ tool is used by organ experts to generate 2D segmentation masks for FTUs and vasculature (see SOP^{98,99}) and DeepCell Label (<https://label.deepcell.org>) is used to get 2D segmentation masks for single cells. Resulting ‘gold standard’ segmentation and annotation data (Fig. 1d) are needed to train machine-learning algorithms so that experimental datasets can be automatically segmented (Fig. 2a(v)).

OMAPs and AVRs. OMAPs (<https://humanatlas.io/omap>) are collections of antibodies designed for a particular sample preservation method and multiplexed imaging technology to allow spatial mapping of the anatomical structures and cell types present in the tissues for which they were validated^{15,100} (Fig. 1e). OMAPs are wet bench validated antibodies, which experts initially identify as candidates for their multiplexed antibody-based imaging experiments by using literature, available antibody search engines and potentially also the ASCT+B Reporter (Supplementary Fig. 4 and SOP¹⁰¹). Antibodies in OMAPs link to expert-generated HuBMAP AVRs (<https://avr.hubmapconsortium.org> and SOP¹⁰²) that provide details on the characterization of individual antibodies for multiplexed antibody-based imaging assays. Antibody validation is expensive and time consuming, so these resources are designed to jump start other researchers to be successful and reduce the time and money required for multiplexed antibody-based imaging studies.

Cell annotation references. A large majority of single-cell data are single-cell or single-nucleus RNA-seq data. Cell-type annotation tools (Fig. 1f) such as Azimuth¹⁴, CellTypist^{71,103} and popular Vote (popV)¹⁰⁴ are commonly used to cluster cells based on their gene expression profiles, followed by assigning those Uniform Manifold Approximation and Projection¹⁰⁵ clusters to cell types based on published gene expression profiles. Supplementary Table 1 shows the number of cell types that these three tools can assign per organ (rightmost columns)—compared to the number of cell types in the ASCT+B tables and 3D reference object library (middle columns); the second column shows the number of datasets available via the HuBMAP, SenNet, GTEx and CZ CELLxGENE data portals. Note that datasets for some organs (for example, urinary bladder) do not exist.

Human expertise is required to compile crosswalks that associate cell labels assigned by these three tools to terms in CL. Mapping cell labels to CL can be partially automated; however, this is more effective if the labels researchers provide are written out rather than listed as abbreviations, as different research groups do not use standardized abbreviations for cell types. Automated mapping to CL is further hindered when the cell type is not yet present in CL, in this case, often a parent cell type is used as a placeholder until the exact cell type can be added to the ontology. For these reasons, it is desirable to construct crosswalks that use the most specific cell type supported by experimental data. Depending upon the number of active editors/curators available for adding the new cell types that single-cell RNA sequencing is discovering, prioritization of new terms and collecting supporting literature takes time. Resulting crosswalks are organ-specific and they are published as cell type annotation specific crosswalks that associate any cell type assigned by the three tools with the corresponding term in the ASCT+B tables, see examples¹⁰⁶.

Experimental data

The HuBMAP Data Portal (<https://portal.hubmapconsortium.org>) uses a microservices architecture (Supplementary Fig. 1) to serve data and code via a hybrid on-premises and cloud approach using federated identity management, UUIDs and full provenance for data management plus data security. Workflow and container support exist for diverse unified analysis pipelines and interactive exploration tools. This architecture makes it possible to ingest data at scale, adjust metadata formats as needed, add new algorithms and workflows as they become available and ensure production phase speed and scalability for all services. On 20 January 2024, the HuBMAP Data Portal provided open access to 2,332 datasets from 213 donors. Overall, 360 of these datasets are sc/snRNA-seq and 79 are spatial OMAP-aligned datasets.

Tissue collection and RUI registration. The RUI⁶⁰ (<https://apps.humanatlas.io/rui>) was implemented to support the spatial registration of tissue blocks into the HRA CCF; Supplementary Fig. 6. It collects

sample ID, donor metadata, plus provenance information (who registered the data and when) in the process. Subject-matter experts with knowledge of the spatial and donor data for the tissue samples use the RUI to register their tissue samples—supported by a designated HRA registration coordinator as needed, see SOP⁶¹. Alternatively, a more collaborative workflow is available in which the registration coordinator plays a more active role in making the registration with guidance from a subject-matter expert. These workflows are explained in two dedicated SOPs detailing how the RUI can be used⁶¹ and the responsibilities of the registration coordinator¹⁰⁷. Next, the registration coordinator uses a location processor tool¹⁰⁸ to combine tissue sample metadata with de-identified donor metadata (sex, age, body mass index, race, etc.) and publication metadata (DOI, authors, publication year, etc.). Once the samples are registered and the metadata has been enriched, the registration coordinator contacts the subject-matter expert to check accuracy and completeness. The registration coordinator then publishes the validated registration set, making it accessible through the EUI (<https://apps.humanatlas.io/eui>; Supplementary Fig. 7).

Tissue block registration can be streamlined and made more reproducible through the use of the ‘millitome,’ (<https://humanatlas.io/millitome>) a device that aids wet bench scientists to cut and register multiple tissue blocks from a single organ in a reproducible manner. This 3D-printable apparatus is designed to secure a freshly procured organ and is fitted with cutting grooves that can direct a carbon steel cutting knife for uniform slicing, see HRA’s millitome catalog (<https://hubmapconsortium.github.io/hra-millitome>) to access and customize organ millitomes based on donor sex, organ laterality, organ size and cutting intervals. Each millitome package contains an STL file for 3D printing the millitome’s reproducible surface geometry plus a lookup sheet correlating millitome locations with tissue sample IDs assigned by the research team. After slicing the organ using the millitome, scientists document the samples on the lookup sheet and submit this data for review by the HRA millitome facilitator. Once the package is complete, data are added to the EUI for review by scientists to verify registration accuracy in terms of tissue size, placement and orientation. SOPs detail millitome construction¹⁰⁹ and usage¹¹⁰.

sc/snRNA-seq transcriptomic data annotation. The sc/snRNA-seq transcriptomic datasets are downloaded from four data portals using the hra-workflows-runner (<https://github.com/hubmapconsortium/hra-workflows-runner>). For data from HuBMAP and SenNet (each dataset comes from exactly one donor), search APIs (HuBMAP, <https://search.api.hubmapconsortium.org/v3>; SenNet, <https://search.api.sennetconsortium.org>) are used to obtain a list of dataset IDs for all existing cell-by-gene matrices in H5AD format and to download these files plus donor metadata. For GTEx, a single H5AD file is downloaded from <https://gtexportal.org/home/singleCellOverviewPage>. For CZ CELLxGENE, datasets are stored in collections and one collection can contain multiple datasets and donors; the workflow runner reads in an index of all healthy adult human collections compiled using the CZI Science CELLxGENE Python API (<https://chanzuckerberg.github.io/cellxgene-census/python-api.html>); it splits the collection into unique donor-dataset pairs; and runs all H5AD files through the three cell type annotation tools: Azimuth¹⁴, CellTypist⁷¹ and popV^{104,111} (Supplementary Table 1). Azimuth (<https://azimuth.hubmapconsortium.org>) serves organ-specific human adult references for ten unique organs (lung and tonsil have a revised v.2 that is used here); for Azimuth, there exists HRA crosswalks¹⁰⁶ for 226 unique cell types in seven organs (3D spatial reference organs do not exist for blood, adipose tissue and bone marrow). For CellTypist (<https://www.celltypist.org>), there are crosswalks for 13 organs and a total of 214 unique cell types. For popV (<https://github.com/YosefLab/PopV>), we provide crosswalks for 22 organs and 134 unique cell types. There are 574 total cell types linked across all three tools. The workflow runner outputs four files: (1) cell summaries for all sc-transcriptomics datasets, subset by cell type annotation tool; (2) a

corresponding metadata file with donor and publication information; (3) cell summaries for all sc-proteomics datasets; and (4) a corresponding metadata file with donor and publication information¹¹². All four files are used during the enrichment phase to construct the atlas-level HRApop data.

Cell and FTU segmentation for spatial data. Antibody-based multiplexed imaging datasets, once uploaded to the HuBMAP Data Portal via the Ingest portal, are processed using a unified CWL workflow for cell and nuclei segmentation. Whole cell segmentation for CODEX datasets (<https://github.com/hubmapconsortium/codex-pipeline>) is conducted using Cytokit¹¹³ and Cell DIVE (<https://github.com/hubmapconsortium/celldive-pipeline>). MIBI (<https://github.com/hubmapconsortium/mibi-pipeline>) datasets are processed using Deepcell’s Mesmer model¹¹⁴. Resulting cell segmentations are assigned a segmentation quality score using CellSegmentationEvaluator⁵¹. Cell segmentation for forthcoming 3D spatial proteomics datasets is provided by 3DCellComposer¹¹⁵ in combination with trained 2D segmenters.

FTU segmentation on periodic acid–Schiff/H&E-stained histology datasets is conducted using code developed via two Kaggle competitions^{77,78}. The current production pipeline includes support for FTUs in the kidney, with large intestine and spleen that will be run when histology datasets become available.

Cell type annotation for spatial proteomic data. After cell segmentation, spatial cell type annotation is performed using the antibody metadata for marker channels for CODEX datasets, soon expanding to MIBI and Cell DIVE. OMAPs link marker panels in the datasets to cell types in the ASCT+B tables. The SPRM package (<https://github.com/hubmapconsortium/sprm>) computes various statistical analyses, including mean marker expression for all cells. The Van Valen laboratory has developed a language-informed vision model, DeepCellTypes¹¹⁶ to classify cell types across tissue types and imaging technologies. This model covers 30+ cell types and will be updated as new multiplexed imaging data become available. DeepCellTypes is available at <https://github.com/vanvalenlab/deepcell-types>. In addition to this model, various teams have been annotating cell types with different approaches such as manual labeling with clustering or graph-based networks such as STELLAR¹¹⁷. The intestine datasets by Hickey et al.^{69,117,118} were annotated using a combination of manual and STELLAR approaches.

Spatial alignment for 2D multi-omics data. Spatial structural alignment of different segmentation masks, see Fig. 2a(iv) in support of multi-omics assay data analysis and/or alignment of spatial transcriptomics data to H&E imaging data can be performed using tools like STalign¹¹⁹. In STalign, segmented cellular spatial positions are rasterized into an image representation to be aligned with structurally matched H&E images. Because tissues may be rotated, stretched or otherwise warped during data collection, both affine and diffeomorphic alignments are performed. Such an alignment is achieved by optimizing an objective function that seeks to minimize the image intensity differences between a target (rasterized cell positions) and source (H&E) image subject to regularization penalties. The resulting learned transformation is applied to the original segmented cellular spatial positions to move the points into an aligned coordinate space. Such 2D spatial alignment facilitates downstream molecular and cell-type compositional comparisons within matched structures as well as integration across technologies.

Spatial data 3D reconstruction. Spatial alignment of multiple subsequent tissue sections in support of 3D tissue block reconstruction (Fig. 2a(vi)) has been performed using MATRICS-A¹²⁰ for skin data. Additional tools for 3D tissue block reconstruction have been developed and include SectionAligner, 3DCellComposer¹¹⁵ and CellSegmentationEvaluator. SectionAligner takes as input a series of images of 2D tissue sections, segments each piece of tissue in each section and aligns

the slices of each piece into a 3D image. 3DCellComposer uses one of various trained 2D cell segmenters (such as Mesmer), to segment each 3D image into individual cells using CellSegmentationEvaluator to automatically optimize parameter settings.

Atlas-enriched data

Mesh-level collision detection. Extraction sites are post-processed via code specifically developed for efficient spatial registration using mesh surfaces¹²¹. To improve performance during tissue registration, the RUI uses bounding-box collision detection to determine (approximate but fast) intersections at runtime. To optimize accuracy, surface mesh collision detection is used during the enrichment phase to determine exact intersection volumes between a given RUI location and any anatomical structures it intersects with based on mesh-level colliders. The '3D Geometry-Based Tissue Block Annotation: Collision Detection between Tissue Blocks and Anatomical Structures' code is available on GitHub¹²² and the API is deployed to AWS¹²³.

HRAlit. The HRA DOs from the sixth release (for 4,499 anatomical structures, 1,295 cell types and 2,098 biomarkers) were linked to 7,103,180 publications, which are associated with 583,117 authors, 896,680 funded projects and 1,816 experimental datasets³². The resulting HRAlit database represents 21,704,001 records as a network with 8,694,233 nodes and 14,096,735 links. It has been mined to identify leading experts, major papers, funding trends or alignment with existing ontologies in support of systematic HRA construction and usage. All data and code are at <https://github.com/cns-iu/hra-literature>.

HRApop. HRApop provides experimental data evidence for the existence of specific cell types and mean biomarker expression values for datasets and anatomical structures for which 3D reference models exist. In the sixth release of the HRA, there are 1,192 anatomical structures of 516 types (unique Uberon IDs) for 65 organs (including male/female and left/right).

There are three criteria that experimental datasets have to meet to be used in HRApop construction: they must (1) be spatially registered using the RUI; (2) have cell type population data (for example, an H5AD file that can be annotated via CTann tools (Supplementary Table 1) or via proteomics workflows); (3) come from a data portal with quality assurance/quality control or have been published in a peer-reviewed paper.

To construct HRApop v.0.10.2, we downloaded 9,613 H5AD single-cell transcriptomics datasets from four data portals: HuBMAP, SenNet, CZ CELLxGENE and GTE_x. Exactly 5,118 H5AD datasets were healthy and could be annotated using Azimuth, CellTypist or popV (Supplementary Table 1). In addition, we downloaded 74 single-cell proteomics datasets from HuBMAP published in two papers^{69,120}). In sum, 553 datasets (479 sc/snRNA-seq transcriptomics and 74 spatial proteomics datasets) satisfied the three criteria and were used for HRApop construction.

The resulting HRApop v.0.10.2 was validated and optimized by making predictions for datasets for which both an RUI-registered extraction site and a cell type population exist. It was then used to predict cell type annotation or spatial origin for 2,004 HuBMAP, 166 SenNet and 4,789 CZ CELLxGENE datasets for which this information was missing.

VCCF distances and Vitesse visualizations in 2D. In support of constructing a VCCF^{90–92}, code that measures and graphs the distance of different cell types to blood vessel cell types in 2D and 3D (see SOP¹²⁴) has been developed. Distance plots can be overlaid on the tissue section using Vitesse^{62,63} for 2D data and explored in 2D and 3D using custom code^{120,125}; Figs. 2a(viii) and 3a provide examples. A new tool, Cell Distance Explorer, is also made available to visualize cell-to-cell graphs and distance distributions for both 2D and 3D datasets (<https://apps.humanatlas.io/cde>).

Hierarchical community analysis of cell types. Hierarchical community analysis of cell types makes it possible to automatically detect multilevel FTUs⁶⁹. The approach uses the single-cell labels and x, y coordinates from spatial datasets. For the preview example featured in this paper (Fig. 3b), the dataset is a CODEX multiplexed imaging^{83,118,126} dataset of the healthy human intestine⁶⁹. The original multiplexed imaging data was segmented, normalized and clustered using z -normalization of the antibody markers used with Leiden unsupervised clustering¹⁰⁰. Cell types were propagated to additional samples using the deep-learning algorithm (STELLAR) for cell type label transfer in spatial single-cell data¹¹⁷.

Once cell type labels were assigned, cell neighborhoods were calculated by clustering nearest neighbor ($n = 10$) vectors surrounding each cell. A similar approach was taken to identify larger structures (termed communities⁶⁹) using neighborhoods as the labels and taking a larger window for the nearest neighbors ($n = 100$). Similarly, to identify major tissue units, community labels were used and an even larger window for nearest neighbors ($n = 300$) before clustering of the vectors. Once all tissue structures were identified, the connections in terms of primary components from various levels of tissue structures can be connected and visualized via a network graph. Currently, each node is organized per level and connected to the next spatial layer (for example, cell type to neighborhood, neighborhood to community). This code is deposited on GitHub⁶⁷.

Atlas validation

Each DO in the HRA is validated either by human expert review or using algorithmic means. HRA DO data formats depending on the type: ASCT+B is in CSV format, 3D Reference Organs in GLB format, 2D FTUs in scalable vector graphics, etc. When these data are normalized to LinkML format, the source data are processed and structural errors in the raw data are identified. Once in a normalized form, LinkML is used to validate the structure of the transformed data, including ensuring that data types and URLs fall within acceptable parameters. This step catches basic errors, including malformed URLs, missing data and incorrect data types that can be a problem downstream. Beyond this, certain DO types go through more advanced semantic checks to be sure that ontology terms used actually exist and that assertions from the DO also appear in trusted ontologies like Uberon and Cell Ontology. Validation of the ASCT+B tables is most rigorous and involves detailed review and reporting from the EBI team¹²⁷. While these tables are being authored, new/updated terms and relationships are published in the latest ontology versions available in Ubergraph (Uberon 2024-01-18 and CL 2024-01-05 for the sixth HRA) and weekly reports are generated at <https://hubmapconsortium.github.io/ccf-validation-tools/> to aid table authors in getting the highest quality data for the HRA.

Flexible hybrid cloud microservices architecture

Hybrid cloud. The IEC developed a hybrid cloud infrastructure that leverages the unique strengths of both on-premises and public cloud resources—each colocating robust and scalable storage with robust and scalable computing—providing the flexibility to proactively adapt to evolving technologies and respond to the needs of the HuBMAP Consortium and the broader atlasing community. As a key piece of this strategy, the HIVE IEC ingested, processed and archived HuBMAP data at PSC. This approach provides flexible access, as the primary copy of HuBMAP data can be stored on-premises at a low cost, but then made available on any public or local resource without incurring substantial industry standard data egress charges, as well as free, low-friction access, as researchers can run basic analyses without having to create a public cloud account or larger analyses by accessing the full HuBMAP data repository colocated with PSC's national supercomputing infrastructure made available without charge to the research community.

Microservices architecture. The HuBMAP microservices architecture (Supplementary Fig. 1) is built via agile development practices based on user-centered design, with microservices that communicate using REST APIs¹²⁸ via Docker orchestration on AWS and on-premise resources. Each microservice is focused to serve specific functionality. Services are packaged into individual Docker containers. This orchestration of Docker containers is routinely built and rebuilt in development, test and production instances which allows for independent operation and monitoring. This microservices architecture supports the plug-and-play of a continuously evolving set of algorithms required for experimental data ingestion, annotation, segmentation, search, filter and visualization, as well as for atlas construction and usage. Supplementary Fig. 1 shows the resource, API and application layers with exemplary modules (the Supplementary Information website shows an interactive version that lets users click on any module to access details). The core service that others are dependent on is the Entity API, backed by a Neo4j graph database, which provides the storage (creation, retrieval, update and deletion) of all provenance and metadata information associated with HuBMAP data. The Search API allows for search of all provenance and metadata via the AWS hosted OpenSearch search engine, which holds a copy of all information maintained by the Entity API. The HuBMAP authentication and authorization model makes use of the Globus Auth service (<https://globus.org>) with login services in compliance with the OAuth2 standard (<https://oauth.net/2>), which provides user tokens that are passed among the services where they can be centrally validated and provide user authorization with linkage to defined groups via the Globus Groups service. The remaining services provide application specific functionality for support of data ingest and management (Ingest API) and unique entity tracking and ID generation (UUID API).

HRA cloud infrastructure. HRA applications, including the HRA Portal, HRA Knowledge Graph, EUI and RUI are all deployed to the web and hosted via AWS or GitHub pages. For applications requiring server-side logic, Docker containers are created, tested and built automatically with continuous integration/continuous deployment via GitHub Actions, published to Amazon Elastic Container Registry and then deployed via AWS AppRunner or Amazon Elastic Container Service. For applications which are served primarily as static files, they are tested and built automatically with CI/CD via GitHub Actions and then copied to Amazon S3 for serving or pushed to a branch for GitHub pages deployment. Except for GitHub Pages, both static and server driven applications have Amazon CloudFront act as the front-end, providing a service mesh that supports serving web requests, tracking usage, proxying requests to services running in AWS AppRunner or Elastic Container Service and caching frequently used files and responses. While AWS are used extensively in the HRA cloud infrastructure, the technology is well suited to be adapted to other platforms.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All HuBMAP data are available via the HuBMAP Data Portal at <https://portal.hubmapconsortium.org>. Azimuth references can be accessed at <https://azimuth.hubmapconsortium.org>. HRA data and code are available at the HRA Portal (<https://humanatlas.io>). HuBMAP and HRA primary and secondary data repositories are listed in Supplementary Table 2 and HRA code repositories are in Supplementary Table 3.

Code availability

Code is available on three different GitHub organizations: (1) <https://github.com/hubmapconsortium> is the HuBMAP organization where HRA started; (2) <https://github.com/cns-iu> is the organization

owned by the Cyberinfrastructure for Network Science Center at Indiana University and initial experimental HRA code starts here; and (3) <https://github.com/x-atlas-consortia> was created recently to host cross-consortia code, including *hra-kg*, *hra-pop*, *hra-apps* and *hra-api*. Supporting information is at <https://cns-iu.github.io/hra-construction-usage-supporting-information>.

References

83. Black, S. et al. CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat. Protoc.* **16**, 3802–3835 (2021).
84. Gerdes, M. J. et al. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc. Natl Acad. Sci. USA* **110**, 11982–11987 (2013).
85. Radtke, A. J. et al. IBEX: an iterative immunolabeling and chemical bleaching method for high-content imaging of diverse tissues. *Nat. Protoc.* **17**, 378–401 (2022).
86. Radtke, A. J. et al. IBEX: a versatile multiplex optical imaging approach for deep phenotyping and spatial analysis of cells in complex tissues. *Proc. Natl Acad. Sci. USA* **117**, 33455–33465 (2020).
87. Bandura, D. R. et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* **81**, 6813–6822 (2009).
88. Quardokus, E. M., Record, E. & Herr II, B. W. Authoring anatomical structures, cell types and biomarkers (ASCT+B) tables. *Zenodo* <https://doi.org/10.5281/zenodo.5746152> (2022).
89. Miller, J. A. et al. Common cell type nomenclature for the mammalian brain. *eLife* **9**, e59928 (2020).
90. Boppana, A. et al. Anatomical structures, cell types, and biomarkers of the healthy human blood vasculature. *Sci. Data* **10**, 452 (2023).
91. Weber, G. M., Ju, Y. & Börner, K. Considerations for using the vasculature as a coordinate system to map all the cells in the human body. *Front. Cardiovasc. Med.* **7**, 29 (2020).
92. Galis, Z. S. Where is Waldo: contextualizing the endothelial cell in the era of precision biology. *Front. Cardiovasc. Med.* **7**, 127 (2020).
93. Schlehlein, H. & Quardokus, E. M. Creating 3D models from datasets. *Zenodo* <https://doi.org/10.5281/zenodo.7384275> (2022).
94. Bajema, R. Creating 2D illustrations for functional tissue units (FTUs). *Zenodo* <https://doi.org/10.5281/zenodo.6703107> (2022).
95. Spitzer, V., Ackerman, M. J., Scherzinger, A. L. & Whitlock, D. The visible human male: a technical report. *J. Am. Med. Inform. Assoc.* **3**, 118–130 (1996).
96. Bajema, R., Bidanta, S. & Quardokus, E. Mapping functional tissue unit (FTU) illustrations to ASCT+B tables. *Zenodo* <https://doi.org/10.5281/zenodo.5748154> (2024).
97. What is QuPath? <https://qupath.readthedocs.io/en/latest/docs/intro/about.html> (2019).
98. Scherschel, L., Ju, Y. & Jain, Y. Manual segmentation of tissue. *Zenodo* <https://doi.org/10.5281/zenodo.5565027> (2022).
99. Jain, Y. Vasculature segmentation in 3D hierarchical phase-contrast tomography images of human kidneys. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.08.25.609595> (2024).
100. Hickey, J. W., Tan, Y., Nolan, G. P. & Goltsev, Y. Strategies for accurate cell type identification in CODEX multiplexed imaging data. *Front. Immunol.* **12**, 727626 (2021).
101. Radtke, A. J., Quardokus, E. M. & Saunders, D. C. Construction of organ mapping antibody panels for multiplexed antibody-based imaging of human tissues. *Zenodo* <https://doi.org/10.5281/zenodo.5749882> (2022).
102. McDonough, E., Saunders, D., Radtke, A. J., Quardokus, E. M. & Caldwell, M. Constructing antibody validation reports (AVRs). *Zenodo* <https://doi.org/10.5281/zenodo.7418623> (2022).

103. Xu C. et al. Automatic cell type harmonization and integration across Human Cell Atlas datasets. *Cell* **186**, 5876–5891.e20 (2023).
104. Ergen, C. Consensus prediction of cell type labels in single-cell data with popV. *Nature* <https://www.nature.com/articles/s41588-024-01993-3> (2024).
105. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).
106. Cyberinfrastructure for Network Science Center. HRA-workflows-runner, crosswalking-tables. *GitHub* <https://github.com/hubmapconsortium/hra-workflows-runner/tree/main/crosswalking-tables> (2024).
107. Wright, D. & Quaroni, D. Managing Human Reference Atlas (HRA) registrations. *Zenodo* <https://doi.org/10.5281/zenodo.10359228> (2023).
108. HuBMAP Consortium. HRA-rui-locations-processor. *GitHub* <https://github.com/hubmapconsortium/hra-rui-locations-processor> (2023).
109. Kienle, P., Quardokus, E. M. & Bueckle, A. Constructing a millitome and generating virtual tissue blocks. *Zenodo* <https://doi.org/10.5281/zenodo.7901004> (2023).
110. Bueckle, A. & Kienle, P. Using a millitome. *Zenodo* <https://doi.org/10.5281/zenodo.7382703> (2022).
111. The Tabula Sapiens Consortium et al. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
112. Cyberinfrastructure for Network Science Center. HRA-pop/input-data/v0.10.2. *GitHub* <https://github.com/x-atlas-consortia/hra-pop/tree/main/input-data/v0.10.2> (2024).
113. Czech, E., Aksoy, B. A., Aksoy, P. & Hammerbacher, J. Cytokit: a single-cell analysis toolkit for high dimensional fluorescent microscopy imaging. *BMC Bioinform.* **20**, 448 (2019).
114. Greenwald, N. F. et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-01094-0> (2021).
115. Chen, H. & Murphy, R. F. 3DCellComposer: a versatile pipeline utilizing 2D cell segmentation methods for 3D cell segmentation. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.03.08.584082> (2024).
116. Wang, X. et al. Generalized cell phenotyping for spatial proteomics with language-informed vision models. Preprint at *bioRxiv* <https://www.biorxiv.org/content/10.1101/2024.11.02.621624v2> (2024).
117. Brbić, M. et al. Annotation of spatially resolved single-cell data with STELLAR. *Nat. Methods* **19**, 1411–1418 (2022).
118. Hickey, J. W. et al. Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging. *Nat. Methods* **19**, 284–295 (2022).
119. Clifton, K. et al. STalign: Alignment of spatial transcriptomics data using diffeomorphic metric mapping. *Nat. Commun.* **14**, 8123 (2023).
120. Ghose, S. et al. 3D reconstruction of skin and spatial mapping of immune cell density, vascular distance and effects of sun exposure and aging. *Commun. Biol.* **6**, 718 (2023).
121. Chen, L. et al. Real-time spatial registration for 3D Human Atlas. In *Proc. 10th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data* 27–35 (Association for Computing Machinery, 2022).
122. HuBMAP Consortium. ccf-tissue-block-annotation. *GitHub* <https://github.com/hubmapconsortium/ccf-tissue-block-annotation> (2023).
123. Cyberinfrastructure for Network Science Center. AWS: tissue block annotation. Collision detection between tissue blocks and anatomical structures <https://pfn8zf2gtu.us-east-2.awsapprunner.com/get-collisions> (2025).
124. Ju, Y. & Jain, Y. Computing cell type to vasculature distance distributions. *Zenodo* <https://doi.org/10.5281/zenodo.10371472> (2023).
125. HuBMAP Consortium. vccf-visualization-2022. *GitHub* <https://github.com/hubmapconsortium/vccf-visualization-2022> (2023).
126. Kennedy-Darling, J. et al. Highly multiplexed tissue imaging using repeated oligonucleotide exchange reaction. *Eur. J. Immunol.* **51**, 1262–1277 (2021).
127. Caron, A. R. et al. A general strategy for generating expert-guided, simplified views of ontologies. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.12.13.628309> (2024).
128. Fielding, R. T. et al. Reflections on the REST architectural style and ‘principled design of the modern web architecture’ (impact paper award). In *Proc. 2017 11th Joint Meeting on Foundations of Software Engineering* 4–14 (ACM, 2017).

Acknowledgements

The authors thank A. Taylor, A. Pillai, A. Ma’ayan, D. Osumi-Sutherland, G. Bader, R. Gonçalves, S. Lobentzner and Z. Galis for their expert comments and suggestions on an earlier version of this paper. We thank C. Ergen for sharing popV code and an initial crosswalk to CL, plus for validating popV results. The HRA is under active development by HuBMAP, the SenNet Consortium, the KPMP, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and the GenitoUrinary Development Molecular Anatomy Project (GUDMAP) projects with expert input by the HRA Editorial Board and in close collaboration with experts from more than 18 other consortia. K.B. and S.A.T. are co-directors of and are funded by the CIFAR MacMillan Multiscale Human program. This research has been supported by the NIH Common Fund through the Office of Strategic Coordination/Office of the NIH Director under awards: OT2OD033756 (K.B., Y.Z., G.M.W., Y.J., D.Q., A.B. and B.W.H.) and OT2OD026671 (K.B., G.M.W., Y.J., D.Q., A.B. and B.W.H.); U54 HL165443 and HLU01148861 (G.P., R.M. and J.P.); 1R03OD036499 (Y.Z.); 3U54AG075936 (J.W.H.); OT2OD026675 and OT2OD033759 (P.B., J.C.S. and A.B.); and 3OT2OD033760 (R.S., G.M. and J.F.); as well as 3OT2OD026682 and 1OT2OD033761 (M.R., S.A.T. and C.X.). Further, this work was supported by: the SenNet CODCC under award number U24CA268108 (K.B., J.C.S., Y.J., D.Q., A.B. and B.W.H.); by the NIDDK under award U24DK135157 (K.B., D.Q. and B.W.H.); by the KPMP grant U2CDK114886 (K.B., Y.J., D.Q., A.B. and B.W.H.); by the National Human Genome Research Institute RM1HG011014 (R.S.); and the NIH National Institute of Allergy and Infectious Diseases, Department of Health and Human Services under BCBB Support Services Contract HHSN316201300006W/HHSN27200002. This research was supported in part by the Intramural Research Program of the US NIH. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

K.B. led the HRA effort, led the writing of this paper and is a corresponding author with A.B. and B.W.H. P.D.B. and J.C.S. led the development of the flexible hybrid cloud microservices architecture. M.R. led the developments of the HuBMAP data analysis pipelines. R.S. and G.M. developed the Azimuth cell-type annotation tool. S.A.T. and C.X. developed the CellTypist annotation tool. G.P., R.S.M., J.F. and J.W.H. contributed atlas previews to this paper. Y.Z. provided valuable input on cell type annotation tools. G.M.W. led the vasculature CCF effort. Y.J. led the HRA image segmentation and other machine-learning efforts. D.Q. conducted interviews to determine HRA user stories. Y.K. led the HRAlit effort. A.B. led the HRA Organ Gallery and HRApop efforts. B.W.H. led the development of the HRA user interfaces. K.B., A.B., B.W.H., P.D.B., M.R., G.P., R.S.M., J.M.P., J.F., J.W.H., G.M., C.X., Y.Z., G.M.W., Y.J., D.Q. and Y.K. wrote the paper.

All other authors reviewed and commented on the paper. All HRA Team authors either attended the HRA WG and/or created and/or reviewed HRA DOs.

Competing interests

The primary authors declare the following competing interests:

R. Satija receives compensation from 10x Genomics, Parse Biosciences and Neptune Bio. R.S. is a co-founder and equity holder of Neptune Bio. S. Teichmann is a remunerated member of the Scientific Advisory Boards of QIAGEN, Foresite Labs and Element Biosciences, a co-founder and equity holder of TransitionBio and EnsoCell Therapeutics and a part-time employee of GlaxoSmithKline since January 2024. The HRA Team authors declare the following competing interests: B. Aronow declares Nexstone Immunology, Uniquity and Advisors. C. Werlein declares speaker fees from Boehringer Ingelheim. M. Snyder declares Personalis, SensOmics, Qbio, January AI, Fodsel, Filtricine, Protos, RTHM, Iollo, Marble Therapeutics, Crosshair Therapeutics, NextThought and Mirvie, Jupiter, Neuvivo, Swaza, Mitrix, Yuvan, TranscribeGlass and Applied Cognition. N. Kelleher declares Thermo Fisher Scientific, Proteinaceous, Integrated Protein Technologies and ImmPro. W. Müller declares Miltenyi Biotec. E. Lundberg is an advisor for the Chan-Zuckerberg Initiative Foundation, Element Biosciences, Cartography Biosciences, Pfizer and Pixelgen Technologies. T. Kendall serves as a consultant or

advisory board member for Resolution Therapeutics, Clinnovate Health, HistoIndex, Fibrofind, Kynos Therapeutics, Perspectum, Concept Life Sciences and Jazz Pharmaceuticals; and has received speakers' fees from Servier Laboratories, Jazz Pharmaceuticals, AstraZeneca, HistoIndex and Incyte Corporation. A. Ropelewski is an equity holder in Illumina, Nanostring, 10x Genomics and Akoya. L. Faló is a cofounder and equity holder in SkinJect. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-024-02563-5>.

Correspondence and requests for materials should be addressed to Katy Börner, Andreas Bueckle or Bruce W. Herr II.

Peer review information *Nature Methods* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Rita Strack, in collaboration with the *Nature Methods* team.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

All HuBMAP data is available via the HuBMAP Data Portal (<https://portal.hubmapconsortium.org>). Azimuth references can be accessed at <https://azimuth.hubmapconsortium.org>. HRA data and code are available at the HRA Portal (<https://humanatlas.io>).

Data analysis

Code is available on three different GitHub organizations: (1) <https://github.com/hubmapconsortium> is the HuBMAP organization where HRA started, (2) <https://github.com/cns-iu> is the organization owned by the Cyberinfrastructure for Network Science Center at Indiana University, and initial experimental HRA code starts here, (3) <https://github.com/x-atlas-consortia> was created recently to host cross-consortia code, including hra-kg (no version number), hra-pop (0.10.2), and hra-api (0.8.0).

HuBMAP and HRA primary and secondary data repositories are listed in Supplemental Table 2 and HRA code repositories in Supplemental Table 3.

Supporting information is at <https://cns-iu.github.io/hra-construction-usage-supporting-information>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All HuBMAP data is available via the HuBMAP Data Portal (<https://portal.hubmapconsortium.org>). Azimuth references can be accessed at <https://azimuth.hubmapconsortium.org>. HRA data and code are available at the HRA Portal (<https://humanatlas.io>).

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No data was collected for this paper, and thus no sample size had to be determined.
Data exclusions	No data was excluded from analysis.
Replication	The paper does not contain any previously unpublished study information. We thus do not deem it relevant or necessary to report on replication.
Randomization	The paper does not contain any previously unpublished study information. We thus do not deem it relevant or necessary to report on randomization.
Blinding	The paper does not contain any previously unpublished study information. We thus do not deem it relevant or necessary to report on binding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|--|
| n/a | Involvement in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involvement in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |