Trends in Genetics

CellPress OPEN ACCESS

Series: Celebrating the Human Genome Project and its outcomes

Opinion

Towards a Human Cell Atlas: Taking Notes from the Past

Rik G.H. Lindeboom [1,*] Aviv Regev,[2,4] and Sarah A. Teichmann[1,3]

Comprehensively characterizing the cellular composition and organization of tissues has been a long-term scientific challenge that has limited our ability to study fundamental and clinical aspects of human physiology. The Human Cell Atlas (HCA) is a global collaborative effort to create a reference map of all human cells as a basis for both understanding human health and diagnosing, monitoring, and treating disease. Many aspects of the HCA are analogous to the Human Genome Project (HGP), whose completion presents a major milestone in modern biology. To commemorate the HGP's 20-year anniversary of completion, we discuss the launch of the HCA in light of the HGP, and highlight recent progress by the HCA consortium.

Highlights

The Human Cell Atlas (HCA) consortium was founded as a collaborative and open effort to create a reference map of the cells in the human body.

Organizing a large-scale project such as the HCA draws inspiration from the Human Genome Project (HGP) that was completed 20 years ago.

Significant progress has been made by the HCA community, including profiling more than 39 million cells from 15 major organs to date.

The expected impact of the HCA is illustrated by its use during the coronavirus disease 2019 (COVID-19) pandemic.

Building a Reference Map of the Human Body

In the past decade, new methods have emerged for single-cell genomics that have revolutionized our ability to identify and characterize the cells that comprise complex tissues. With these tools at hand, the HCA project was launched as an international collaborative effort to create comprehensive reference maps of all human cells – the fundamental units of life – as a basis for both understanding human health and diagnosing, monitoring, and treating disease [1]. The foundation for organizing large-scale consortium efforts such as the HCA leads back to the HGP. To commemorate the completion of the HGP 20 years ago, we lay out organizational considerations and the latest progress of the HCA community.

Lessons from the HGP

The HGP was launched in 1990 as a scientific effort of unprecedented magnitude to create a reference map of the human genome. The success of this ambitious project depended on an interdisciplinary approach that bridged teams specialized in computation, engineering, and biology, in an international collaboration between institutions in the USA, Europe, and Asia. The focus on international and interdisciplinary collaboration inspired numerous large-scale consortium-based research ventures that followed, including the HCA initiative. Recognizing the broad importance of ethics by dedicating 5% of its funding to ensure proper ethical practice and explore its societal impact, the HGP contributed an important aspect that inspired many large-scale biological projects. While it was not fully clear at the time of the launch how the HGP would succeed in its ambitious goal, the focus on intermediate milestones and technology development eventually led to a finished human genome reference 2 years ahead of schedule and with budget to spare. This achievement underlines the importance of phasing long-term initiatives into graspable intermediate goals to refine future plans and exploit the inevitable increase in throughput and resolution that technological advances bring.

Organizing the HCA

Similar to the aim of the HGP to build a reference map of the genome, the goal of the HCA initiative is to create reference maps that chart the cells in human tissues and organs. Building such maps

[1]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK
[2]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[3]Cavendish Laboratory, JJ Thomson Avenue, University of Cambridge, Cambridge, UK
[4]Current Address: Genentech, South San Francisco, CA, USA

*Correspondence: rl21@sanger.ac.uk (R.G.H. Lindeboom).

requires collaboration between research groups and institutes, so the HCA was launched as an open international initiative. To maximize the benefits from collaboration and data sharing, the HCA is organized as an open and community-driven venture with more than 2000 members to date, and growing. Any scientist that shares its ambitions, goals, and values can become a member at any point by registering online[i].

To ensure scientific leadership and deep engagement from the broad scientific community, HCA's working groups take on its core challenges[ii], and include the Biological Networks, each taking on a specific tissue, organ, or system, the Analysis Working Group, focused on computational and analytical challenges, the Standards and Technologies Working Group, focused on the needed experimental assays. Notably, when creating a human reference resource such as the HCA, it is essential to ensure an equal benefit is gained worldwide, to both the participating scientists and the representation of humanity, requiring the incorporation of extensive diversity in sex and ethnicity. HCA engaged in this goal early and in an ongoing manner through the Ethics Working Group, and more recently the Equity Working Group [2].

The relatively late realization of the importance of data analysis in the HGP presented a bottleneck for the HGP to construct a genome reference [3]. Learning from this past experience, the continuous development of computational approaches has thus been a major area of focus of the HCA community[iii], where the Analysis Working Group – the first working group of HCA – is dedicated to the key computational challenge of building and querying and atlas. While the core product of the HGP was essentially a single DNA sequence, the multimodal and complex nature of the data generated within the HCA will require a modular and multifaceted approach to standardize, integrate, and share data. To this end, the HCA Data Coordination Platform was established in 2017, and is under continuous development to accommodate standardized processing and broad access to HCA data through both graphic and programming interfaces[iv]. In addition, a burgeoning community of tertiary data portals now enable users to easily access and analyze HCA data without the need for sophisticated bioinformatics expertise. Examples of these portals include the cellxgene software [4], EBI's Single Cell Expression Atlas[v], the Cambridge Portal[vi], the Broad Single Cell Portal[vii], and the UCSC Cell Portal[viii], each offering interactive access to datasets from a wide range of HCA studies, with distinct analysis features. Other dedicated portals, such as the COVID-19 Cell Atlas[ix], Developmental Cell Atlas[x], and the Human Tumor Atlas[xi] data portals, are HCA-related portals dedicated to specific aspects of human biology and/or disease.

Unlike the single coordinated funding structure for the HGP, the HCA involves a more distributed structure, reflecting the democratization of technology, computation, and growth of the biomedical scientific community itself, especially in genomics and computational biology, over the past decades. At the time of the HGP, only a few, large, and heavily funded centers could perform the needed work, but both single cell genomics technologies and associated computation has become much more broadly accessible, partly due to the innovation and efforts of HCA members. As a result, the funding and organizational structure is distinct: scientists participate in HCA irrespective of their specific funding source, many funders support atlas construction activities, and HCA is allied with several formally funded consortia focused on specific aspects, providing a scientific community open to all.

Progress towards a Draft of the HCA

The HCA has already made significant progress towards the goal set for the first draft of a cell atlas – profiling the common cell types in tissues from the major human organs[xii]. Currently, HCA scientists have profiled more than 39 million cells using suspension cell genomics from 15

major organ systems, including for example 11.1 million nervous system cells, 5.8 million embryonic and fetal cells, 3.4 million lung cells and 7.2 million immune cells. These atlases also cover important human diseases, including nearly 4.8 million cells derived from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-infected individuals. These statistics are collected by the HCA Executive Office through regular quarterly surveys of its members, and thus reflect data currently available in a multiple range of sources, including data consented for open access stored at the HCA Data Coordination Platform[iv], and HCA data generated under consent for data sharing by managed or controlled access only, which are currently distributed across various databases such as dbGAP, DUOS, and EGA. These cell counts also include unpublished datasets for which the cell numbers have been shared with us by HCA members.

The data collected across HCA is leading to exciting scientific discoveries. For example, a recently published cardiac single cell reference highlights the cellular heterogeneity of the atrial and ventricular chambers, and gender-specific differences in the cellular composition of the heart [5]. The emerging lung atlas discovered a host of new cell types, from the ionocyte, a new cell type expressing the cystic fibrosis gene *CFTR* [6], to endothelial cell subsets that may play a role in COVID-19 [7]. The gut atlas is recovering many dozens of cell types in the small and large intestine [8,9], including rare cells such as enteric neurons [10]. Similarly, multiple atlases of the human liver, an organ historically known for its homogenous cell type composition, reveal heterogeneity in epithelial progenitors [11] and provide broad insights in hematopoietic development that occurs in the fetal liver [12]. Single-nucleus RNA-sequencing (seq) of the neurons of the cerebral cortex uncovered extensive differences in the cellular composition and characteristics between human and mouse models, highlighting the importance of generating a cell atlas for humans [13]. A spatial cell atlas of healthy and diseased pancreas tissues reveals how the morphological organization of this organ features cell-type-specific neighborhoods and unexpected cell–cell interactions [14]. In the thymus, the dynamic cellular composition across the human lifespan unraveled the development and repertoire of T cells and thymic stroma at unprecedented detail [15]. Profiling the cellular composition of the maternal–fetal interface of the placenta unveiled many regulatory interactions that govern the cellular organization during early human pregnancy [16].

While the first steps towards a HCA are to create a reference of healthy cells, many efforts have also already examined implications in disease. A human single-cell atlas of the lung has identified novel epithelial cell types, including asthma-related cell populations [17], and similar atlases in the gut have helped understand cells related to inflammatory bowel disease [8,9]. Charting the dynamic cellular composition of the fetal, pediatric, and adult human kidney has uncovered that pediatric and adult kidney cancers originate from different and previously little-known cell types [18]. Systematic interrogation of tumor cell landscapes with complementary single-cell RNA-seq techniques has furthermore enabled scientists to study single-cell biology at a pancancer scale [19]. Several cell atlases have detailed organ-specific subsets of tissue-resident immune cells [5,12,15,18], underscoring the impact of spatial and environmental influences of cells for their identity and function. In disease, the HCA approach has inspired dedicated initiatives such as the Human Tumor Atlas Network [20], whereas efforts such as the Kidney Precision Medicine Program (KPMP) are tackling multiple kidney diseases. Similarly, the COVID-19 pandemic sparked a large-scale joint effort of HCA scientists to shed light on this new pathology at single-cell resolution (see later).

Impact of the HCA

As highlighted earlier, contributions by the HCA community have already led to numerous insights ranging from basic human physiology and fundamental biology to discoveries with direct clinical

applications such as pinpointing disease-associated cell types and pathology-induced cell states (Figure 1). Beyond these direct insights that individual studies bring, the long-term goal of the HCA is to provide a comprehensive reference of the identities and characteristics of the cells in a human body (see Outstanding Questions). The HGP provided a reference where biologists could look up the origin of, for example, their isolated fragment of DNA, RNA, or protein, and how this differed in a disease context. The HCA aspires to become a similar tool to accelerate both fundamental and translational science. Equivalent look-ups for HCA will include determining which cells express a gene of interest, what cell types are present in a tissue/organ, and which cell types co-occur in close spatial proximity. In addition, key marker genes that identify a cell type of interest can be derived from the HCA, which can be the starting point for numerous experimental assays.

The impact of the HCA on understanding human disease was powerfully illustrated during the dawn of the COVID-19 pandemic. To obtain early insights into the pathology of COVID-19, the HCA community used the existing single-cell RNA-seq data in the atlas to study important aspects of viral infection and responses at single cell resolution [21]. This quickly led to a comprehensive overview of the cells and organs that express key viral entry genes, such as *ACE2*, and are therefore susceptible to SARS-CoV-2 infection. HCA scientists have since turned to studying samples from COVID-19 patients and deceased donors directly, to shed light on disease pathology at the single cell and spatial level. This work presents a clear case-in-point of how the HCA will be useful for biologists and society.

The HCA also holds a translational promise, where the cell atlas can be applied to a range of medical questions related to, for example, disease mechanisms, diagnostics, regenerative medicine, and drug discovery and toxicity (Figure 1). Here, the HCA can be used to identify disease-associated cell phenotypes that represent drug targets, predict cell-type specific effects and understand on-target effects in other cell types. In regenerative medicine, the cell atlas can provide guidance on how to steer differentiation into desired cell fates. In the clinic, new biomarkers for disease could be identified and interpreted at unprecedented resolution.

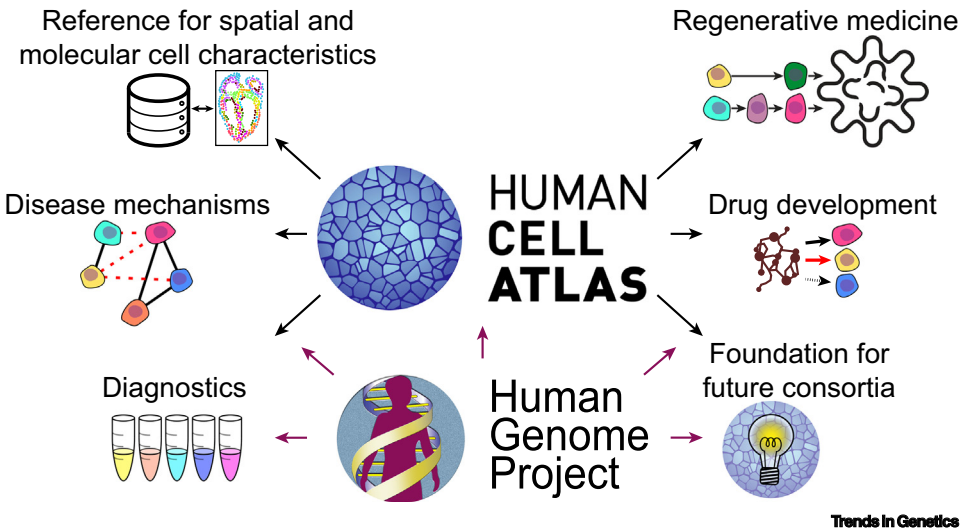

Figure 1. Schematic Overview of the Impact of the Human Cell Atlas (HCA). Different biomedical fields that are expected to benefit from the HCA are highlighted around the HCA logo. These fields (indirectly) also greatly benefited from the Human Genome Project, which has provided insights and inspiration for organizing and predicting the impact of consortia such as the HCA.

While the earlier-mentioned examples provide insights in the direct and future utility of the HCA, we can take notes from other consortia such as the HGP to predict the longer-term future impact of generating the HCA (Figure 1). The impact of the HGP was much greater than the direct discoveries about the human genome. With the genome reference as a basis, many new large-scale consortia such as HapMAP, 1000 Genomes, GWAS, ENCODE, and TCGA/ICGC projects were launched, resulting in unprecedented scientific achievements. We expect that the HCA can have a similar impact, where the fundamental knowledge about the cellular organization of our body will act as a foundation for future projects and consortia to investigate human physiology and disease at even higher resolution and in the spatial context of the human body.

Concluding Remarks and Future Perspectives

As we progress towards a first draft of the HCA, exciting technological advances are enabling the community to characterize cells at higher throughput, and in a more detailed and comprehensive manner. While the cellular maps highlighted earlier are mostly based on single-cell RNA-seq, these are now complemented with chromatin profiles, protein profiles, and spatial information. Simultaneous measurements of multiple modalities in the same single cell, such as the proteome, transcriptome, and epigenome, will greatly advance our understanding of cell identities and phenotypic characteristics.

As laid out in the HCA White Paper [22], the first draft of the HCA aims to profile 30–100 million human cells from all major organs in ethnically diverse males and females. In addition to cell suspension-based transcriptome profiling – which is currently the most mature technology to profile cells for the HCA – these tissues will also be profiled with spatial profiling technologies to map identified cell types onto the tissue architecture of the human body. Rapid developments in the throughput and molecular resolution of spatial transcriptomics and other spatial profiling methods have enabled the establishment of the spatial branch of the HCA.

Lastly, mirroring the advances in sequencing technologies at the time of the HGP, there is a sustained dramatic increase in throughput of single-cell profiling techniques; a trend that will allow HCA to reach its ambitious goal to profile and ultimately characterize billions of cells from human organs in health, as a reference map of the human body.

Outstanding Questions

What is the comprehensive compendium of cell types and cell states in the human body?

Are there differences in tissue architecture between males and females (besides reproductive tissues)?

Are there stereotypical patterns of ageing in human tissues?

Can HCA data reveal new units of tissue architecture (i.e., recurring 3D motifs of cell types) that are hitherto unappreciated?

Can HCA data reveal new, previously unknown adult stem cells and pinpoint regeneration occurring as part of homeostasis within adult tissues?

How similar are immune, fibroblast, and vascular lineage cells across tissues and organs?

What is the impact of gradients of signaling molecules (either proteins or small molecules, including, e.g., oxygen gradients) in development and in adult tissues?

Are cellular responses to gradients and other challenges (e.g., oxidative stress) restricted to a single cellular compartment or concerted across, for example, immune, epithelial, and vascular compartments?

Acknowledgments

We gratefully acknowledge Jennifer E. Rood for critical reading and editing of this manuscript, and all authors of HCA papers that we have not been able to cite due to space constraints. This publication is part of the Human Cell Atlas – www.humancellatlas.org/publications.

Declaration of Interests

In the last 3 years, S.A.T. has consulted for Genentech and Roche, and is a member of SABs of Biogen, GlaxoSmithKline, and Foresite Labs. A.R. is a cofounder and equity holder of Celsius Therapeutics, an equity holder of Immunitas and was an SAB member of Neogene Therapeutics, Thermo Fisher Scientific, Asimov, and Syros Pharmaceuticals until July 31, 2020. Since August 1, 2020, A.R. is an employee of Genentech, a member of the Roche group. A.R. is an inventor on multiple patents to the Broad Institute in the area of single cell genomics. R.G.H.L. has no interests to declare.

Resources

[i] www.humancellatlas.org/join-hca/
[ii] www.humancellatlas.org/learn-more/working-groups/
[iii] https://openproblems.bio/
[iv] https://data.humancellatlas.org
[v] www.ebi.ac.uk/gxa/sc/home
[vi] www.cambridgecellatlas.org

[vii]https://singlecell.broadinstitute.org/single_cell
[viii]https://cells.ucsc.edu
[ix]www.covid19cellatlas.org
[x]https://developmentcellatlas.ncl.ac.uk
[xi]https://data.humantumoratlas.org
[xii]www.humancellatlas.org/publications/

References

1. Regev, A. *et al.* (2017) The Human Cell Atlas. *eLife* 6, e27041
2. Majumder, P.P. *et al.* (2020) The Human Cell Atlas and equity: lessons learned. *Nat. Med.* 26, 1509–1511
3. Green, E.D. *et al.* (2015) Human Genome Project: twenty-five years of big biology. *Nature* 526, 29–31
4. Li, K. *et al.* (2020) cellxgene VIP unleashes full power of interactive visualization, plotting and analysis of scRNA-seq data in the scale of millions of cells. *bioRxiv* Published online August 31, 2020. https://doi.org/10.1101/2020.08.28.270652
5. Litviňuková, M. *et al.* (2020) Cells of the adult human heart. *Nature* 588, 466–472
6. Montoro, D.T. *et al.* (2018) A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 560, 319–324
7. Travaglini, K.J. *et al.* (2020) A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 587, 619–625
8. Smillie, C.S. *et al.* (2019) Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell* 178, 714–730.e22
9. Martin, J.C. *et al.* (2019) Single-cell analysis of Crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy. *Cell* 178, 1493–1508
10. Drokhlyansky, E. *et al.* (2020) The human and mouse enteric nervous system at single-cell resolution. *Cell* 182, 1606–1622.e23
11. Aizarani, N. *et al.* (2019) A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* 572, 199–204
12. Popescu, D.M. *et al.* (2019) Decoding human fetal liver haematopoiesis. *Nature* 574, 365–371
13. Hodge, R.D. *et al.* (2019) Conserved cell types with divergent features in human versus mouse cortex. *Nature* 573, 61–68
14. Tosti, L. *et al.* (2020) Single nucleus and in situ RNA sequencing reveals cell topographies in the human pancreas. *Gastroenterology* 160, 1330–1344.e11
15. Park, J.E. *et al.* (2020) A cell atlas of human thymic development defines T cell repertoire formation. *Science* 367, eaay3224
16. Vento-Tormo, R. *et al.* (2018) Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* 563, 347–353
17. Vieira Braga, F.A. *et al.* (2019) A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* 25, 1153–1163
18. Young, M.D. *et al.* (2018) Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* 361, 594–599
19. Slyper, M. *et al.* (2020) A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* 26, 792–802
20. Rozenblatt-Rosen, O. *et al.* (2020) The Human Tumor Atlas Network: charting tumor transitions across space and time at single-cell resolution. *Cell* 181, 236–249
21. Teichmann, S. and Regev, A. (2020) The network effect: studying COVID-19 pathology with the Human Cell Atlas. *Nat. Rev. Mol. Cell Biol.* 21, 415–416
22. Regev, A. *et al.* (2018) The Human Cell Atlas White Paper. *arXiv* Published online October 11, 2018. http://arxiv.org/abs/1810.05192