The human body at cellular resolution: the NIH Human Biomolecular Atlas Program

HuBMAP Consortium*

Transformative technologies are enabling the construction of three-dimensional maps of tissues with unprecedented spatial and molecular resolution. Over the next seven years, the NIH Common Fund Human Biomolecular Atlas Program (HuBMAP) intends to develop a widely accessible framework for comprehensively mapping the human body at single-cell resolution by supporting technology development, data acquisition, and detailed spatial mapping. HuBMAP will integrate its efforts with other funding agencies, programs, consortia, and the biomedical research community at large towards the shared vision of a comprehensive, accessible three-dimensional molecular and cellular atlas of the human body, in health and under various disease conditions.

The human body is an incredible machine. Trillions of cells, organized across an array of spatial scales and a multitude of functional states, contribute to a symphony of physiology. While we broadly know how cells are organized in most tissues, a comprehensive understanding of the cellular and molecular states and interactive networks resident in the tissues and organs, from organizational and functional perspectives, is lacking. The specific three-dimensional organization of different cell types, together with the effects of cell–cell and cell–matrix interactions in their natural milieu, have a profound impact on normal function, natural ageing, tissue remodelling, and disease progression in different tissues and organs. Recently, new technologies have enabled the molecular characterization of a multitude of cell types1–4 and mapping of their spatial relationships in complex tissues at unprecedented scale and single-cell resolution. These advances create the opportunity to build a high-resolution atlas of three-dimensional maps of human tissues and organs.

HuBMAP (https://commonfund.nih.gov/hubmap) is an NIH-sponsored program with the goals of developing an open framework and technologies for mapping the human body at cellular resolution as well as generating foundational maps for several tissues obtained from normal individuals across a wide range of ages. A previous NIH-sponsored project, GTEx5, examined DNA variants and bulk tissue expression patterns across approximately a thousand individuals, but HuBMAP is a distinct project focused on generating molecular maps that are spatially resolved at the single-cell level but using samples from a more limited number of people. To achieve these goals, HuBMAP has been designed as a cohesive and collaborative organization, with a culture of openness and sharing using team science-based approaches6. The HuBMAP Consortium (https://hubmapconsortium.org/) will actively work with other ongoing initiatives including the Human Cell Atlas7, Human Protein Atlas8, LifeTime (https://lifetime-fetflagship.eu/), and related NIH-funded consortia that are mapping specific organs (including the brain9, lungs (https://www.lungmap.net/), kidney (https://kpmp.org/about-kpmp/), and genitourinary (https://www.gudmap.org/) regions) and tissues (especially pre-cancer and tumours10, https://humanitumor-atlas.org), as well as other emerging programs.

HuBMAP organization and approaches

The HuBMAP consortium comprises members with diverse expertise (for example, molecular, cellular, developmental, and computational biologists, measurement experts, clinicians, pathologists, anatomists, biomedical and software engineers, and computer and data information scientists) and is organized into three components: (1) tissue mapping centres (TMCs); (2) HuBMAP integration, visualization and engagement (HIVE) collaborative components; and (3) innovative technologies groups (transformative technology development (TTD) and rapid technology implementation (RTI)) (Fig. 1). Throughout the program, HuBMAP will increase the range of tissues and technologies studied through a series of funding opportunities that have been designed to be synergistic with other NIH-funded and international efforts. In the later stages of HuBMAP, demonstration projects will be added to show the utility of the generated resources and, importantly, to engage the wider research community to analyse HuBMAP data alongside data from other programs or their own labs.

Tissue and data generation

The HuBMAP TMCs will collect and analyse a broad range of largely normal tissues, representing both sexes, different ethnicities and a variety of ages across the adult lifespan. These tissues (Fig. 2) include: (1) discrete, complex organs (kidney, ureter, bladder, lung, breast, small intestine and colon); (2) distributed organ systems (vasculature); and (3) systems comprising dynamic or motile cell types with distinct microenvironments (lymphatic organs: spleen, thymus, and lymph nodes). Tissue will be collected at precisely defined anatomical locations (when possible, photographically recorded) according to established protocols that preserve tissue quality and minimize degradation. Beyond meeting standard regulatory requirements, to the greatest extent possible, consent will be obtained so that the generated data is available for open-access data sharing (that is, public access without approval by data committees), to maximize their usage by the biomedical community.

To achieve spatially resolved, single-cell maps, the TMCs will use a complementary, iterative, two-step approach (Fig. 3). First, ’omic assays, which are extremely efficient for data acquisition, will be used to generate global genome sequence and gene expression profiles of dissociated single cells or nuclei in a massively parallel manner. The molecular state of each cell will be revealed by single-cell transcriptomic11 and, in many cases, chromatin accessibility12,13 assays; imitation of transcription

* A list of participants and their affiliations appears at the end of the paper.
tools change, and as HuBMAP interacts and collaborates with other techniques, storage and shipping conditions, assays, and data-processing standards and be interoperable with other platforms, such as the HCA Data Coordination Platform, making it possible to readily add, update, and use new software modules (for example, as with Dockstore and Toil). The portion of HuBMAP data that will be open source can be used to choose probes for RNA or proteins), which will provide spatial information for up to hundreds of molecular targets of interest. These data will allow the computational registration of cell-specific epigenomic or transcriptomic profiles to cells on a histological slide to reveal various microenvironmental states. They will potentially include information about protein localization to cytoplasm, nucleus, or cell surface; phosphorylation; complex assembly; extracellular environment; and cellular phenotype determined by protein marker coexpression. Registration and computational integration of complex imaging data will provide biological insights beyond any single imaging modality. The powerful combination of single-cell profiling and multiplexed in situ imaging will provide a pipeline for constructing multi-omics spatial maps for the various human organs and their cellular interactions at a molecular level.

The TMCs will use complementary methods for data collection with an emphasis on processes to ensure the generation of high-quality data and standardized metadata annotations. Benchmarking, quality assurance and control standards, and standard operating procedures, where appropriate, will be developed for each stage of the methodological process and be made available to promote rigor, reproducibility and transparency. It is expected that quality assurance and control standards for both biospecimens and data will evolve as tissue collection, processing techniques, storage and shipping conditions, assays, and data-processing tools change, and as HuBMAP interacts and collaborates with other related efforts, as they have for other consortium projects. Where possible, metadata related to preanalytical variables (for example, annotations and nomenclature) and technologies will be harmonized, and protocols and standards will be shared with the wider research community.

Building an integrated tissue map across scales

The diversity of data generated by HuBMAP, ranging across macroscopic and microscopic scales (for example, anatomical, histological, cellular, molecular and genomic) and multiple individuals, is essential to its core mission. Exploring each of these valuable datasets collectively will yield an integrated view of the human body. Hence, HuBMAP will develop analytical and visualization tools to bridge spatial and molecular relationships in order to help to generate a high-resolution three-dimensional molecular atlas of the human body.

The volume of data generated and collected by HuBMAP will require the utilization, extension and development of tools and pipelines for data processing. While we expect that initial data-processing tools will be based on methods developed by consortium members, HuBMAP will also work with and incorporate algorithms developed by other programs and the wider research community to supplement, enhance or update its pipelines. To this end, HuBMAP will develop one or more portals tailored to emerging use cases identified through a series of user needs. These open source portals will use recognized standards and be interoperable with other platforms, such as the HCA Data Coordination Platform, making it possible to readily add, update, and use new software modules (for example, as with Dockstore and Toil). The portion of HuBMAP data that will be open source can live on or be accessed from multiple platforms, enhancing its utility. This infrastructure will enable external developers to apply their codes, applications, open application programming interfaces, and data schema to facilitate customized processing and analysis of HuBMAP data in concert with other data sources. Furthermore, by actively working with other global and NIH initiatives, the consortium will seek to
reduce the barriers to browsing, searching, aggregating, and analysing data across portals and platforms.

To fully integrate spatial and molecular data across individuals, HuBMAP will create a common coordinate framework (CCF) that defines a three-dimensional spatial representation, leveraging both an early consortium-wide effort to standardize technologies and assays using a single common tissue and the broader range of tissues of the human body analysed across multiple scales (whole body to single cells). This spatial representation will serve as an addressable scaffold for all HuBMAP data, enabling unified interactive exploration and visualization (search, filter, details on demand) and facilitating comparative analysis across individuals, technologies, and laboratories.44,35. To achieve these objectives, HuBMAP envisions a strategy inspired by other tissue atlas efforts36–38 that leverages the identification of ‘landmark’ features, including key anatomical structures and canonical components of tissue organization (for example, epidermal boundaries and normally spatially invariant vasculature) that can be identified in all individuals. These landmarks will enable a ‘semi-supervised’ strategy for aligning and assembling an integrated reference, upon which HuBMAP investigators can impose diverse coordinate systems, including relative representations and zone-based projections. As one example, an open-source, computational histology topography cytometry analysis toolbox (histoCAT49) currently facilitates two-dimensional visualization and will soon also be applicable to three-dimensional reconstruction. Ontology-based frameworks will be explored in parallel to effectively categorize, navigate, and name multiscale data; synergies are expected between these two approaches. Whenever available, medical imaging, such as CT and MRI information, will serve as a basis for landmarking and constructing the CCF techniques to specimens collected in a reproducible manner from specific sites in the body. These data will be then be integrated to arrive at a high-resolution, high-content three-dimensional map for any given tissue. To ensure inter-individual differences will not be confounded with collection heterogeneity, a robust CCF will be developed.

**Technology development and implementation**

Quantitative imaging of different classes of biomolecule in the same tissue sample with high spatial resolution, sensitivity, specificity, and throughput is central to the development of detailed tissue maps. Although no single technique can fully address this challenge at present, the development and subsequent multiplexing of complementary capabilities provides a promising approach for accelerating tissue mapping efforts. The HuBMAP innovation technologies groups aim to develop several innovative approaches that will address the limitations of existing state-of-the-art techniques. For example, transformative technologies such as signal amplification by exchange reaction (SABER)30,41, seqFISH18,42,43, and Lumiphore probes44 will be refined to improve multiplexing, sensitivity, and throughput for imaging RNA and proteins across multiple tissues. Furthermore, new mass spectrometry imaging techniques will enable the quantitative mapping of hundreds of lipids, metabolites, and proteins from the same tissue section with high spatial resolution and sensitivity45,46. There is also scope within the program to develop and test new technologies. These efforts will benefit from the development of new computational tools and machine learning algorithms, optimized first from data generated from a common tissue during the pilot phase, for data integration across modalities.

**Challenges**

Previous programs such as GTEx have faced the challenge of optimizing the collection, preservation, and processing of a wide variety of tissue types from multiple donors. However, one of the goals of HuBMAP to generate comprehensive, interactive high-resolution maps using a wide variety of assays, introduces an added level of complexity. Mapping functionally important biomolecules, including some of
which we may not even be aware and for which sensitive, specific, and high-throughput assays are still lacking, will require close attention. Moreover, the program will produce an unprecedented volume and diversity of datasets for comprehensive data capture, management, mining, modelling, visual exploration and communication. The integration of data from different modalities is required for generating robust maps; it will be necessary to develop corresponding analysis and interactive visualization tools to ensure that the data and atlas are accessible to the entire life-sciences community. Finally, given the enormity of a human atlas, HuBMAP faces the challenges of prioritization of tissues and technologies, sampling across tissues and donors, and optimally synergizing its efforts with international efforts. Determining the number of cells, fields of view, and samples needed to capture rare cell types, states or tissue structures is an important challenge, but can be tackled with adaptive power analyses, leveraging the growing amount of data available both within HuBMAP and from other consortia as well as individual groups.

Resources and community engagement

HuBMAP is an important part of the international mission to build a high-resolution cellular and spatial map of the human body, and we are firmly committed to close collaboration and synergy with the aforementioned initiatives to build an easy-to-use platform and interoperable datasets that will accelerate the realization of a high-resolution human atlas. Shared guiding principles around open data, tools, and access will enable collaborative and integrated analyses of data produced by diverse consortia. To achieve this synergy, HuBMAP and other consortia will work together to tackle common computational challenges, such as cellular annotation, through formal and informal gatherings focused on addressing these problems, planned joint benchmarking and hands-on jamborees and workshops. Another example of the potential for close collaboration is in the study of the colon; multiple projects funded by HuBMAP, the Human Tumour Atlas Network, and the Wellcome Trust will be complemented by projects funded by the Leona M. and Harry B. Helmsley Charitable Trust. With projects focusing on partly distinct regions and diseases (for example, normal tissue, colon cancer, and Crohn’s disease), it will be important for all of the programs to ensure that data are collected and made available in a consistent manner, and HuBMAP will play an active part in such efforts. As a concrete next step, HuBMAP, in collaboration with other NIH programs, plans to hold a joint meeting with the Human Cell Atlas initiative to identify and work on areas of harmonization and collaboration during the spring of 2020. In parallel, HubMAP participants engage in the meetings and activities of other consortia, such as the Human Cell Atlas or the Human Tumour Atlas Network, thus forming tight connections. We have started a series of open meetings to develop the CCF, with the first of these recently held in collaboration with the Kidney Precision Medicine Program and focused on the kidney.

HuBMAP will provide capabilities for data submission, access, and analysis following FAIR (findable, accessible, interoperable, and reusable) data principles47. We will develop policies for prompt and regular data releases in commonly used formats, consistent with similar initiatives. We anticipate that the first round of data will be released in the summer of 2020, with subsequent releases at timely intervals thereafter. Robust metadata will comprise all aspects of labelling and provenance, including de-identified donor information (both demographic and clinical), details of tissue processing and protocols, data levels, and processing pipelines.

Indeed, engagement and outreach to the broader scientific community and other mapping centres is central to ensure that resources generated by HuBMAP will be leveraged broadly for sustained impact. To ensure that browsers and visualization tools from HuBMAP are valuable, the consortium will work closely with anatomists, pathologists, and visualization and user experience experts, including those with expertise in virtual or augmented reality. As described above, we expect that the diversity of normal samples included in this project will facilitate valuable comparative analyses, pinpointing how cells and tissue structures vary across individuals, throughout the lifespan, and in the emergence of dysfunction and disease. The program will build its resources with these use cases in mind and provide future opportunities, such as the demonstration projects, for close collaboration with domain experts. We also anticipate that these data will be highly useful for the generation of new biomedical hypotheses, tissue engineering, the development of robust simulations of spatiotemporal interactions, machine learning of tissue features, and educational purposes.

Conclusions

Analogous to the release of the first human genome build, we anticipate that the first reference three-dimensional tissue maps will represent the tip of the iceberg in terms of their ultimate scope and eventual impact. HuBMAP, working closely with other initiatives, aspires to help build a foundation by generating a high-resolution atlas of key organs in the normal human body and capturing inter-individual differences, as well as acting as a key resource for new contributions in the growing fields of tissue biology and cellular ecosystems. Given the focus of HuBMAP on spatial molecular mapping, the consortium will contribute to the community of efforts seeking similar goals, with a special emphasis on providing leadership in the development of analytical methods for its data types and for developing a common coordinate framework to integrate data. Ultimately, we hope to catalyse novel views on the organization of tissues, regarding not only which types of cells are neighbouring one another, but also the gene and protein expression patterns that define these cells, their phenotypes, and functional interactions. In addition to encouraging the establishment of intra- and extra-consortium collaborations that align with HuBMAP’s overall mission, we envision an easily accessible, publicly available user interface through which data can be used to visualize molecular landscapes at the single-cell level, pathways and networks for molecules of interest, and spatial and temporal changes across a given cell type of interest. Researchers will also be able to browse, search, download, and analyse the data in standard formats with rich metadata that, over time, will enable users to query and analyse datasets across similar programs.

Importantly, we believe that the project’s compilation of different types of multi-omic information at the single-cell level in a spatially resolved manner will represent an important step in the advancement of our understanding of human biology and precision medicine. These data have the potential to redefine types or subtypes of cells and their relationships within and between tissues beyond the traditional understanding that can be obtained through standard methods (for example, microscopy and flow cytometry). We hope this work will be part of a foundation that enables diagnostic interrogation, modelling, navigation, and targeted therapeutic interventions at such an unprecedented resolution to be transformative for the biomedical field.

Received: 22 February 2019; Accepted: 9 September 2019; Published online 9 October 2019.

1. Svensson, V. et al. Power analysis of single-cell RNA-sequencing experiments. Nat. Methods 14, 381–387 (2017).
2. Schwartzman, O. & Tanay, A. Single-cell genomics: techniques and emerging applications. Nat. Rev. Genet. 16, 716–726 (2015).
3. Tanay, A. & Regev, A. Scalable single-cell genomics from phenomenology to mechanism. Nature 541, 331–338 (2017).
4. Norris, J. L. & Caprioli, R. M. Analysis of tissue sections by matrix-assisted laser desorption/ionization imaging mass spectrometry in biological and clinical research. Chem. Rev. 113, 2399–2342 (2013).
5. GTEx Consortium. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).
6. National Research Council of the National Academies, Enhancing the Effectiveness of Team Science (National Academies Press, 2015).
7. Regev, A. et al. The Human Cell Atlas. eLife 6, e27041 (2017).
8. Interactive human protein atlas launches. Cancer Discov. 5, 339 (2015).
9. Eck, J. R. et al. The BRAIN Initiative Cell Census Consortium: lessons learned toward generating a comprehensive brain cell atlas. Neuron 96, 542–557 (2017).
10. NCI Board of Scientific Advisors and the National Cancer Advisory Board. Human Tumor Atlas (HTA) Network, https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/funding/upcoming/hta-foa-video (National Cancer Institute, 2017).
11. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357, 661–667 (2017).
12. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523, 486–490 (2015).
13. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348, 910–914 (2015).
14. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. Nature 540, 1393–1394 (2018).
15. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. Neuron 92, 342–357 (2016).
16. Gottes, Y. et al. Deep profiling of mouse splenic architecture with CODEX multiplexing. Cell 174, 968–981 (2018).
17. Lubeck, E., Coskun, A. F., Zhiventsev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. Nat. Methods 11, 360–361 (2014).
18. Eng, C. L. et al. Cell-type-specific super-resolution imaging in tissues by RNA seqFISH. Nature 568, 235–239 (2019).
19. Van de Plas, R., Yang, J., Spraggins, J. & Caprio, R. P. Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping. Nat. Methods 12, 366–372 (2015).
20. Spraggins, J. M. et al. Next-generation technologies for spatial proteomics: Integrating ultra-high speed MALDI-TOF and high mass resolution MALDI FTICR imaging mass spectrometry for protein analysis. Proteomics 16, 1678–1689 (2016).
21. Rapsomaniki, M. A. et al. CellCycleTRACER accounts for cell cycle and volume in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Phil. Trans. R. Soc. B 2288–2295 (2011).
22. Schulz, D. et al. Simultaneous multiplexed imaging of mRNA and proteins with subcellular resolution in breast cancer tissue by mass cytometry. Cell Syst. 6, 612–620 (2018).
23. Schulz, D. et al. Simultaneous multiplexed imaging of mRNA and proteins with subcellular resolution in breast cancer tissue by mass cytometry. Cell Syst. 6, 612–620 (2018).
24. Chevrier, S. et al. An immune atlas of clear cell renal cell carcinoma. Cell 169, 736–749 (2017).
25. Patterson, N. H. et al. Next generation histology-directed imaging mass spectrometry-genotyping by autofluorescence microscopy. Anal. Chem. 90, 12404–12413 (2018).
26. Teng, M. et al. A benchmark for RNA-seq quantification pipelines. Genome Biol. 17, 74 (2016).
27. Petryszak, R. et al. Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. Nucleic Acids Res. 44, D746–D752 (2016).
28. Hong, E. L. et al. Principles of metadata organization at the ENCODE data coordination center. Database (Oxford) 2016, baw001 (2016).
29. Campbell-Thompson, M. et al. Network for Pancreatic Organ Donors with Diabetes (nPOD): developing a tissue biobank for type 1 diabetes. Diabetes Metab. Res. Rev. 28, 608–612 (2012).
30. Pugliese, A. et al. The Juvenile Diabetes Research Foundation Network for Pancreatic Organ Donors with Diabetes (nPOD) Program: goals, operational model and emerging findings. Pediatr. Diabetes 15, 1–9 (2014).
31. Phillips, T. et al. Factors that influence the quality of RNA from the pancreas of organ donors. Pancreas 46, 252–259 (2017).
32. O’Connor, B. D. et al. The Dockstore: enabling modular, community-focused sharing of Docker-based genomics tools and workflows. F1000Res. 6, 52 (2017).
33. Vianun, J. et al. Toil enables reproducible, open source, big biomedical data analyses. Nat. Biotechnol. 35, 314–316 (2017).
34. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 33, 203–209 (2015).
35. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. 36, 421–427 (2018).
36. Magliozzi, J. et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Phil. Trans. R. Soc. Lond. B 356, 1326–1322 (2001).
37. Fonseca, C. G. et al. The Cardiac Atlas Project—an imaging database for computational modeling and statistical atlas generation of the heart. Bioinformatics 27, 2288–2295 (2011).
38. Fürth, D. et al. An interactive framework for whole-brain maps at cellular resolution. Nat. Neurosci. 21, 139–149 (2018).
39. Sc {:} ioppo, D. et al. The Jumprocytometry: analysis of cell phenotypes and interactions in multiplex image cytometry data. Nat. Methods 14, 873–876 (2017).
40. Saka, S. K. et al. Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. Nat. Biotechnol. 37, 1080–1090 (2019).
41. Ko, S. J. et al. SABER amplifies FISH: enhanced multiplexing of imaging of RNA and DNA in cells and tissues. Nat. Methods 16, 533–544 (2019).
42. Shah, S., Lubeck, E., Zhou, W. & Cai, L. seqFISH accurately detects transcripts in single cells and reveals robust spatial organization in the hippocampus. Neuron 94, 752–759 (2017).
43. Zhu, Q., Shah, S., Dries, R., Cai, L. & Yuan, G.-C. Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. Nat. Biotechnol. 36, 1183–1190 (2018).
44. Cho, U. et al. Ultrasensitive optical imaging with lanthanide lumiphores. Nat. Chem. Biol. 14, 15–21 (2018).
45. Yin, R. et al. High spatial resolution imaging of mouse pancreatic islets using nanodroplet processing platform with deep and quantitative proteome profiling of 10–100 mammalian cells. Nat. Commun. 9, 882 (2018).
46. Zhu, Y. et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. Nat. Commun. 9, 882 (2018).
47. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018 (2016).
