Advances and prospects for the Human BioMolecular Atlas Program (HuBMAP)

Jain, Sanjay; Pei, Liming; Spraggins, Jeffrey M.; Angelo, Michael; Carson, James P.; Gehlenborg, Niels; Ginty, Fiona; P. Gonçalves, Joana; Van de Plas, Raf; More Authors

Advances and prospects for the Human BioMolecular Atlas Program (HuBMAP)

Sanjay Jain [1,2,3,79] ✉, Liming Pei [4,79] ✉, Jeffrey M. Spraggins [5,79] ✉, Michael Angelo[6], James P. Carson[7], Nils Gehlenborg[8], Fiona Ginty[9], Joana P. Gonçalves [10], James S. Hagood [11], John W. Hickey [12], Neil L. Kelleher [13], Louise C. Laurent[14], Shin Lin[15], Yiing Lin [16], Huiping Liu [17], Alexandra Naba [18], Ernesto S. Nakayasu [19], Wei-Jun Qian [19], Andrea Radtke [20], Paul Robson [21], Brent R. Stockwell[22], Raf Van de Plas[23], Ioannis S. Vlachos[24,25], Mowei Zhou[26], HuBMAP Consortium*, Katy Börner [27,79] ✉ & Michael P. Snyder [28,79] ✉

The Human BioMolecular Atlas Program (HuBMAP) aims to create a multi-scale spatial atlas of the healthy human body at single-cell resolution by applying advanced technologies and disseminating resources to the community. As the HuBMAP moves past its first phase, creating ontologies, protocols and pipelines, this Perspective introduces the production phase: the generation of reference spatial maps of functional tissue units across many organs from diverse populations and the creation of mapping tools and infrastructure to advance biomedical research.

HuBMAP was founded with the goal of establishing state-of-the-art frameworks for building spatial multiomic maps of non-diseased human organs at single-cell resolution[1]. During the first phase (2018–2022), the priorities of the project included the validation and development of assay platforms; workflows for data processing, management, exploration and visualization; and the establishment of protocols, quality control standards and standard operating procedures. Extensive infrastructure was established through a coordinated effort among the various HuBMAP integration, visualization and engagement teams, tissue-mapping centres, technology and tools development and rapid technology implementation teams and working groups[1]. Single-cell maps, predominantly consisting of two-dimensional (2D) spatial data as well as data from dissociated cells, were generated for several organs. The HuBMAP Data Portal (https://portal.hubmapconsortium.org) was established for open access to experimental tissue data and reference atlas data.

The infrastructure was augmented with software tools for tissue data registration, processing, annotation, visualization, cell segmentation and automated annotation of cell types and cellular neighbourhoods from spatial data. Computational methods were developed for integrating multiple data types across scales and interpretation[2]. Standard reference terminology and a common coordinate framework spanning anatomical to biomolecular scales were established to ensure interoperability

across organs, research groups and consortia[3]. Guidelines to capture high-quality multiplexed spatial data[4] were established including validated panels of cell- and structure-specific antibodies[5]. The first phase produced a large number of manuscripts (https://commonfund.nih.gov/publications?pid=43) including spatially resolved single-cell maps[6–11].

The production phase of HuBMAP was launched in the autumn of 2022. The focus is on scaling data production spanning diverse biological variables (for example, age and ethnicity) and deployment and enhancement of analytical, visualization and navigational tools to generate high-resolution 3D accessible maps of major functional tissue units from more than 20 organs. This phase involves over 60 institutions and 400 researchers with opportunities for active intra- and inter-consortia collaborations and building a foundational resource for new biological insights and precision medicine. Below, we summarize major accomplishments and challenges encountered from the first phase of HuBMAP and describe the future roadmap of HuBMAP in the production phase and beyond.

Key resources and insights from the first phase of HuBMAP
Data types, organs and technologies
Since the second quarter of 2023, HuBMAP datasets using 18 different analytical technologies or assay types have been publicly available

---

A full list of affiliations appears at the end of the paper. ✉e-mail: sanjayjain@wustl.edu; lpei@pennmedicine.upenn.edu; jeff.spraggins@vanderbilt.edu; katy@indiana.edu; mpsnyder@stanford.edu

(Figs. 1a and 2). These encompass 1,672 publicly available datasets from regions with no histopathologic abnormality from 14 organs across 180 donors. Protocols were optimized to minimize the ischaemia time of healthy donors. All published data are freely available via the HuBMAP portal (https://portal.hubmapconsortium.org). These assays span spatial scales (from ~100 nm subcellular resolution to approximately centimetre and organ level) and interrogate macromolecules using diverse technologies (Figs. 1 and 2, https://software.docs.hubmapconsortium.org/assays.html). The assays include single-cell and single-nucleus transcriptomics and chromatin accessibility mapping (single-cell and single-nucleus RNA sequencing (sc/snRNA-seq) and assay for transposase-accessible chromatin using sequencing (ATAC-seq)), spatial transcriptomics and proteomics, imaging mass spectrometry (IMS) and fluorescence in situ hybridization (FISH)-based and antibody-based highly multiplexed fluorescence imaging assays as well as bulk tissue assays. Publicly available datasets include 2D or 3D spatially resolved assays surveying RNA[12,13], proteins[14] and metabolites (light-sheet microscopy[15], co-detection by indexing (CODEX)[16], matrix-assisted laser desorption and ionization, desorption electrospray ionization, secondary ion mass spectrometry (SIMS) and IMS[17,18]) (Fig. 2).

Detailed 2D single-cell maps have been generated for several organs, including the small and large intestine, kidney, placenta, lymph node, liver and spleen, and 3D maps exist for the skin. Informed consents permit the broad sharing of genomic data with the research community, through either controlled-access databases for raw sequencing files (for example, the Database of Genotypes and Phenotypes, dbGAP) or completely open access. Data that are not directly identifiable are freely available from the HuBMAP portal. Extensive metadata are included for each donor, including clinical and epidemiological data, such as sex, race (self-reported), ethnicity, age, lab results, medications, comorbidities, tissue morphology and processing parameters. The demographics of the donors can be explored at the portal.

Human Reference Atlas

An important goal of the first phase of HuBMAP was establishing standard terminology, as well as 2D and 3D anatomical structure reference frameworks for mapping the human body across scales[1]: from the entire human body to organs, functional tissue units (FTUs), cells and molecular markers (Fig. 1b). FTUs are defined as the smallest multicellular tissue organization that performs a unique physiologic function in concert with the surrounding microenvironment and is replicated multiple times in a whole organ[19]; examples are alveoli in the lung, crypts in the large intestine, glomeruli in the kidney and islets in the pancreas.

To this end, HuBMAP developed the Human Reference Atlas (HRA)[3], which aims to unify terminology and data infrastructure

across consortia, such as the Human Tumor Atlas Network (HTAN)[20] and Human Cell Atlas[21], and to name and spatially characterize major anatomical structures, cell types and biomarkers (ASCT+B). ASCT+B tables have been established for 26 organs (https://humanatlas.io/asctb-tables). For 14 organs, there exist anatomically correct 3D models of the male and female reference bodies (Fig. 1c, left). Additional organs will be added during the production phase, including the female reproductive system (ovary, fallopian tube, breast and uterus), eye, and lymphatic vessels (Fig. 1c, right of the two 3D reference bodies and https://humanatlas.io/3d-reference-library). ASCT+B table entities are linked to existing ontologies, Uberon[22] and Foundational Model of Anatomy for anatomical structures[23,24], Cell Ontology for cell types[25] and HUGO Gene Nomenclature for gene and protein biomarkers (https://www.genenames.org). Standardization across these domains is necessary for quality control and integrated analysis of ASCT+B across different systems in the body. Publication evidence exists for approximately 80% of the 2,582 anatomical structures, 898 cell types, 2,548 biomarkers and the 13,882 interlinkages between anatomical structures, cell types and biomarkers. More than 6,000 experimental datasets have been registered into this spatially and semantically explicit HRA framework via the Registration User Interface[19]. HuBMAP tools and methods are being used to map high-quality experimental data for healthy human adults from other consortia to increase the quality and coverage of the HRA. These include 47 kidney datasets from HuBMAP, 196 datasets from the Chan Zuckerberg CELLxGENE Discover Portal (https://cellxgene.cziscience.com), 25 datasets of 8 organs from the Genotype–Tissue Expression Portal[26] (https://gtexportal.org), 21 kidney datasets from the Kidney Precision Medicine Project[27] (KPMP) and 14 brain datasets from the Neuroscience Multiomic Archive portal (https://nemoarchive.org).

A web- and R-based application called Azimuth was developed to automate the processing, analysis and interpretation of scRNA-seq experiments using annotated reference datasets[2]. Azimuth performs normalization, visualization, cell annotation and differential expression (biomarker discovery) on a user-provided count matrix of gene expression in single cells. As of May 2023, Azimuth references exist for ten healthy human organs that can be explored using Vitessce[28]; these references have been used to automatically annotate approximately 280 million cells in about 20,000 datasets. For seven organs, standardized Organ Mapping Antibody Panels (https://humanatlas.io/omap) have been established that make it possible to map new experimental data to the evolving atlas[5].
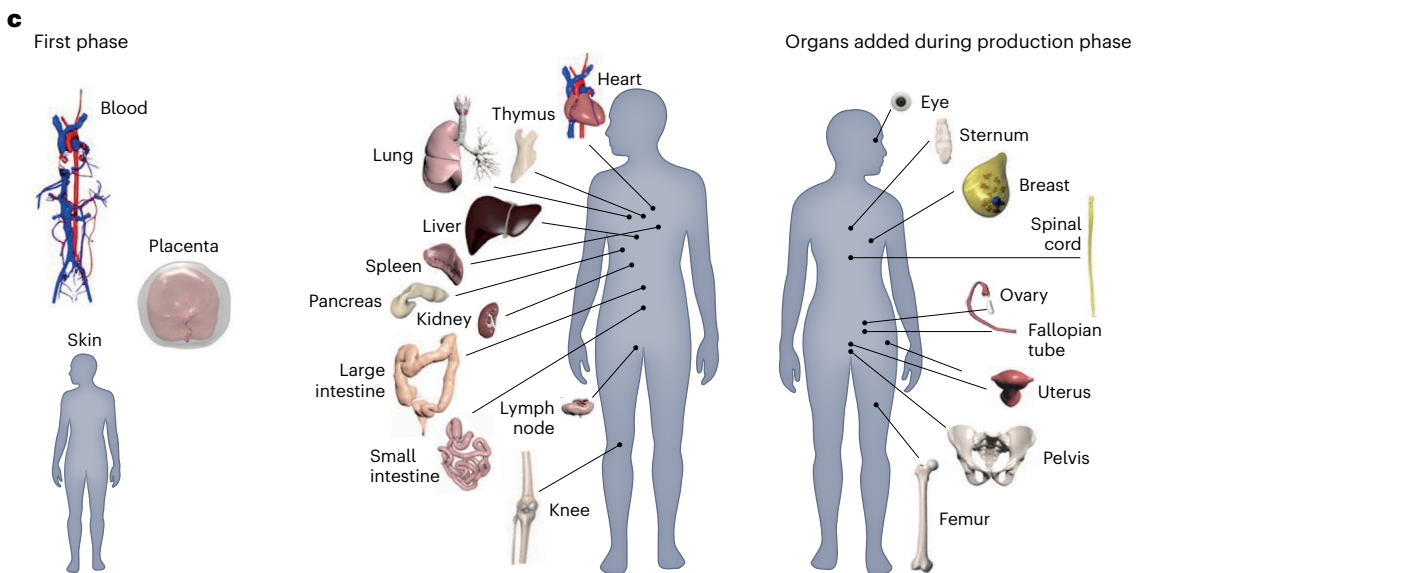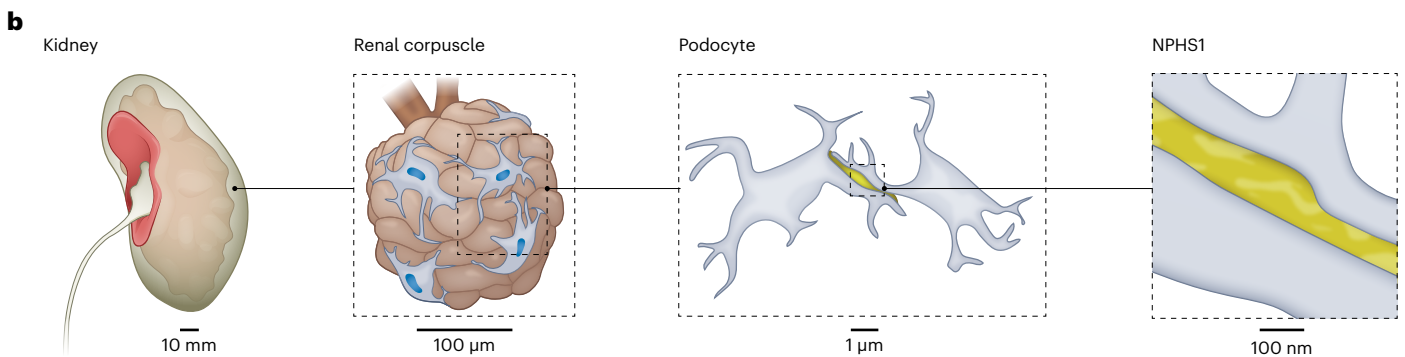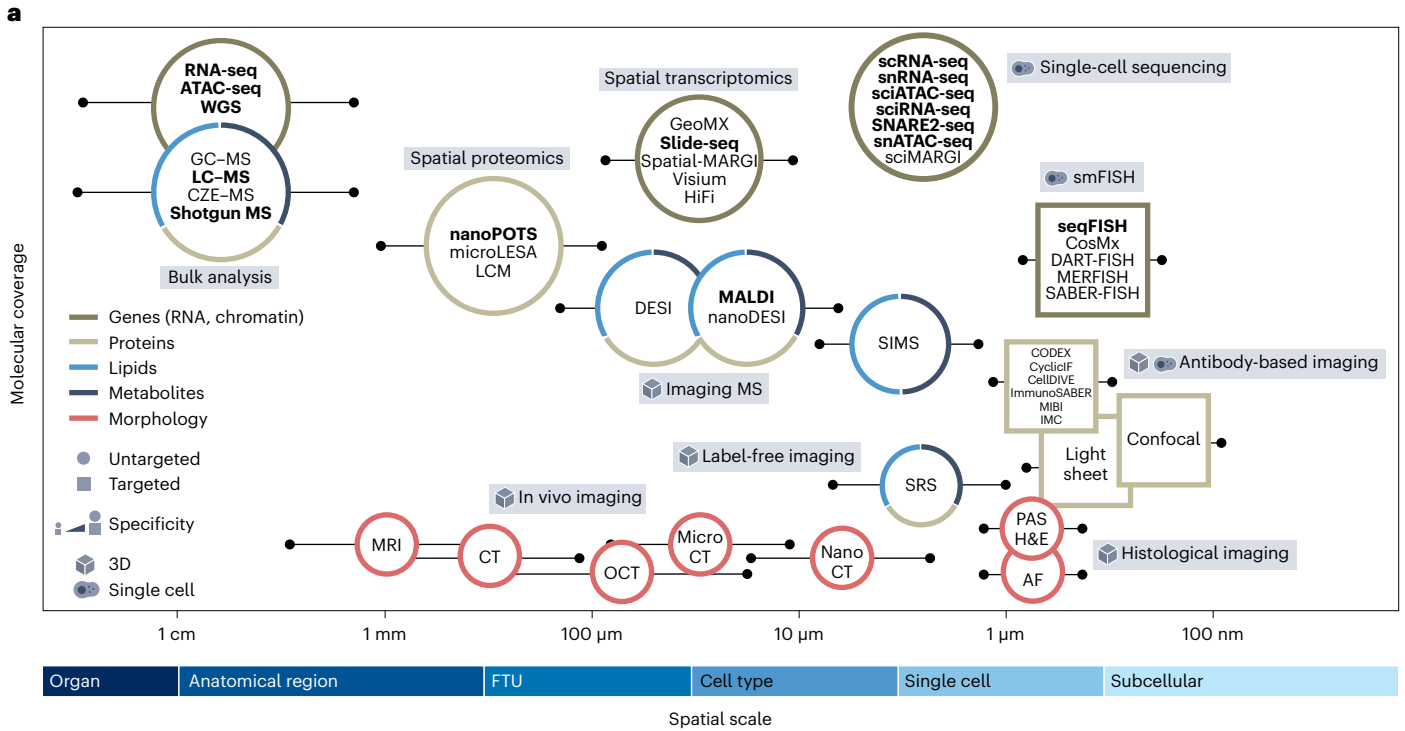
The utility of healthy HRA data for understanding human development, aging[29] and tissue function and dysfunction in health and disease has been showcased in several publications. Ref. 11 provides

---

Fig. 1 | Molecular coverage and spatial scale of different assay types. About 40 different analytical technologies are used in HuBMAP. **a**, Molecular coverage versus spatial scale. Data publicly available via the HuBMAP portal are rendered in bold. AF, autofluorescence; CellDIVE, an antibody-based multiplexed imaging technology; CT, computed tomography; CyclicIF, cyclic immunofluorescence; CZE–MS, capillary zone electrophoresis–mass spectrometry; DART-FISH, decoding amplified targeted transcripts with fluorescence in situ hybridization; DESI, desorption electrospray ionization; GC–MS, gas chromatography–mass spectrometry; GeoMX, a spatial omics technology from Nanostring; H&E, hematoxylin and eosin; IMC, imaging mass cytometry; ImmunoSABER, immunostaining with signal amplification by exchange reaction; LCM, laser capture microdissection; MALDI, matrix-assisted laser desorption/ionization; MERFISH, multiplexed error-robust fluorescence in situ hybridization; MIBI, multiplex ion beam imaging; MRI, magnetic resonance imaging; nanoDESI, nanospray desorption electrospray ionization; OCT, optical coherence tomography; PAS, Periodic Acid-Schiff; SABER-FISH, signal amplification by exchange reaction fluorescence in situ hybridization; sciATAC-seq, single-cell combinatorial indexing assay for transposase-accessible chromatin with sequencing; sciMARGI, single-cell combinatorial indexing mapping of RNA–

genome interactions; sciRNA-seq, single-cell combinatorial indexing RNA sequencing; seqFISH, sequential fluorescence in situ hybridization; Slide-seq, a spatial transcriptomics technology; SNARE2-seq, single-nucleus chromatin accessibility and mRNA expression sequencing; Spatial-MARGI, spatial mapping of RNA–genome interactions; SRS, stimulated Raman scattering. **b**, The multiscale HRA covers more than 1,500 anatomical structures in the male and female body. A zoom into the kidney (10 mm level) reveals a representative view of a renal corpuscle (100 μm level), a subsegment of one of the approximately one million FTUs (nephrons) of the kidney that is important in filtration. Podocytes, one of the cells important in filtration (micrometre level) with nuclei (in blue) and protein NPHS1 that maintains the structural integrity of the filtration barrier (yellow), are illustrated. **c**, Reference objects exist in 3D for 53 organs (counting left and right and female and male organs). Shown on the left of the 3D reference bodies are female and male organs for which HRA data exist on the HuBMAP portal; note that the placenta is full term; on the right of the reference bodies are female and male organs that will be added during the production phase. The 3D reference organs are used during tissue registration to automatically assign anatomical structure tags and serve as landmarks during the spatial search for tissue datasets with specific anatomical structures, FTUs, cell types or biomarkers.

high-quality data for the kidney Azimuth reference and demonstrates how healthy and diseased tissue can be compared. Ref. 9 provides linked single-cell multiome and spatial CODEX multiplexed imaging data to compare how cellular differentiation and organization change across the large and small intestine and code for hierarchical cell neighbourhood analysis and visualization. Ref. 10 provides the first spatio-temporal atlas of the human maternal–foetal interface and a statistical model of trophoblast invasion and vascular remodelling in
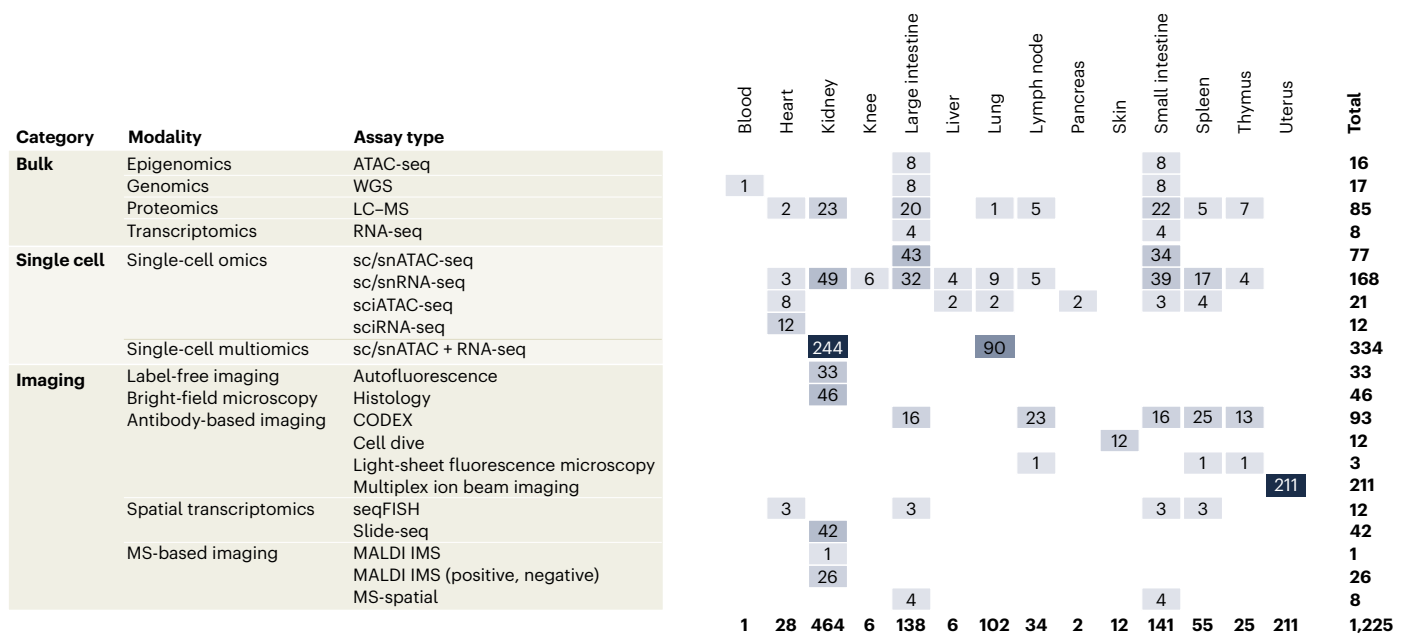
Category	Modality	Assay type	Blood	Heart	Kidney	Knee	Large intestine	Liver	Lung	Lymph node	Pancreas	Skin	Small intestine	Spleen	Thymus	Uterus	Total
Bulk	Epigenomics	ATAC-seq					8						8				**16**
	Genomics	WGS	1				8						8				**17**
	Proteomics	LC–MS		2	23		20		1	5			22	5	7		**85**
	Transcriptomics	RNA-seq					4						4				**8**
Single cell	Single-cell omics	sc/snATAC-seq					43						34				**77**
		sc/snRNA-seq		3	49	6	32	4	9	5			39	17	4		**168**
		sciATAC-seq			8			2	2		2		3	4			**21**
		sciRNA-seq			12												**12**
	Single-cell multiomics	sc/snATAC + RNA-seq					244		90								**334**
Imaging	Label-free imaging	Autofluorescence			33												**33**
	Bright-field microscopy	Histology			46												**46**
	Antibody-based imaging	CODEX					16			23			16	25	13		**93**
		Cell dive										12					**12**
		Light-sheet fluorescence microscopy							1					1	1		**3**
		Multiplex ion beam imaging														211	**211**
	Spatial transcriptomics	seqFISH		3			3						3	3			**12**
		Slide-seq			42												**42**
	MS-based imaging	MALDI IMS			1												**1**
		MALDI IMS (positive, negative)			26												**26**
		MS-spatial					4						4				**8**
			1	**28**	**464**	**6**	**138**	**6**	**102**	**34**	**2**	**12**	**141**	**55**	**25**	**211**	**1,225**

Fig. 2 | Organs and assay types publicly available via the HuBMAP portal (as of May 2023). Organs are sorted alphabetically. Assay types are grouped into bulk, single cell and imaging. LC–MS includes imaging mass cytometry (2D), imaging mass cytometry (3D), LC–MS, LC–MS bottom up and LC–MS top down. WGS, whole-genome sequencing.

the first half of human pregnancy. Ref. 7 provides the first 3D map of the skin showing shorter distances between immune and endothelial cells and higher immune cell densities in 3D versus 2D, highlighting the benefits of 3D for cell spatial analysis.

HuBMAP portal

All data generated in HuBMAP are made publicly available after rigorous quality assessment and control. A cloud-based database hosts raw experimental output as well as processed data and published downstream analyses. Extensive metadata concerning biospecimens, assays and protocols are available (https://portal.hubmapconsortium.org). Users can search the HuBMAP portal for donors, samples and datasets by sex, organ, data type and so on. Search results are displayed as a list with further filter options and data visualizations in Vitessce, with links to experimental, computational and metadata information. Users can run spatial searches using the Exploration User Interface[19] (https://portal.hubmapconsortium.org/ccf-eui) by placing a probing sphere to define a 3D space, enabling the exploration of tissue blocks and associated datasets in this space. Resulting datasets are listed on the right side of the Exploration User Interface together with provenance and donor metadata along with thumbnails indicating assay types such as CODEX or liquid chromatography–mass spectrometry (LC–MS). The spatial search can be performed across more than 1,500 anatomical structures and more than 6,000 spatially registered datasets to retrieve cell types commonly located in these spatial areas.

HuBMAP outreach

An important part of HuBMAP has been the dissemination of its data, code, methods and results. Methodologic and didactic resources have been made available to the community. As of May 2023, HuB-MAP published 215 protocols (https://www.protocols.io/workspaces/human-biomolecular-atlas-program-hubmap-method-development) and more than 20 standard operating procedures (https://humanatlas.io/standard-operating-procedures). More than 70 public talks, demos and other videos are available from HuBMAP members on its YouTube channel (https://www.youtube.com/@hubmapconsortium4358).

Biological insights

Several biological concepts and insights emerged from initial atlasing efforts including 2D and 3D cellular relationships in spatially resolved FTUs, rare or previously undocumented cell types and mechanisms associated with clinical outcomes, homeostasis and disease.

Unique microenvironments in healthy and diseased FTUs and remarkable cellular diversity associated with underlying pathobiology. A consistent theme has been that the 2D and 3D maps reveal a unique spatial association of parenchymal cells with different subtypes of immune or stromal cells in distinct FTUs of different tissues (Fig. 3).

For example, integrated analyses through an inter-consortium effort between HuBMAP and KPMP using sc/snRNA-seq, spatial transcriptomics of ~300,000 cells and 3D neighbourhoods of more than a million cells with imaging cytometry described unique immune and stromal cell microenvironments associated with healthy (reference or mature) and altered (non-reference) cell states[11]. Genes associated with altered tubular epithelium cell states were indicative of failed repair and associated with progression to kidney disease[11] (Fig. 3a). Molecular maps of metabolites and lipids of key FTUs spanning the entire nephron along the kidney cortico-medullary axis revealed distinct patterns associated with filtration, secretion, absorption, solute transport, glucose metabolism and water balance, indicative of the unique functional organization of the nephron[6].

In the human maternal–foetal interface atlas[10], changes in decidual composition and structure were dependent on both temporal and microenvironmental queues (Fig. 3b). For example, time-dependent changes in the frequency of natural killer (NK) cells, T cells and tolerogenic macrophages were sufficiently robust such that gestational age could be predicted based on these parameters alone. By contrast, the progression of spiral artery remodelling was locally regulated in the nearby tissue microenvironment by placentally derived extravillous trophoblasts (EVTs) that encircled and invaded each vessel. Using spatial transcriptomics, the investigators developed an integrated model of spiral artery remodelling supporting an intravasation mechanism in which EVT vessel invasion is accompanied by the upregulation of pro-angiogenic, immunoregulatory programmes, promoting
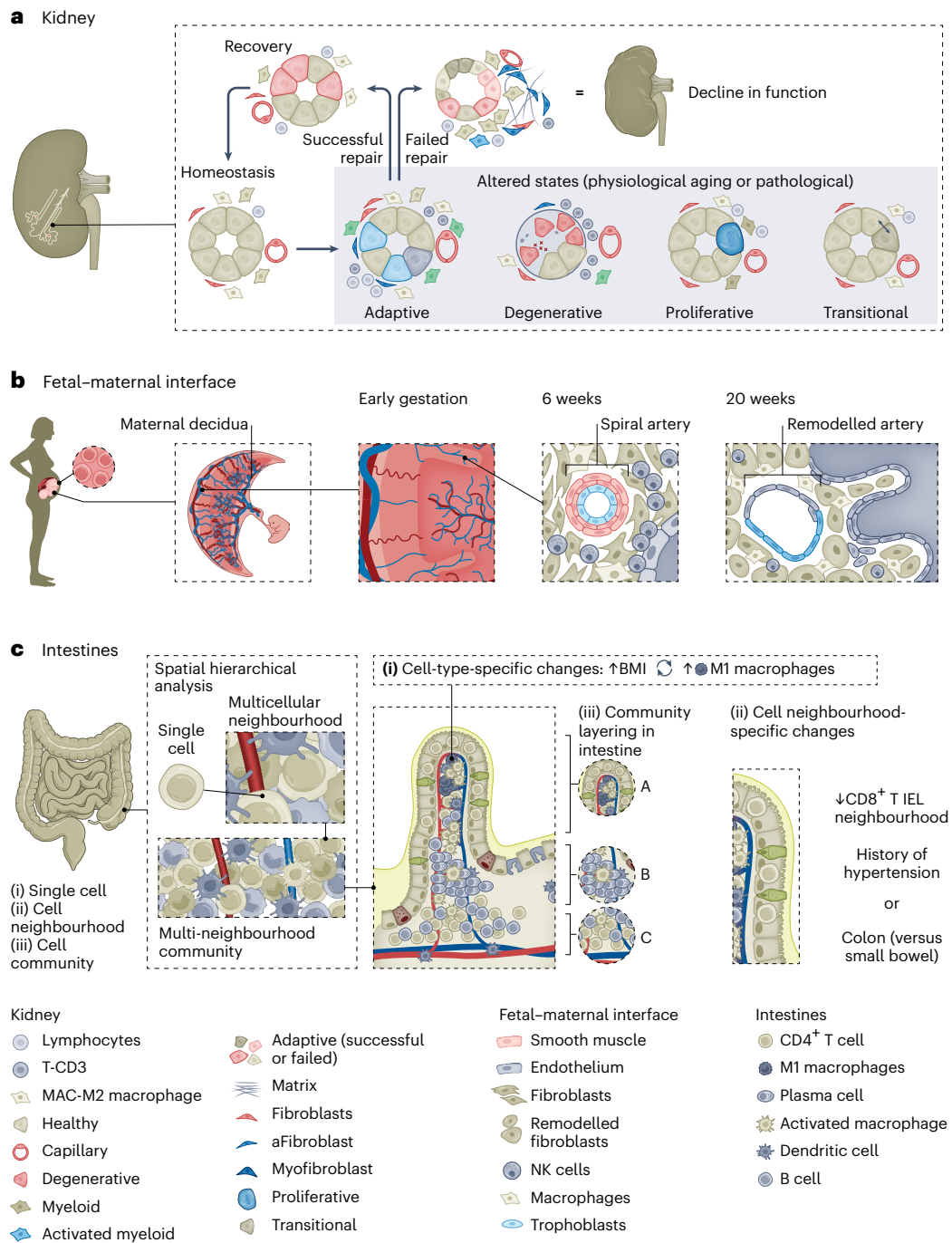
Fig. 3 | Unique FTU neighbourhoods in different biological contexts in the human body. Illustrated are three themes demonstrating functional and neighbourhood relationships revealed through HuBMAP atlas data. **a**, Remarkable differences in the interstitial neighbourhood of kidney tubules in homeostasis and altered states. Single-cell analysis of healthy and diseased kidneys uncovered healthy and several altered cell states. Spatial analysis (2D and 3D) led to the discovery of differential enrichment of immune and stromal cells in adaptive (successful or failed repair of tubules) and degenerative (severely injured cells with degenerative changes) states. The tubules with failed repair show enrichment of macrophages, adaptive fibroblasts (aFibroblasts) and myofibroblasts with fibrosis due to collagen deposition, whereas degenerative tubule cells associate more with CD3[+] lymphocytes. **b**, A distinct immune cell neighbourhood of immature and remodelled spiral arteries at the maternal–foetal interface at different gestational ages coincides with maturing placental villi. As gestation progresses, the immune cell composition changes from high NK cells and low macrophages to high macrophages and low NK cells, a pattern

that can ascertain gestational age and is associated with paracrine interactions between the placental villous trophoblasts and spiral artery endothelial cell-mediated remodelling. **c**, The healthy human intestine was analysed at the single-cell level with spatial resolution using a multi-hierarchical approach to define cellular neighbourhoods and multi-neighbourhood communities. Biological insights were obtained at three scales: (i) at the cell-type level, M1 macrophages positively correlate with body mass index; (ii) at the neighbourhood level, the 'CD8[+] T cell intraepithelial lymphocyte' multicellular neighbourhood decreased in the colon compared with that in the small intestine and for patients with a history of hypertension; (iii) at the cellular community level, distinct cellular communities were identified and found to be layered from the submucosa to the lumen, driven by distinct compositions of epithelial, immune and mesenchymal cells in these communities. These illustrations relay concepts and are not intended to be truly reflective of actual biological scale and cellular composition. Figure elements have been created with BioRender.com.

interactions with vascular endothelium while avoiding the activation of circulating maternal immune cells. Together, these results support a coordinated model in which gestational age drives a transition to an immune-permissive niche that is conducive to decidual invasion and vascular remodelling by genetically foreign EVTs.

Similarly, the intestine atlas[9] provides a detailed overview of the relationships of epithelial, mesenchymal and immune cells across spatial hierarchies (Fig. 3c). At the cellular level, CD8+ T cell counts decreased from the small intestine to the large intestine, although primarily within the 'CD8+ T cell-enriched intraepithelial lymphocyte' neighbourhood. Interestingly, these intraepithelial-located CD8+ T cells were decreased in subjects with a history of hypertension. The investigators identified intestinal crypt neighbourhoods containing adaptive immune cells in the small and large intestine. These aligned with Paneth cell-enriched neighbourhoods in the small intestine, whereas in the large intestine, the stem cell crypt was identified with increasing neuroendocrine cell density toward the bottom of the crypt. Even beyond the stem cell zone, multi-neighbourhood communities of cells were found based on the composition of immune, epithelial and mesenchymal cells in each zone of the intestine and interestingly were distinctly layered with increasing proximity to the lumen of the intestine. These results reflect the complex and distinct cell-type organization in the human intestine.

Unique epigenetic profiles associated with cell states and FTUs. The kidney and intestine atlases interrogated the epigenetic regulation at cellular resolution and identified transcription factor binding site (TFBS) activities unique to different FTUs and cell states using chromatin accessibility studies[9,11]. As cells transitioned from healthy to altered states in the kidney, TFBS activities and corresponding gene regulatory networks and pathways changed. For example, TFBS motifs of the oestrogen-related receptor, a key regulator of the healthy thick ascending limb (TAL) gene network, were inactivated in the altered TAL cells of the kidney. Furthermore, TFBS activities of pathways associated with fibrosis (transforming growth factor beta, TGFβ) and inflammation (rel avian reticuloendotheliosis viral oncogene homolog A/nuclear factor kappa-light-chain-enhancer of activated B cells, REL/NF-κB) were increased in a maladaptive TAL cell state. In the intestine, ETS variant transcription factor 6 (ETV6) TFBS motifs were enriched in differentiated colon cells whereas achaete-scute family bHLH transcription factor 2 (ASCL2) TFBS motifs were enriched in more undifferentiated cells[9].

Translational potential of multimodal single-cell atlases. One of the key advantages of healthy atlas efforts is that the healthy state can be used as a benchmark to understand the progression to altered states due to aging, infections, injury or biological variables. These comparisons inform biomarkers, mechanisms and disease course. For example, joint analyses of healthy and diseased specimens in the kidney atlas identified cell states, neighbourhoods and genes in different FTUs that drive kidney injury and are associated with a decline in kidney function[11]. This work identified a senescence-associated phenotype in injured cells that likely drives maladaptive repair. Comparative analysis of the papilla cells of healthy individuals and patients with kidney stone disease identified increased expression of matrix metalloproteinases (MMP7 and MMP9), genes associated with cell injury and matrix remodelling, in a number of cell types in patients with active stone disease and in regions of mineralization in the papilla[8]. Furthermore, both of these secreted proteins were increased in the urine of patients and might become clinically informative markers of kidney stone disease[8]. In the intestine, M1 macrophages were associated with changes in body mass index, which potentially indicates an early stage of gastrointestinal disease[9].

Multimodal analysis enhances opportunities for discovery of cell types and empowers comparisons among organs. Multimodal interrogation of cellular and molecular features including different cell states and genes using different technologies (sc/snRNA-seq, sc/snATAC-seq, multiplex immunofluorescence and spatial transcriptomics) enabled the discovery of cell types and tissue sub-compartmentalization at the FTU level, orthogonal validation, spatial resolution of high-dimensional single-cell data and discovery of similarities among gene expression patterns across organs. While single-cell transcriptomics data helped uncover cellular diversity, detected rare cell types and associated expression profiles, concomitant analysis of specific regions of interest using CODEX identified a mucin 6 positive (MUC6+) mucous-producing cell type in a specific region of the intestine[9]. The transcription factor ETV6 was expressed in the colon, only in differentiated absorptive cells, and in the kidney, where its chromatin accessibility and expression levels were high in altered tubules and not in differentiated tubules[11].

Future steps, challenges and opportunities

The HuBMAP production phase (2022–2026) has several goals, many of which are unique to HuBMAP (Fig. 4). First, it will generate many reference datasets, including from several new technologies, with an emphasis on building 3D maps as well as on collecting data from diverse donors that represent a range of demographic features (sex, race or ethnicity, and age). Second, it will develop methods to overcome challenges and gaps in aligning outputs of the different technologies. Finally, it will improve metadata standards, analytical and visualization tools, and data integration and interpretation to enhance the tissue atlases of the human body through collaborative efforts. The production phase plans to deliver 3D multi-scale molecular tissue maps at single-cell (or even subcellular − nucleus, mitochondria) resolution from human organs, as well as across systems that span many organs throughout the entire body (for example, the vascular and lymphatic systems). We will adapt workflows for collecting and processing data at scale, as well as creating tools for the community to navigate and leverage the HuBMAP data and resources. HuBMAP aims to establish procedures and build resources to address several challenges that are expected in integrating diverse data types from many organs by generating robust, orthogonally validated spatial maps by standardizing nomenclature, metadata and identifying 'bridges or anchors' across assays for integrated analysis.

Expanding production and molecular technologies for comprehensive maps

HuBMAP is implementing cutting-edge pipelines that bring together single-cell and spatial transcriptomics workflows[30], as well as expanding existing assays (for example, multiplex ion beam imaging[31,32], highly multiplexed antibody-based imaging strategies) and adding imaging approaches. Newly added assays include highly accurate long high-fidelity (HiFi)-Slide sequencing[33], CosMx Spatial Molecular Imager (CosMx)[31] and Xenium (10x Genomics) spatially resolved transcriptomics, secondary ion mass spectrometry imaging[32], stimulated Raman spectroscopy[34,35], optical coherence tomography[36,37], spatially driven mass spectrometry technologies and quantitative proteome imaging[38]. Single-cell proteome and phosphoproteome analyses can be readily related to phenotype and are amenable to integration with other multiomic analyses including lipidomes and metabolomes. Spatial mapping protocols, datasets and analytical pipelines will be enhanced to achieve single-cell or near-single-cell resolution of several organs using these technologies.

One of the challenges in various atlas efforts is the need to standardize vocabularies and enhance interoperability for successful integration. To this end, HuBMAP will work to strengthen collaborations with other atlasing efforts (for example, HTAN[20] (National Institutes of Health; NIH), Cellular Senescence Network[29] (NIH), KPMP[27] (NIH), LungMAP[39] (NIH), Human Cell Atlas[21] (HCA, CZI), Human Protein Atlas[40], Genotype–Tissue Expression project[41] (NIH)), and create an ecosystem of data and code for mapping the human body at high resolution. These
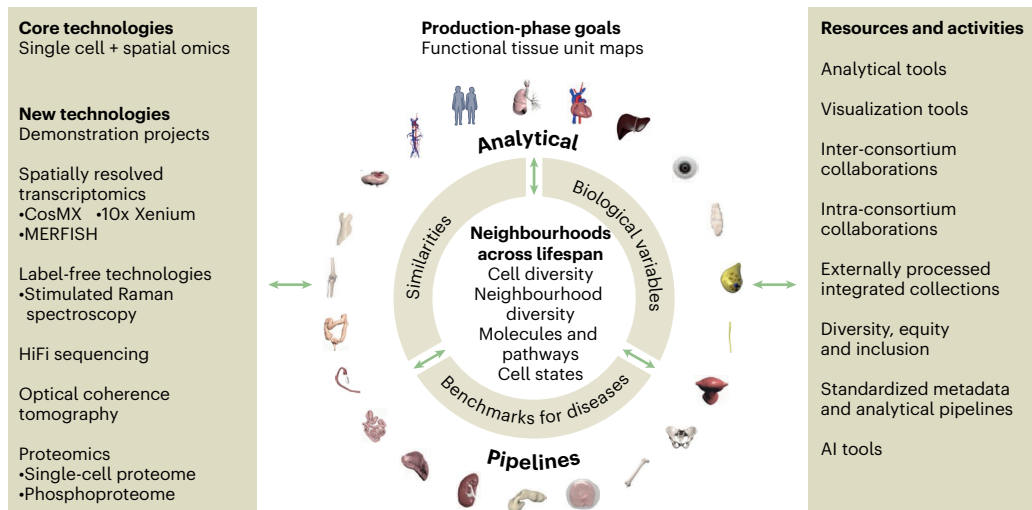
Fig. 4 | New technologies, resources, integrated knowledge base and mapping in the production phase. Left, all organs generate molecular data from core technologies that include sc/snRNA-seq, morphology, antibody-based protein expression and metabolomics. Several new technologies will be applied in addition. Right, resources for the community and atlasing efforts. Analytical tools include Azimuth for ATAC-seq and RNA-seq, cell segmentation and neighbourhood mapping. Visualization tools include enhancement of Vitessce for 2D and 3D viewing of single-cell and spatial data and tracking of specimens with associated data and ASCT+B on the HRA portal. Inter-consortium efforts include harmonizing nomenclature via ASCT+B tables and 2D and 3D anatomical structure references, tissue exchange and community tools for mapping and visualization. Intra-consortium efforts include collaborations among the various components of HuBMAP, such as common antibody or targeted transcriptome panels, and cross-organ comparative analysis such as vasculature or extracellular matrix. External processed integrated collections will provide a mechanism that imports external quality-control data into the HuBMAP portal and create opportunities for synergies in organ-mapping projects and discovery. Diversity,

equity and inclusion efforts are dedicated to establishing infrastructure to attract individuals from underrepresented and disadvantaged backgrounds and provide opportunities for cutting-edge research and to foster their career goals. Standardized data and analytical pipelines will enhance quality control by harmonizing metadata standards across tissue-mapping centres and technologies with detailed documentation of metadata, assays and analytical parameters. Artificial intelligence tools are resources in which users will be able to visualize omics data integrated with histology, quantify tissue components at the level of the cell state and predict gene or protein expression or cell identity from histology slides. Centre, the information from technologies and resources will be leveraged to create spatial maps of FTUs and 2D and 3D neighbourhoods across HuBMAP organs. The knowledge base created will enable comparative analysis of cell types and neighbourhoods across organs to identify similarities, understand how diversity (age, sex and race) affects FTU maps and create a benchmark reference atlas of FTUs for studying changes in disease and targets for therapies. Figure elements have been created with BioRender.com.

collaborations will align preanalytical metadata, tissue procurement, processing, preservation and analytical methods to enable running multiple assay types and data analysis workflows on the same tissue across institutions. The organs will be extensively sampled with a goal of 20–50 donors per organ using detailed protocols and operating procedures to ensure reproducibility. This will enhance quality control and facilitate the generation of comprehensive organ and body reference maps with several layers of orthogonal validation, increasing confidence in results that are essential for new biological insights, hypotheses and benchmarks for disease.

Data analysis, tools and integration across scales and modalities

HuBMAP is developing data integration and analytical approaches during its production phase. Machine learning and deep learning tools are being developed and applied to single-cell and spatial data to identify cell types, to match their spatially registered molecular profiles with morphology and to describe relationships with the microenvironment[42] (Fig. 4). Innovative approaches and tools are being developed to automate molecular and cell-type predictions and segmentation of morphological features in light-microscopy images[42–44]. HuBMAP data and analytical pipelines will enable the comparison of molecular profiles and neighbourhoods of the same general cell type across different regions and organs to identify shared functions and features for a given cell type across organs in homeostasis (Fig. 4). We will also focus on organ vasculature (veins, arteries and lymphatics) and rigorous evaluation of its spatial relationships between different cell types (epithelial, immune, stroma and so on). Defining these spatial relationships in conjunction with multiomic data will enable the identification of biologically relevant ligand–receptor interactions

in these microenvironments (for example, cells within glomeruli and their relationship with the extracellular matrix in kidneys) and help determine how the molecular and cellular functions for a given cell type compare across organs (for example, genes important in water transport across the kidneys, intestines and lungs). Furthermore, we will extend and generalize the automated discovery of cross-modality relationships and their capture into computational models for prediction[45] as well as interpretation. One example of the latter is the automated biomarker candidate discovery for FTUs, in which imaging mass spectrometry measurements covering hundreds of molecular species are integrated with coregistered microscopy-derived tissue annotations and interpretable supervised machine learning is used to automatically discern FTU-specific molecular markers[46]. An important aspect of the HuBMAP production phase will be help to determine how the diverse donor demographics affect cellular diversity and spatial relationships. Thus, efforts will be made to collect samples from diverse populations through enhanced local or consortium-wide procurement efforts and share them across groups for atlas construction and comparative analyses. Single donors contributing multiple organs will be pursued to enable comparisons of tissues from the same donor.

Enabling broad utility for different users, education and outreach

The production phase of HuBMAP is committed to improving data access and utilization tools (Fig. 4). These include enhanced spatial data registration and indexing of all tissue datasets using anatomical structure tags. Single-cell and FTU segmentation workflows will be produced and automated for all major organs. Azimuth and Organ Mapping Antibody Panels will be expanded to include all organs and

also within existing mapped organs to include cell states and associated biomarkers in single cells, FTUs and anatomical structures. In this way, HuBMAP experimental data will be linked (spatially and semantically) to the HRA and atlas user interfaces. A vasculature-based common coordinate framework will be developed to facilitate comparison and coordination across tissues[3,47]. As splicing results in multiple forms of protein, or proteoforms, from a single gene, the concept of proteoform biomarkers will be implemented to capture this greater proteome diversity in the HRA, recently shown to correlate strongly with cell types[48]. Importantly, it will be possible to explore the resulting data and ontology linkages not only via visual interactive user interfaces[19,47], but also via application programming interfaces. The full integration of the HRA into the HuBMAP portal will make it possible to define the exact experimental data used to construct the HRA (full data provenance, transparency and reproducibility). In addition, it will allow querying different experimental data and results generated by multiple teams using a unified metadata schema, such as spatial data, cell, FTU annotation and segmentation, and donor information. This infrastructure will enable a clinician or scientist to readily localize the site of action of a drug target in multiple organs of diverse populations. In addition, we envision HuBMAP data and tools becoming an important part of general education both in classrooms and for public engagement.

A resource for understanding development, aging and disease

The HuBMAP knowledge base and tools are expected to both catalyse and promote the generation of biological and clinical hypotheses as well as provide insights into human biology and disease. For example, technologies applied to biological samples covering lifespan will allow investigations of individual cells and neighbourhoods in healthy aging. Such information can be integrated with projects such as Cellular Senescence Network[29], which aims to define and map human senescent cells across human organs (https://sennetconsortium.org). Using the HuBMAP reference, one can better understand changes in the context of acute and chronic diseases (for example, diabetes, cardiovascular disease) and during the human lifespan. For example, KPMP[27] aims to understand the molecular basis of acute kidney injury and chronic kidney disease at single-cell resolution and help glean insights into the molecular trajectories and changes in neighbourhoods during transition from healthy to disease states.

HuBMAP will serve as a valuable reference for cancer atlas maps. Tumour formation, progression and treatment are influenced by the cellular environment[49]. 2D and 3D single-cell maps are being generated by the HTAN and others and rely on a normal-tissue reference for comparison. These maps are built on both tumour tissue and early-stage precancerous samples (for example, ductal carcinoma in situ and precancerous polyps). In addition to spatial relationships, such maps reveal early events involved in cancer formation or progression, such as the presence of stem cells or transcriptional regulatory events[50].

Another area of opportunity is to expand beyond HuBMAP's focus on adult tissues to include foetal and paediatric organs, which are at earlier stages of development with different physiology and biology than adult organs. Inclusion of comprehensive and longitudinal foetal and paediatric atlases of various organs will help in the understanding of physiological maturation, aging and progression of disease, thereby providing a holistic perspective of human biology across the entire lifespan.

Demonstration use cases

In the production phase, HuBMAP initiated demonstration projects highlighting the utility of HuBMAP data. These include (1) investigating mitochondrial DNA variant accumulation with aging, (2) building the human extracellular matrix atlas, (3) identifying organotypic and disease-specific vascular cell populations and (4) reverse engineering the extracellular neighbourhood to restore ovarian function. As an example, somatic mitochondrial DNA (mtDNA) mutations accumulate with age and contribute to aging-related diseases such as immune disorders, cardiovascular diseases and neurodegeneration[51,52]. Whether accumulation of specific mtDNA variants during aging occurs similarly across different cell types, organs, sexes and races is unclear. Demonstration project 1 aims to analyse mtDNA information in human single-cell genomics datasets to obtain insights into how cell or organ type, aging, sex, race and other factors impact mtDNA and how mtDNA variants affect nuclear gene expression and cell function.

Conclusions

In the next 3 years, HuBMAP will develop and validate analytical and visualization tools and apply emerging technologies, generating a wealth of molecular data spanning major aspects of the central dogma (DNA, RNA, protein) and defining cellular phenotypes and FTU neighbourhoods that will serve as benchmarks to understand homeostasis, aging and disease. This granular characterization of cells, including their chromatin, transcriptomic, proteomic and metabolomic states, and spatially resolved molecular maps of FTU microenvironments, will define the biologically relevant cellular and extracellular communities for each tissue and organ, providing opportunities to explore interrelatedness across organ systems. Such knowledge will be critical in understanding disease or syndromes that affect multiple organs and help design better informed and more precise drug targets (Fig. 4). The molecular landscape of human tissues in biologically diverse populations across the lifespan created in HuBMAP will provide instrumental insights for tissue engineering and regenerative medicine. The atlas of each organ will provide an important healthy reference to identify molecules associated with altered cell states that are linked to disease or repair defects. Scientists can use the HuBMAP data to develop experimental models to further understand the underlying biology and mechanisms. HuBMAP data and resources will also enable the discovery of relevant autocrine and short- and long-range paracrine effects that regulate organ crosstalk necessary for homeostasis and potentially impact clinical practice in predicting disease outcomes. To fully realize the potential of these datasets, we must continue to improve access to data and resources with minimum restrictions and to develop methods that use minimal amounts of tissue for interrogation by multiple single-cell and spatial technologies and application to clinical specimens.

Code availability

The HuBMAP Consortium GitHub at https://github.com/hubmapconsortium has 120 repositories that support data ingest, analysis, visualization and search plus HRA construction and usage. Documentation is available at https://software.docs.hubmapconsortium.org/apis.html.

References

1. Snyder, M. P. et al. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).
2. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
3. Börner, K. et al. Anatomical structures, cell types and biomarkers of the Human Reference Atlas. *Nat. Cell Biol.* **23**, 1117–1128 (2021).
4. 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).
5. Jain, S. Organ Mapping Antibody Panels (OMAPs): a community resource for standardized multiplexed imaging. *Nat. Methods* (in the press).
6. Neumann, E. K. et al. A multiscale lipid and cellular atlas of the human kidney. Preprint at *bioRxiv* https://doi.org/10.1101/2022.04.07.487155 (2022).
7. Ghose, S. et al. Human digital twin: automated cell type distance computation and 3D atlas construction in multiplexed skin biopsies. Preprint at *bioRxiv* https://doi.org/10.1101/2022.03.30.486438 (2022).

8. Canela, V. H. et al. A spatially anchored transcriptomic atlas of the human kidney papilla identifies significant immune injury and matrix remodeling in patients with stone disease. *Nat. Commun.* https://doi.org/10.1038/s41467-023-38975-8 (2023).

9. Hickey, J. et al. High resolution single cell maps reveals distinct cell organization and function across different regions of the human intestine. *Nature* https://doi.org/10.1038/s41586-023-05915-x (2023).

10. Greenbaum, S. et al. Spatiotemporal coordination at the maternal–fetal interface promotes trophoblast invasion and vascular remodeling in the first half of human pregnancy. *Nature* https://doi.org/10.1038/s41586-023-06298-9 (2023).

11. Blue, B. L. et al. An atlas of healthy and injured cell states and niches in the human kidney. *Nature* https://doi.org/10.1038/s41586-023-05769-3 (2023).

12. Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).

13. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

14. Guilliams, M. et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379–396.e38 (2022).

15. Stelzer, E. H. K. Light-sheet fluorescence microscopy for quantitative biology. *Nat. Methods* **12**, 23–26 (2015).

16. Black, S. et al. CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat. Protoc.* **16**, 3802–3835 (2021).

17. Neumann, E. K. et al. Protocol for multimodal analysis of human kidney tissue by imaging mass spectrometry and CODEX multiplexed immunofluorescence. *STAR Protoc.* **2**, 100747 (2021).

18. Spraggins, J. M. et al. High-performance molecular imaging with MALDI trapped ion-mobility time-of-flight (timsTOF) mass spectrometry. *Anal. Chem.* **91**, 14552–14560 (2019).

19. Börner, K. et al. Tissue registration and exploration user interfaces in support of a human reference atlas. *Commun. Biol.* **5**, 1369 (2022).

20. Rozenblatt-Rosen, O. et al. The Human Tumor Atlas Network: charting tumor transitions across space and time at single-cell resolution. *Cell* **181**, 236–249 (2020).

21. 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).

22. 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).

23. Rosse, C. & Mejino, J. L. V. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J. Biomed. Inform.* **36**, 478–500 (2003).

24. 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).

25. Meehan, T. F. et al. Logical development of the cell ontology. *BMC Bioinform.* **12**, 6 (2011).

26. Eraslan, G. et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science* **376**, eabl4290 (2022).

27. de Boer, I. H. et al. Rationale and design of the Kidney Precision Medicine Project. *Kidney Int.* **99**, 496–510 (2021).

28. Manz, T. et al. Viv: multiscale visualization of high-resolution multiplexed bioimaging data on the web. *Nat. Methods* **19**, 515–516 (2022).

29. Lee, P. J. et al. NIH SenNet Consortium to map senescent cells throughout the human lifespan to understand physiological health. *Nat. Aging* **2**, 1090–1100 (2022).

30. Keren, L. et al. MIBI-TOF: a multiplexed imaging platform relates cellular phenotypes and tissue structure. *Sci. Adv.* **5**, eaax5851 (2019).

31. He, S. et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* **40**, 1794–1806 (2022).

32. Tian, H. et al. Successive high-resolution $(H_2O)_n$-GCIB and C_{60}-SIMS imaging integrates multi-omics in different cell types in breast cancer tissue. *Anal. Chem.* **93**, 8143–8151 (2021).

33. Xu, T.-Y., Calandrelli, R., Lin, P., Zhu, D. & Zhong, S. HiFi-Slide spatial RNA-sequencing. *Protocol.io* https://doi.org/10.17504/protocols.io.3byl4by18vo5/v1 (2022).

34. Fung, A. A. & Shi, L. Mammalian cell and tissue imaging using Raman and coherent Raman microscopy. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **12**, e1501 (2020).

35. Zhang, W. et al. Multi-molecular hyperspectral PRM-SRS imaging. Preprint at *bioRxiv* https://doi.org/10.1101/2022.07.25.501472 (2022).

36. Balaratnasingam, C. et al. Histologic and optical coherence tomographic correlates in drusenoid pigment epithelium detachment in age-related macular degeneration. *Ophthalmology* **124**, 644–656 (2017).

37. Spaide, R. F. & Curcio, C. A. Anatomical correlates to the bands seen in the outer retina by optical coherence tomography: literature review and model. *Retina* **31**, 1609 (2011).

38. Zhu, Y. et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **9**, 882 (2018).

39. 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).

40. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

41. Lonsdale, J. et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

42. Brbić, M. et al. Annotation of spatially resolved single-cell data with STELLAR. *Nat. Methods* **19**, 1411–1418 (2022).

43. Govind, D. et al. PodoSighter: A cloud-based tool for label-free podocyte detection in kidney whole-slide images. *J. Am. Soc. Nephrol.* **32**, 2795–2813 (2021).

44. Lutnick, B. et al. A user-friendly tool for cloud-based whole slide image segmentation with examples from renal histopathology. *Commun. Med.* **2**, 105–109 (2022).

45. Van de Plas, R., Yang, J., Spraggins, J. & Caprioli, R. M. Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping. *Nat. Methods* **12**, 366–372 (2015).

46. Tideman, L. E. M. et al. Automated biomarker candidate discovery in imaging mass spectrometry data through spatially localized Shapley additive explanations. *Anal. Chim. Acta* **1177**, 338522 (2021).

47. 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).

48. Melani, R. D. et al. The Blood Proteoform Atlas: a reference map of proteoforms in human hematopoietic cells. *Science* **375**, 411–418 (2022).

49. Mackenzie, N. J. et al. Modelling the tumor immune microenvironment for precision immunotherapy. *Clin. Transl. Immunol.* **11**, e1400 (2022).

50. Becker, W. R. et al. Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. *Nat. Genet.* **54**, 985–995 (2022).

51. Wallace, D. C. Mitochondrial DNA variation in human radiation and disease. *Cell* **163**, 33–38 (2015).

52. Wallace, D. C. Mitochondrial genetic medicine. *Nat. Genet.* **50**, 1642–1649 (2018).

Author contributions

S.J., L.P., J.M.S., K.B. and M.P.S. conceived the idea and wrote the first draft of the manuscript and revisions. M.A., J.P.C., N.G., F.G., J.P.G., J.S.H., J.W.H., N.L.K., L.C.L., S.L., Y.L., H.L., A.N., E.S.N., W.-J.Q., A.R., P.R., B.R.S., R.V., I.S.V. and M.Z. contributed to manuscript revision.

Competing interests

The authors declare the following competing interests. F.G. is an employee of GE Research. B.R.S. is an inventor on patents and patent applications involving small-molecule-drug discovery and the 3F3 anti-Ferroptotic Membrane (3F3-FMA) antibody; co-founded and serves as a consultant to Inzen Therapeutics, Nevrox, Exarta Therapeutics and ProJenX; and serves as a consultant to Weatherwax Biotechnologies and Akin Gump Strauss Hauer and Feld. M.P.S. is a co-founder and an advisory board member of Personalis, Qbio, January AI, Mirvie, Filtricine, Fodsel, Lollo and Protos. I.S.V. consults for Guidepoint Global, Cowen, Mosaic and NextRNA. N.G. is a co-founder and an equity owner of Datavisyn. H.L. is a co-founder and an equity owner of ExoMira Medicine. The remaining authors declare no competing interests.

Additional information

Correspondence should be addressed to Sanjay Jain, Liming Pei, Jeffrey M. Spraggins, Katy Börner or Michael P. Snyder.

Peer review information *Nature Cell Biology* thanks the anonymous reviewers for their contribution to the peer review of this work.

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

[1]Department of Medicine, Washington University School of Medicine, St Louis, MO, USA. [2]Department of Pediatrics, Washington University School of Medicine, St Louis, MO, USA. [3]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA. [4]Center for Mitochondrial and Epigenomic Medicine, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, and Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [5]Department of Cell and Developmental Biology and the Mass Spectrometry Research Center, Vanderbilt University School of Medicine, Nashville, TN, USA. [6]Department of Pathology, Stanford School of Medicine, Stanford, CA, USA. [7]Texas Advanced Computing Center, University of Texas at Austin, Austin, TX, USA. [8]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [9]GE Research Center, Niskayuna, NY, USA. [10]Department of Intelligent Systems, Delft University of Technology, Delft, Netherlands. [11]Department of Pediatrics (Pulmonology) and Program for Rare and Interstitial Lung Disease, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [12]Department of Microbiology and Immunology, Stanford University, Stanford, CA, USA. [13]Departments of Medicine, Chemistry and Molecular Biosciences, Northwestern University, Evanston, IL, USA. [14]Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Diego, La Jolla, CA, USA. [15]Division of Cardiology, University of Washington School of Medicine, Seattle, WA, USA. [16]Department of Surgery, Washington University School of Medicine, St Louis, MO, USA. [17]Departments of Pharmacology, Medicine (Hematology and Oncology), Lurie Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. [18]Department of Physiology and Biophysics, University of Illinois at Chicago, Chicago, IL, USA. [19]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA. [20]Lymphocyte Biology Section and Center for Advanced Tissue Imaging, Laboratory of Immune System Biology, NIAID, NIH, Bethesda, MD, USA. [21]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA. [22]Department of Biological Sciences and Department of Chemistry, Columbia University, New York, NY, USA. [23]Delft Center for Systems and Control, Delft University of Technology, Delft, Netherlands. [24]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [25]Spatial Technologies Unit, Harvard Medical School Initiative for RNA Medicine, Department of Pathology, Beth Israel Deaconess Medical Center, and Harvard Medical School, Boston, MA, USA. [26]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA. [27]Department of Intelligent Systems Engineering, Indiana University, Bloomington, IN, USA. [28]Department of Genetics, Stanford School of Medicine, Stanford, CA, USA. [79]These authors contributed equally: Sanjay Jain, Liming Pei, Jeffrey M. Spraggins, Katy Börner, Michael P. Snyder. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: sanjayjain@wustl.edu; lpei@pennmedicine.upenn.edu; jeff.spraggins@vanderbilt.edu; katy@indiana.edu; mpsnyder@stanford.edu

HuBMAP Consortium

Kyung Jin Ahn[29], Jamie Allen[5], David M. Anderson[30], Christopher R. Anderton[26], Christine Curcio[31], Alessia Angelin[4], Constadina Arvanitis[17], Lyla Atta[32], Demi Awosika-Olumo[33], Amir Bahmani[34], Huajun Bai[29], Karol Balderrama[35], Leandro Balzano[36], Gautam Bandyopadhyay[37], Shovik Bandyopadhyay[38], Ziv Bar-Joseph[39], Kurt Barnhart[38], Daria Barwinska[40], Michael Becich[41], Laren Becker[34], Winston Becker[34], Kenneth Bedi[38], Sean Bendall[34], Kathy Benninger[42], David Betancur[42], Keith Bettinger[34], Sunteasja Billings[42], Philip Blood[42], Daniel Bolin[43], Samuel Border[36], Marc Bosse[34], Lisa Bramer[26], Maya Brewer[44], Maigan Brusko[36], Andreas Bueckle[43], Karl Burke[41], Kristin Burnum-Johnson[26], Eugene Butcher[34], Elizabeth Butterworth[36], Long Cai[45], Riccardo Calandrelli[46], Michael Caldwell[13], Martha Campbell-Thompson[36], Dongfeng Cao[31], Ivan Cao-Berg[42], Richard Caprioli[30], Chiara Caraccio[34], Anita Caron[47], Megan Carroll[42], Chrystal Chadwick[9], Angela Chen[48], Derek Chen[34], Fei Chen[35], Haoran Chen[39], Jing Chen[36], Li Chen[49], Lu Chen[50], Kenneth Chiacchia[42], Sanghee Cho[9], Peter Chou[51], Lisa Choy[8], Cecilia Cisar[52], Geremy Clair[19], Laura Clarke[53], Kelly A. Clouthier[44], Madeline E. Colley[5], Kristin Conlon[1], John Conroy[8], Kevin Contrepois[34], Anthony Corbett[37], Alex Corwin[54], Daniel Cotter[34], Elise Courtois[21], Aaron Cruz[41], Christopher Csonka[42], Kimberley Czupil[52], Vicky Daiya[43], Kali Dale[55], Shakeel Ahamed Davanagere[43], Monica Dayao[39], Mark P. de Caestecker[44], Aubrianna Decker[22], Stephen Deems[42], David Degnan[26], Tushar Desai[34], Vikrant Deshpande[43], Gail Deutsch[56], Michelle Devlin[42], Dinh Diep[45,46,57], Carla Dodd[32], Sean Donahue[39], Weixiu Dong[58], Rafael dos Santos Peixoto[32], Michael Duffy[38], Martin Dufresne[5], Thu Elizabeth Duong[46], Jennifer Dutra[37], Michael T. Eadon[40], Tarek M. El-Achkar[40], Archibald Enninful[59], Gokcen Eraslan[35], Diane Eshelman[42], Almudena Espin-Perez[34], Edward D. Esplin[34], Allison Esselman[60], Louis D. Falo[61], Louis Falo[41], Jean Fan[32], Rong Fan[59], Melissa A. Farrow[30], Negin Farzad[59], Patricia Favaro[34], Jamie Fermin[36], Ferda Filiz[34], Shane Filus[42], Kathleen Fisch[14], Eyal Fisher[62], Stephen Fisher[38], Katelyn Flowers[35], William F. Flynn[21], Agnes B. Fogo[63], Dongtao (Ann) Fu[36], James Fulcher[26], Anthony Fung[46], Derek Furst[41], Michael Gallant[43], Fu Gao[59], Yu Gao[18], Kyle Gaulton[46], Joseph P. Gaut[3], James Gee[38], Reetika R. Ghag[1], Shila Ghazanfar[53], Soumya Ghose[9], Debora Gisch[40], Ilan Gold[8], Aashay Gondalia[43], Brittney Gorman[26], William Greenleaf[34], Noah Greenwald[34], Brian Gregory[38], Rong Guo[38], Rajat Gupta[58], Hunter Hakimian[36], Jeff Haltom[4], Marc Halushka[32], Kyu Sang Han[32], Casey Hanson[34], Pehr Harbury[34], Josef Hardi[34], Linda Harlan[36], Raymond C. Harris[44], Austin Hartman[53], Elyas Heidari[64], Jesse Helfer[41], David Helminiak[51], Martin Hemberg[58], Nathaniel Henning[13], Bruce W. Herr II[43], Jonhan Ho[61], Jeanne Holden-Wiltse[37], Seung-Hyun Hong[65], Young-Kwon Hong[66], Brendan Honick[42], Greg Hood[42], Po Hu[29], Qiwen Hu[58], Molly Huang[14], Heidie Huyck[37], Tamjid Imtiaz[38], Olof Gerdur Isberg[5], Maxim Itkin[38], Dana Jackson[15], Marni Jacobs[14], Yashvardhan Jain[43], David Jewell[36], Lihua Jiang[34], Zhenghui G. Jiang[48], Sarah Johnston[38], Pujan Joshi[65], Yingnan Ju[43], Audra Judd[5], Adam Kagel[34], Ari Kahn[33], Nikolaos Kalavros[48], Kian Kalhor[46], Dimitra Karagkouni[48], Thomas Karathanos[34], Arivarasan Karunamurthy[41], Suhas Katari[36], Heather Kates[36], Madhurima Kaushal[1], Nicholas Keener[52], Mark Keller[8], Mariah Kenney[42], Colin Kern[46], Peter Kharchenko[8], Junhyong Kim[38], Carl Kingsford[39], Jessica Kirwan[36], Vladimir Kiselev[39], Jocelyn Kishi[67], Reta Birhanu Kitata[26], Amanda Knoten[1], Charles Kollar[41], Praveen Krishnamoorthy[1], Angela R. S. Kruse[5], Kuang Da[38], Anshul Kundaje[34], Eric Kutschera[4], Yumi Kwon[19], Blue B. Lake[46,57], Samuel Lancaster[34], Jonah Langlieb[35], Roy Lardenoije[68], Monica Laronda[69], Julia Laskin[70], Ken Lau[71], Hayan Lee[34], Maria Lee[66], Mejeong Lee[65], Yulia Levites Strekalova[36], Dongshunyi Li[39], Jennifer Li[34], Jilong Li[35], Xiangtang Li[70], Zhi Li[46], Yen-Chen Liao[26], Tiffany Liaw[8], Pei Lin[46], Yulieh Lin[38], Scott Lindsay[14], Chunjie Liu[4], Yang Liu[59], Yuan Liu[46], Marie Lott[4], Martin Lotz[65], Lisa Lowery[9], Peiran Lu[4], Xinyue Lu[39], Nicholas Lucarelli[36], Xiaokang Lun[58], Zhifei Luo[46], Jian Ma[39], Evan Macosko[35], Mayank Mahajan[58], Libby Maier[43], Danika Makowski[29], Morad Malek[5], David Manthey[36], Trevor Manz[8], Kenneth Margulies[38], John Marioni[47], Matthew Martindale[43], Cayla Mason[14], Clayton Mathews[36], Peter Maye[65], Chuck McCallum[8], Elizabeth McDonough[9], Liz McDonough[9], Hannah Mcdowell[72], Morgan Meads[14], Miguel Medina-Serpas[36], Ricardo Melo Ferreira[40], Jeffrey Messinger[31], Kay Metis[41], Lukasz G. Migas[23], Brendan Miller[32], Sayat Mimar[36], Brittany Minor[1], Ravi Misra[37], Alsu Missarova[53], Christopher Mistretta[48], Roger Moens[10], Eric Moerth[8], Jeffrey Moffitt[73], Gesmira Molla[53], Matthew Monroe[19], Emma Monte[34], Mike Morgan[53], Daniele Muraro[74], Bob (Robert) Murphy[39], Evan Murray[35], Mark A. Musen[34], Ahmed Naglah[36], Chanond Nasamran[14], Taruna Neelakantan[22], Stephanie Nevins[34], Hieu Nguyen[38], Nam Nguyen[39], Tram Nguyen[8], Tri Nguyen[46], Deb Nigra[42], Michel Nofal[58], Garry Nolan[34], Gerald Nwanne[1], Martin O'Connor[34], Kenichi Okuda[75], Merissa Olmer[49], Kathleen O'Neill[38], Nancy Otaluka[48], Minxing Pang[38], Mana Parast[14], Ljiljana Pasa-Tolic[26], Benedict Paten[52], Nathan Heath Patterson[30], Ting Peng[4], Gesina Phillips[42], Mina Pichavant[34], Paul Piehowski[26], Hannah Pilner[56], Ellie Pingry[5], Yered Pita-Juarez[48], Sylvia Plevritis[34], Athanasios Ploumakis[48], Alison Pouch[38], Gloria Pryhuber[37], Juan Puerto[42], Danial Qaurooni[43], Ling Qin[38], Ellen M. Quardokus[43], Presha Rajbhandari[22], Rebecca Rakow-Penner[14], Ramalakshmi Ramasamy[21], David Read[15], Elizabeth G. Record[43], David Reeves[5], Allyson Ricarte[39], Ana Rodríguez-Soto[14], Alexander Ropelewski[42], Jean Rosario[38], Morla-Adames Roselkis[8], David Rowe[49], Tarun Kanti Roy[76], Matt Ruffalo[39], Nancy Ruschman[43], Angela Sabo[40], Nina Sachdev[35], Sinem Saka[67], Diane Salamon[1], Pinaki Sarder[36], Hiroshi Sasaki[67], Rahul Satija[53], Diane Saunders[44], Riley Sawka[29], Kevin Schey[30], Heidi Schlehlein[43], David Scholten[17], Sarah Schultz[39], Lauren Schwartz[38], Melissa Schwenk[41], Robin Scibek[42], Ayellet Segre[58], Matthew Serrata[67], Walter Shands[52], Xiaotao Shen[34], Jay Shendure[56], Holly Shephard[42], Lingyan Shi[46], Tujin Shi[26], Dong-Guk Shin[49], Bill Shirey[41], Max Sibilla[41], Michal Silber[38], Jonathan Silverstein[42], Derek Simmel[42], Alan Simmons[41], Dhruv Singhal[48], Santhosh Sivajothi[21], Thomas Smits[8], Francesca Soncin[46], Qi Song[39], Valentina Stanley[14], Tim Stuart[53], Hanquan Su[67], Pei Su[13], Xin Sun[46], Christine Surrette[9], Hannah Swahn[49], Kai Tan[29], Sarah Teichmann[74], Abhiroop Tejomay[43], George Tellides[59], Kathleen Thomas[38], Tracey Thomas[38], Marissa Thompson[42], Hua Tian[41], Leonoor Tideman[23], Cole Trapnell[56], Albert G. Tsai[34], Chia-Feng Tsai[26], Leo Tsai[48], Elizabeth Tsui[69], Tina Tsui[5], Jason Tung[34], Morgan Turner[8], Jackie Uranic[42], Eeshit Dhaval Vaishnav[35], Sricharan Reddy Varra[34], Vasyl Vaskivskyi[39], Dusan Velickovic[26], Marija Velickovic[26], Jamie Verheyden[46], Jessica Waldrip[41], Douglas Wallace[4], Xueyi Wan[46], Allen Wang[46], Fusheng Wang[50], Meng Wang[34], Shuoshuo Wang[48], Xuefei Wang[45], Clive Wasserfall[36], Leonard Wayne[36], James Webber[35], Griffin M. Weber[58], Bei Wei[34], Jian-Jun Wei[17], Annika Weimer[34], Joel Welling[42], Xingzhao Wen[46], Zishen Wen[39], MacKenzie Williams[36], Seth Winfree[77], Nicholas Winograd[78], Abashai Woodard[38], Devin Wright[43], Fan Wu[32], Pei-Hsun Wu[32], Qiuyang Wu[46], Xiaodong Wu[76], Yi Xing[4], Tiangyang Xu[46], Manxi Yang[70], Mingyu Yang[59],

Joseph Yap[48], Dong Hye Ye[70], Peng Yin[67], Zhou Yuan[41], Chi (Jina) Yun[45], Ali Zahraei[5], Kevin Zemaitis[26], Bo Zhang[1], Caibin Zhang[49], Chenyu Zhang[49], Chi Zhang[46], Kun Zhang[46,57], Shiping Zhang[4], Ted Zhang[39], Yida Zhang[58], Bingqing Zhao[34], Wenxin Zhao[46], Jia Wen Zheng[58], Sheng Zhong[46], Bokai Zhu[34], Chenchen Zhu[34], Diming Zhu[46], Quan Zhu[46] & Ying Zhu[19]

[29]Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA, USA. [30]Department of Biochemistry and the Mass Spectrometry Research Center, Vanderbilt University School of Medicine, Nashville, TN, USA. [31]University of Alabama at Birmingham, Birmingham, AL, USA. [32]Johns Hopkins University, Baltimore, MD, USA. [33]The University of Texas at Austin, Austin, TX, USA. [34]Stanford University, Stanford, CA, USA. [35]Broad Institute, Boston, MA, USA. [36]University of Florida, Gainesville, FL, USA. [37]University of Rochester Medical Center, Rochester, NY, USA. [38]University of Pennsylvania, Philadelphia, PA, USA. [39]Carnegie Mellon University, Pittsburgh, PA, USA. [40]Indiana University, Indianapolis, IN, USA. [41]University of Pittsburgh, Pittsburgh, PA, USA. [42]PSC (Pittsburgh Supercomputing Center)/CMU, Pittsburgh, PA, USA. [43]Indiana University, Bloomington, IN, USA. [44]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [45]California Institute of Technology, Pasadena, CA, USA. [46]University of California San Diego, San Diego, CA, USA. [47]European Bioinformatics Institute, Hinxton, UK. [48]Beth Israel Deaconess Medical Center, Boston, MA, USA. [49]University of Connecticut/Scripps, Mansfield, USA. [50]Stony Brook University, Stony Brook, NY, USA. [51]Marquette University, Milwaukee, WI, USA. [52]University of California Santa Cruz, Santa Cruz, CA, USA. [53]New York Genome Center, New York City, NY, USA. [54]GE Research Cente, Niskayuna, NY, USA. [55]Native BioData Consortium, Oklahoma City, OK, USA. [56]University of Washington, Seattle, WA, USA. [57]Altos Labs, San Diego Institute of Science, San Diego, CA, USA. [58]Harvard Medical School, Boston, MA, USA. [59]Yale University, New Haven, CT, USA. [60]Department of Chemistry, Vanderbilt University, Nashville, TN, USA. [61]Department of Dermatology, University of Pittsburgh, Pittsburgh, PA, USA. [62]University of Cambridge, Cambridge, UK. [63]Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA. [64]University of Zurich, Zurich, Switzerland. [65]University of Connecticut, Mansfield, CT, USA. [66]University of Southern California, Los Angeles, CA, USA. [67]Harvard University, Boston, MA, USA. [68]Delft University of Technology, Delft, Netherlands. [69]Lurie Children's Hospital, Chicago, IL, USA. [70]Purdue University, Purdue, IN, USA. [71]Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, TN, USA. [72]Northwestern University, Chicago, IL, USA. [73]Boston Children's Hospital, Boston, MA, USA. [74]Wellcome Sanger Institute, Hinxton, UK. [75]University of North Carolina, Chapel Hill, NC, USA. [76]University of Iowa, Iowa City, IA, USA. [77]University of Nebraska Medical Center, Omaha, NE, USA. [78]Pennsylvania State University, State College, PA, USA.