Toward mapping the human body at a cellular resolution

Ananda L. Roy* and Richard S. Conroy*

Office of Strategic Coordination, Division of Program Coordination, Planning, and Strategic Initiatives, Office of the Director, National Institutes of Health, Bethesda, MD 20892

ABSTRACT The adult human body is composed of nearly 37 trillion cells, each with potentially unique molecular characteristics. This *Perspective* describes some of the challenges and opportunities faced in mapping the molecular characteristics of these cells in specific regions of the body and highlights areas for international collaboration toward the broader goal of comprehensively mapping the human body with cellular resolution.

INTRODUCTION

Mapping the human body

The adult human body is composed of ~37 trillion cells of human origin and at least as many again as part of the human microbiota (Bianconi *et al.*, 2013; Sender *et al.*, 2016). Historically, cells have been defined based on their origin and morphology and more recently based on cell surface proteins, resulting in 200–500 defined major cell types (Valentine *et al.*, 1994; Vickaryous and Hall, 2006). Increasingly, however, the limitations of this approach are becoming apparent. Intrinsic and extrinsic factors such as epigenomic modifications, chromatin structure, cell phase, chronobiology, cell signaling, extracellular environment, and cellular neighborhood introduce spatial-temporal modulation of the cell state that can significantly alter its physical, molecular, and functional properties. Thus, single-time-point studies of isolated cells outside their natural environmental context do not provide accurate structural, molecular, or functional information.

One of the major goals of mapping human tissues at the cellular level is to better understand this spatial-temporal context. Centuries of anatomical and pathological work has highlighted many of the principles underlying the spatial arrangement of cells in human tissues and identified repeating neighborhood structures with complex function. However, the extent to which cells within these structures represent unique cell types or provide distinct functions has not been known. With the emergence of high-throughput techniques for the molecular profiling of single cells along with improved high-content, high-resolution imaging methods, there is now the opportunity to expand the existing structural, functional, and physiological maps of the human body with quantitative molecular information. Specific details about the distribution of nucleic acids, proteins, metabolites, lipids and other biomolecules in the intra- and extracellular environment is likely to transform our understanding of cell type and cell state and how tissue organization varies across individuals, the lifespan, and the health–disease continuum.

Why do we need to understand the human body with single-cell resolution?

Genetic variation, epigenetic modifications, and chromatin structure drive phenotypic variation at the level of individual cells. Understanding human biology at the level of individual cells is therefore necessary for understanding the impact of genetic variants and epigenetic modifiers on human health and disease. Causally relevant cells for any particular health or disease condition cannot be known unless the many cell states are defined. Similarly, specific therapeutic targeting of disease-causing cells cannot be accomplished if molecular profiles that distinguish these cells from healthy cells are unavailable.

Cellular heterogeneity from perturbations in biomolecular profiles, the tissue environment or changes in cellular signals often dictate the emergence of diseased or dysfunctional states. This is recently exemplified by the heterogeneity observed in neurodegeneration models and in neuromuscular disease (Rodriguez-Muela *et al.*, 2017; Ajami *et al.*, 2018). Two cells nominally of the same type in the same organ or tissue can behave differently to therapeutic intervention depending on their molecular and functional states. Their state is dependent on many factors, including the spatiotemporal environment of the cell, and is influenced by local and systemic signaling, extracellular structure, and previous internal states. Furthermore, the role of rare and motile cells, such as immune cells or cells in a pluripotent state, can significantly alter the phenotypes of cellular neighborhoods. Given these considerations, deep molecular information at cellular resolution can significantly enhance and complement positional information of tissue and organ-resident cells to define health and disease.

THE OPPORTUNITIES FOR SINGLE-CELL ANALYSIS

The scale and complexity of assembling a biomolecular atlas of the human body is daunting. Currently, it takes months to measure and analyze the transcriptional profiles of millions of cells using state-of-the-art techniques, so to date work has focused on specific areas or organs and several sampling approaches employed. Another practical consideration is the trade-off between the molecular depth, the spatial resolution, and the volume of tissue to be analyzed. For example, existing sequencing-based assays provide detailed transcriptional profiles but limited spatial information. Conversely, most spatially resolved assays are limited in the scope of biomolecular data they can provide, particularly those using probes. Ideally, in the short-term, iteratively applying these two broad classes of assays should provide positional information of cells in resident tissues as well as their molecular signatures. However, to build a comprehensive view of all the types of biomolecules present in the body overall and to link the organization of cells in tissue to its overall function will require a new generation of assays and computational methods. The dividends for this effort though are potentially high. Linking together anatomical, molecular, and functional information at different scales will be an opportunity to decipher the influence of cell–cell, cell–neighborhood, and cell–organism communication, the plasticity of cell types, the robustness and sensitivity of functional circuits to molecular perturbations (and vice versa), and how organization, cyclic states, and aging influence the emergence of dysfunction and disease.

Here we describe existing and emerging technologies that are amenable to single-cell analysis and the opportunities they present for constructing a biomolecular atlas. These technologies are not widely validated yet and indeed may not be usable or provide insightful information for all cell and tissue types. Each technique may also have limitations that we have tried to highlight to indicate where future work may be needed.

Genomic assays
Demonstrating reliable single-nucleotide variation detection across the genome of a single cell is difficult due to technical limitations, including sensitivity (low copy number variation) and the need for amplification of the DNA to levels sufficient for sequencing (low amplification fidelity) (Eberwine, 2017). To address these challenges, Xie and colleagues developed multiple annealing and looping-based amplification cycles and, more recently, linear amplification via transposon insertion (LIANTI) to demonstrate whole-genome amplification of single-cell genomic DNA using linear amplification

(Huang *et al.*, 2015; Chen *et al.*, 2017). The authors reported that LIANTI outperforms existing methods, thereby enabling micro–copy number variation detection with kilobase resolution. This allowed direct observation of stochastic firing of DNA replication origins as well as showed that cytosine-to-thymine mutations observed in single-cell genomics often arise from the artifact of cytosine deamination on cell lysis (Chen *et al.*, 2017). Using carefully designed FISH probes, Levesque *et al.* (2013) demonstrated that it was possible to quantify allele-specific expression in single cells in culture. These methods hold significant promise for understanding structural, copy number, and single-nucleotide variations and lineage, though cost, multiplexing these techniques, detecting multiple SNPs concurrently, and working with human tissue samples preserved in different ways and with many different cell types are significant challenges to generating high-throughput data.

Genomic assays are also increasingly providing functional insights. For example, single-cell DNA methylomes have been used to identify 16 mouse and 21 human neuronal subpopulation of cells in the frontal cortex from isolated single nuclei (Luo *et al.*, 2017). Likewise, using a single-nucleus drop-seq method combining with single-cell transposome hypersensitive site sequencing, recent studies mapped >60,000 single cells for regulatory elements and transcription factor binding sites from human adult visual cortex, frontal cortex, and cerebellum (Lake *et al.*, 2016). Nucleus isolation from tissues has been found to be more useful than cell dissocation, particularly for tissues with complex and variable extracellular structures; however, analysis of nuclei alone may provide a skewed view of the transcriptional state of a cell.

An assay for transposase-accessible chromatin using sequencing (ATAC-seq) has enabled analysis of flash-frozen primary tissue samples, with a recent article describing how more than 15,000 nuclei were analyzed and used to identify 20 distinct cell populations and further delineated transcriptional regulatory sequences as well as developmental programs through eight developmental stages (Preissl *et al.*, 2018). Combining single-cell resolution with chromatin capture assays, understanding three-dimensional folding of genome in great depth has been acheived. Ramani *et al.* (2017) used single-cell combinatorial indexed Hi-C to separate cells by karyotyping as well as cell-cycle state differences and identifying cell-to-cell heterogeneity in mammalian cells. Integrating single-cell combinatorial indexing RNA sequencing (sci-RNA-seq) with whole-animal chromatin immunoprecipitation sequencing revealed cell-type effects of transcription factors (Cao *et al.*, 2017). Although these studies were performed in nematodes, applying similar approaches to human tissues would open new avenues to determine cell-type-specific transcription programs and pathways of differentiation at unprecedented resolution. There is also the potential to improve the efficiency and depth of these techniques and to compare single-cell and bulk analysis.

These studies provide a glimpse of deciphering the genome at the single-cell level and insights into genetic programs and regulatory pathways. However, these techniques rely on isolating individual cells and do not capture spatial information. They are also susceptible to the biases, sensitivity, and specificity limitations of their underlying sequencing approach that are particularly important for single cells where individual transcripts or interactions may be present and critical to defining cell state.

Transcriptomic assays
There has been a dramatic rise in the size of studies analyzing single-cell RNA sequencing data due to automation and improvement of sample preparation, sequencers, and analytical tools (Rozenblatt-Rosen *et al.*, 2017). A wide variety of sequencing approaches have

also emerged over the past few years, each with its own strengths and weaknesses (Haque *et al.*, 2017; Ziegenhain *et al.*, 2017). Further, the dramatic rise in variation and throughput of sequencing techniques has been mirrored by increased multiplexing and throughput of image-based techniques. Single-molecule FISH (smFISH) has been the gold standard for quantifying individual transcript abundances (Coleman *et al.*, 2015). FISH-based systems provide direct measurement of spatial organization and can provide semiquantitative readout; however, they have limitations in dynamic range and in density of transcripts that can be resolved simultaneously. Recently, smFISH has been multiplexed to the transcriptome level and profile 10,212 different mRNAs from mouse fibroblast and embryonic stem cells (Eng *et al.*, 2017). This method, called RNA sequential probing of targets, provides an accurate, flexible, and low-cost alternative to sequencing for profiling transcriptomes. Other recent innovations include multiplexed error-robust fluorescence in situ hybridization (Moffitt and Zhuang, 2016), single-molecule hybridization chain reaction for amplification of signal (Shah *et al.*, 2016), and fluorescence in situ sequencing (Lee *et al.*, 2015). Together these innovative approaches are extending beyond more traditional FISH assays, though the extent of any bias, the density of transcripts that can be probed simultaneously, and sensitivity of these techniques when applied to diverse human tissues are yet to be well established.

While RNA abundance could indicate individual (static) cellular states, it does not directly reveal dynamic processes such as cellular differentiation (La Manno *et al.*, 2017). However, RNA velocity, the time derivative of RNA abundance, can be estimated by distinguishing unspliced and spliced mRNAs in standard single-cell RNA sequencing protocols that can predict the future state of individual cells on a timescale of hours. The accuracy of RNA velocity in the neural crest lineage has been recently validated, further demonstrating its use on multiple technical platforms. This appears to be an exciting development that could greatly aid the analysis of developmental lineages and cellular dynamics, particularly in preserved human tissues (La Manno *et al.*, 2017).

Proteomic assays

RNA analysis at a single-cell resolution currently far exceeds that of protein analysis; however, there have been several recent advances for analyzing proteins at a single-cell resolution. Two commonly used methods for massively parallel protein measurements in ensembles of cells are mass spectrometry and barcoded antibodies coupled with flow cytometry. These have now increasingly been applied to the single-cell domain. For example, mass cytometry (fluorescence-based flow cytometry with labeled antibodies combined with mass-spectrometry) or cytometry by time-of-flight (CyTOF) provide high-dimensional quantitative analysis that can delineate cell types based on both surface markers and intracellular signaling molecules (Bjornson *et al.*, 2013). CyTOF uses stable isotopes instead of fluorophore-conjugated antibodies as reporters, thus limiting noise due to autofluorescence. Furthermore, CyTOF is amenable to Formalin-fixed, paraffin-embedded tissues, which is often the major source of pathology biospecimens. Mass cytometry also allows us to interrogate protein posttranslation modifications (e.g., phosphoflow cytometry), thereby measuring basal versus induced cellular states (as in signal transduction) at a single-cell resolution (Bandyopadhyay *et al.*, 2017). Another method, multiparameter ion beam imaging, which integrates imaging mass spectrometry and labeled antibodies, can analyze up to 100 targets simultaneously over a five-log dynamic range and thereby has the potential to establish

three-dimensional tissue maps (Angelo *et al.*, 2014). Serial fluorescence imaging allows multiplex analysis of proteins, and several variants have been developed over the past decade for multiplexed imaging (Schubert *et al.*, 2006), either by binding and stripping the probes sequentially (Gerdes *et al.*, 2013) or by barcoding the probes and sequencing the fluorophores. These techniques have been used to generate single-cell resolution maps in hundreds of human tissue samples using 61 markers (Gerdes *et al.*, 2013) and to study the architecture of mouse spleen with 66 markers that is compatible with any three-color fluorescence microscope (Goltsev *et al.*, 2018). These techniques provide spatial information directly, though a significant challenge is the availability of consistent, high-quality human antibodies and the potential for bias based on the density and modification of proteins of interest. Another potentially relevant development is the "top-down" mass-spectrometry approach that promises to provide readout for intermolecular codes imprinted on histone tails (thus measuring epigenetic changes), although the application to single-cell resolution remains to be firmly proven (Zheng *et al.*, 2016). Together, these techniques are providing an increasingly strong portfolio of tools for studying protein expression in tissues at the single-cell level.

Challenges in the use of single-cell analysis

There are also several challenges with pursuing a single-cell analysis strategy to mapping tissues. Here we describe three of those challenges in more detail: collection and preanalytical processing of tissues, measurement uncertainty, and multiplexing. The collection of tissues introduces many ethical, legal, and social challenges, particularly for broad, open consent of how those tissues and associated data may be used. As part of the collection process, ideally the position and orientation of the tissue is collected, so that it can be easily referenced to a common coordinate framework, though for many reasons this may not be possible; the smaller the biospecimen and the less contextual information that is available, the harder the challenge for integrating any associated data into bodywide map. Ideally, we can also capture details and the location of all biomolecule present in a given tissue; however, in practice, we can only study a small subset of the biomolecules present, in part driven by assay availability and ease of multiplexing.

Preanalytical processing may induce distortions or potential skew measurements. For example, isolating single cells from certain tissues for sequencing is often very difficult, whereas isolating single nuclei in these instances is far more feasible. This resulted in discovery of single nuclei sequencing that has yielded a wealth of information about the new subtypes of resident neuronal cells in the brain (Lake *et al.*, 2016). Single-nuclei sequencing (sNuc-seq/snRNA-seq) has also been specifically used to enrich for newly transcribed genes in the broader context (Lacar *et al.*, 2016). Moreover, it profiles rapid cellular responses and measures newly synthesized nuclear RNAs devoid of rRNAs, which can be overrepresented in total transcriptome (Lake *et al.*, 2016). Further, sNuc-seq can facilitate profiling of difficult biosamples, such as frozen or preserved tissues that often exhibit inseparable cells. For these samples, it is easier to generate a nuclear suspension rather than a single-cell suspension. Although the overall data quality might be slightly inferior to typical single-cell transcriptomic data, the output still allows for discovery of potential new RNA species and/or new cell types. Hence, for analyzing challenging biospecimens (e.g., tissue biopsies or healthy tissue sections), sNuc-seq is better suited than the traditional single-cell sequencing, although the potential for skewing results presents a challenge.

Another area of challenge is in the accuracy, sensitivity, and specificity of measurements. For instance, batch effects and variability in biological samples could be confounding and problematic for further downstream analysis. Generally, there are two broad types of variability: technical variability and biological variability. Technical variability arises due to changes in sample quality and processing but could also arise, for example, due to changes in library preparation or sequencing technology that might vary from lab to lab. Biological variability, on the other hand, might arise due to differences in patient-derived samples or environmental or natural genetic perturbation of derived biological samples. Multiplexing assays may be one method for addressing some of these challenges (Krutzik and Nolan, 2006; Zheng et al., 2018). First, it might eliminate technical batch effects as cells from different samples could be mixed together and tested in a single experiment, followed by mapping of each cell types back to its sample of origin (demultiplexing). Second, multiplexing generally detects and accounts for confounding cell "doublets" as often two cells remain associated even under most stringent conditions (Kang et al., 2018). Because doublets from two samples will exhibit both sample "barcodes" in a multiplex experiment, they can be identified even from the pool (Butler et al., 2018). Furthermore, in multiplex experiments, engineered perturbation methods like CRISPR can be used to generate genome/epigenome diversity across a variety of single cells. A recent study shows the potential of coupling CRISPR perturbations and single-cell RNA sequencing for pooled genetic screens by optimizing CROP-seq via guide RNA barcoding (Hill et al., 2018). Given that measurements are done in a small volume on a diverse range of biomolecules over a large dynamic range and in a high-throughput manner, it is difficult to balance this against achieving unit sensitivity. Likewise, it is difficult to achieve specificity for all relevant biological variation and to accurately map complex molecular structures (Hill et al., 2018).

A further challenge is the choice of which biomolecules to study. For example, transcript levels by themselves may not accurately predict protein levels in many scenarios and to thus explain genotype–phenotype relationships, high-quality data quantifying different levels of gene expression are indispensable for the complete understanding of biological processes (Liu et al., 2016). Furthermore, given the dynamic nature of cell states, variations in the genome, epigenome, or protein content can profoundly alter cell function. Therefore, it is insufficient to rely only on analyzing total transcriptome to accurately define dynamics of cellular state or function. For instance, often posttranscriptional modifications are critical for cell function and define the state of a given cell, such as changes in protein phosphorylation during cell signaling. It is also well documented that variation in genomic sequences can affect epigenetic changes and that alterations in chromatin states (e.g., changes in DNA and/or histone methylation) profoundly affect functional gene expression (Rozenblatt-Rosen et al., 2017). Owing to these complications, it is desirable to multiplex different assays together for simultaneous measurement of multiple molecular phenotypes (multimodal analysis) where possible. Thus, analyzing epigenetic changes or protein modifications together with RNA could provide a greater clarity as to the identity as well as function of a cell. Emerging methods attempt to combine various parameters, such as single-cell methylation profiling (by bisulfite sequencing) or nucleosome/chromatin accessibility (ATAC-seq) and transcriptomics, although none have demonstrated single-cell resolution in human tissue to date. In addition to total transcriptomics, incorporation of scNuc-seq (nuclear RNA) into these multimodal assays, particularly in the case of challenging tissues, could provide a better understanding of the relationship between epigenetic/chromatin changes and transcription initiation (Rozenblatt-Rosen et al., 2017).

A final challenge is that analysis may be biased by limitations on what we can measure. We do not currently have single-cell assays for studying posttranscriptional and posttranslational modifications, lipids, metabolites, or exogenous molecules such as drugs or chelated heavy metals. Given the large number of cells that need to be studied for a comprehensive view of the human body, analytical techniques are likely to be skewed toward lower cost, high-throughput techniques with lower spatial resolution that may have lower sensitivity and specificity and have limited dynamic range. This may limit our understanding of low-expression, heavily modified, and multiple variants of proteins or nucleic acids and the role of subcellular localization in defining cell state. The snapshot, observational nature of preserved tissues is also a limitation to the information collected, and there is a significant need for functional assays to decipher the link between function and molecular state.

BUILDING A HUMAN BIOMOLECULAR ATLAS

The postgenomic revolution in high-content, high-throughput technologies has led to a shift in scientific approach, moving more to include "discovery-driven" approaches as well as "hypothesis-driven" approaches. The rapid dissemination and iteration on single-cell RNA sequencing assays combined with increasingly powerful computational methods to characterize and impute cellular networks highlights how potential understanding can be transformed by discovery-driven approaches. The emergence of spatial transcriptomics (Lein et al., 2017) and in situ sequencing (Lee et al., 2014) enhance these deep molecular profiles with spatial information providing an unprecedented view of the organization of tissue, cellular heterogeneity in situ, and the corresponding functional complexity. Although there are significant challenges for analyzing metabolites, lipids, and comprehensive, unbiased proteomics at cellular resolution, single-cell analysis has reached a point where we can assay the complexity of cellular organization in tissue in a high-throughput, high-content, and reproducible manner and generate high-resolution spatial maps of molecular profiles of resident cells and their extracellular environment that have the potential to provide new insights into how the organization of cell types and their state influence overall tissue function and dysfunction.

There is a significant hurdle, however, in going from a tissue specimen to the full human body, and there are many challenges in integrating the data into a holistic view. Building a comprehensive biomolecular atlas would require not only far-reaching experimental approaches but also sophisticated computational methods, integration of existing knowledge, and coordination and collaboration across many communities.

Ongoing and emerging atlas projects

There is a long history of anatomical atlases of the human body dating back more than 500 years. It is now been more than 30 years since the National Library of Medicine started the Visible Human Project (Ackerman, 2017), which generated submillimeter resolution computed tomography, magnetic resonance imaging, and photographic images of both a male and a female body. More recently, the Visible Korean and Chinese Visible Human Projects expanded this data set to include additional human bodies (Dai et al., 2012). These projects highlighted some of the challenges with constructing an atlas of the human body, notably that preservation techniques distort tissue, preanalytical processing can destroy structures or create uneven effects, that it is difficult to prevent artifacts related to the collection and preservation of the tissue, that standards are

hard to follow uniformly on large-scale projects, and that pathologies can be identified in tissues expected to be normal. Similar challenges were faced by the National Institutes of Health (NIH) Common Fund Gene-Tissue Expression (GTEx) program that started in 2009 and has mapped gene expression levels in more than 50 tissue types across more than 500 donors (Consortium, 2013). This rich data set, based on bulk measurements from tissue blocks, found that local variation affects gene expression for many genes and has been able to identify new disease-associated variations, potential drug targets, and tissue-specific hereditary disorders.

Concurrently over the past decade, the rapid advance of information communication technology; the development of sophisticated controlled vocabularies, ontologies, and semantics; as well as tools for visualizing and modelling multiscale, multidimensional data have dramatically changed our understanding of cell types. The limitations of commonly used cell nomenclatures have become particularly apparent in uniquely identifying functionally distinct cell types and states. Over the past decade, a number of projects have emerged, such as CELLPEDIA (Hatano et al., 2011), CellFinder (Stachelscheid et al., 2014), and LifeMap (Edgar et al., 2013), that have taken a systematic, data-driven approach to identifying and defining the molecular characteristics of cells. These projects have highlighted the complexity of interpreting biomolecular data to define cell type and state and the need for unbiased, high-throughput biomolecular analysis of at least millions of cells.

Identification of cells in the immune system has progressed most rapidly in this respect because of the ease through which fluids can be analyzed. The development of high-content, high-throughput flow cytometry and enhancements such as CyTOF combined with investment in collaborative projects such as the Immunological Genome Project (Shay and Kang, 2013) have resulted in a systematic approach to reconstructing the gene regulatory networks present in the immune cells and the relationship between cell types and state based on many factors, including lineage and activation. The work of this project has primarily focused on mice, though increasingly human samples are being analyzed, resulting in a deeper understanding of how disease and therapies influence the whole immune system.

Collaborative projects on biomolecular mapping of cells in solid mammalian tissues and organs are rapidly expanding our knowledge. Consortia such as the GenitoUrinary Development Molecular Anatomy Project (Harding et al., 2011) and LungMAP (Ardini-Poleske et al., 2017) and projects like the Salivary Gland Molecular Anatomy Project (Musselmann et al., 2011) are developing molecular profiles of tissues across different mammalian systems, life stages and pathologies. The brain has long been an organ of interest, with significant work by the Allen Institute for Brain Science among others in mapping first the anatomical characteristics of the brain and then adding functional, morphological, and molecular information to provide a more detailed view of the cell types present. The BRAIN Initiative Cell Census Network (Ecker et al., 2017), launched in 2017, aims to expand this work to build a comprehensive reference of cell types present in the human, monkey, and mouse brain using an integrated and multiplexed molecular, anatomical, and physiological approach.

Complementing GTEx, over the past decade the Functional Annotation of the Mammalian Genome (FANTOM) Consortium (Lizio et al., 2017) and the Human Protein Atlas (HPA) project (Uhlen et al., 2010) have both collected an increasingly extensive molecular data set from multiple tissues across the human body to build a comprehensive view of biomolecular variation, RNA in the case of FANTOM and proteins in the case of HPA, across different regions of the body

as well as across individuals. Building on the work of these consortia and projects, a grass-roots community of researchers came together in 2016 to form the Human Cell Atlas Consortium that seeks to build reference maps for all human cells, with a first draft composed of gene expression profiles for at least 30 million cells expected to be released in 2018 (Consortium, 2017). To complement these approaches, the NIH recently launched a new collaborative project starting in 2018 called the Human BioMolecular Atlas Program with the goal of catalyzing development of an open, global framework for comprehensively mapping the human body at a cellular resolution (National Institutes of Health, 2018). The focus of this program is to study the rich spatial context of cells in situ and develop technologies and techniques that overcome the need to remove cells from their complex tissue environment for analysis. This program will work through collaboration with the other programs to support the development of community standards for data management and analysis that enable cross-querying of data from multiple sources, new technologies for increasing the throughput and diversity of biomolecules that can be mapped, and single-cell resolution maps of diverse tissue types.

The need for international coordination and collaboration

Although there has been rapid progress in the developing and validating new technologies that can generate high-resolution, high-content, and high-throughput data, no individual project has the time or resources to map the entire human body. With the emergence of many programs working on different aspects of the human body, there is a timely opportunity to work together and share resources, methods, data, and knowledge to spur further scientific discoveries. By integrating expertise from multiple domains into a synergistic research plan and pursuing a mission that addresses a shared need, international partnerships can result in knowledge gains that would not have happened otherwise.

Increasingly, international consortia have been established to promote coordination and collaboration among national projects, such as the International Human Epigenetics Consortium (Stunnenberg et al., 2016), the International Mouse Phenotyping Consortium (Brown and Moore, 2012), and the International Cancer Genome Consortium (Zhang et al., 2011). Many of these consortia focus on creating an environment for data exchange based on all partners contributing data that is accessible to all. Creating this environment requires discussing and addressing many issues that are typically outside the purview of individual grants, including multijurisdiction policy, ethical and legal issues; data and metadata types, formats, and standards; establishing and managing shared resources to ensure equality, usability, integrity, and security; establishing systematic and rapid sharing of information without significant transaction costs; and developing mechanisms for the swift release of contributed data to make it findable, accessible, interoperable, and reusable by the research community.

In building international collaborations around mapping the human body at cellular resolution, there are several key challenges that need to be resolved. On the policy, ethical, and legal side, we need to establish donor consent processes that maximize reuse of biospecimens and data for research purposes, while respecting privacy, the right to anonymity, and shared interest in the research results. Experimentally, we need to establish protocols for minimizing degradation of tissue once collected from a donor and prepared for analysis, while also maximizing its use for multiple assays. Assays need to be calibrated and validated, so the data generated are trustworthy and reproducible, and this will likely require sharing protocols, carrying out multisite comparison studies,

and personnel exchanges. Tissue specimens also need to be collected and examined with sufficient information that the location of the orientation of a small volume of analyzed tissue can be placed back into the large context of the human body using a common coordinate system and integrated ontologies. This system needs to be robust against assays at different spatial resolutions and across many different types of measures as well as variation in anatomy. Integrating multidimensional spatial and molecular information for analysis is a complex challenge that will also require international cooperation and flexibility. The volume of data and the resources required to analyze, visualize, and model it as well as make access flexible and sustainable will also be a significant challenge for any international collaboration.

Arguably there is no end to information that can be added to a human atlas as new assays are developed and sensitivity and resolution improved; so what would make a successful atlas? An early goal of cell census programs is to provide deeper and more robust biomolecular descriptions to commonly defined cell types. One practical outcome of this goal may be more rigorous validation and description of cell types and states in the peer-reviewed literature. An atlas defining how to identify and where to find these cells in the human body would provide a reference for researchers collecting primary human cells or evaluating diseased or dysfunctional tissues. Going beyond this linkage to existing cell types, another goal is to complete an atlas that methodically and robustly identifies and describes in a hierarchical taxonomy all human-origin cells. One practical outcome of this goal would be the identification of the molecular characteristics of cells that play an active role in cellular circuits and establishing linkage between the presence and quantify of particular cell types or biomolecules and nonspecific, nondestructive measurements in clinical imaging that can be used for diagnostic purposes. As with the sequencing projects, a draft atlas may involve releases identify cell types within specific organs and using a subset of biomolecules, for example, based on transcriptomes, that gets expanded and refined with subsequent releases. Moving beyond spatial information and cell type, a third goal further in the future could be to create a reference atlas that maps changes over time, linking together the lineage of cells throughout the lifespan as well capturing details of how cyclic chronobiology influences different cells and tissues. Complementing these normal reference atlases will be atlases describing and mapping disease and dynfunctional tissues and their associated biomolecular states. Ultimately, the success of any atlas is whether the community finds value in the information it contains, maintains and refines it, and uses it for new discoveries or improves existing processes.

There is a significant and opportune moment for the current and emerging projects to work together toward establishing a framework that will result in integrated analysis of data from multiple contributors. Current technologies limit any individual project to examining significantly less than 1% of the cells in the human body at any significant molecular depth; however, by working together over the next decade there is a potential to establish the framework for realizing a draft of what the human body looks like at the individual cell level.

REFERENCES

Ackerman MJ (2017). The Visible Human Project: from body to bits. IEEE Pulse 8, 39–41.

Ajami B, Samusik N, Wieghofer P, Ho PP, Crotti A, Bjornson Z, Prinz M, Fantl WJ, Nolan GP, Steinman L (2018). Single-cell mass cytometry reveals distinct populations of brain myeloid cells in mouse neuroinflammation and neurodegeneration models. Nat Neurosci 21, 541–551.

Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, Borowsky AD, Levenson RM, Lowe JB, Liu SD, Zhao S, et al. (2014). Multiplexed ion beam imaging of human breast tumors. Nat Med 20, 436–442.

Ardini-Poleske ME, Clark RF, Ansong C, Carson JP, Corley RA, Deutsch GH, Hagood JS, Kaminski N, Mariani TJ, Potter SS, et al. (2017). LungMAP: The Molecular Atlas of Lung Development Program. Am J Physiol Lung Cell Mol Physiol 313, L733–L740.

Bandyopadhyay S, Fisher DAC, Malkova O, Oh ST (2017). Analysis of signaling networks at the single-cell level using mass cytometry. Methods Mol Biol 1636, 371–392.

Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, Vitale L, Pelleri MC, Tassani S, Piva F, et al. (2013). An estimation of the number of cells in the human body. Ann Hum Biol 40, 463–471.

Bjornson ZB, Nolan GP, Fantl WJ (2013). Single-cell mass cytometry for analysis of immune system functional states. Curr Opin Immunol 25, 484–494.

Brown SD, Moore MW (2012). Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. Dis Model Mech 5, 289–292.

Butler A, Hoffman P, Smibert P, Papalexi E, Satija R (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36, 411–420.

Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357, 661–667.

Chen C, Xing D, Tan L, Li H, Zhou G, Huang L, Xie XS (2017). Single-cell whole-genome analyses by linear amplification via transposon insertion (LIANTI). Science 356, 189–194.

Coleman RA, Liu Z, Darzacq X, Tjian R, Singer RH, Lionnet T (2015). Imaging yranscription: past, present, and future. Cold Spring Harb Symp Quant Biol 80, 1–8.

Consortium GT (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580–585.

Consortium TH (2017). The Human Cell Atlas White Paper, vol. 2018.

Dai JX, Chung MS, Qu RM, Yuan L, Liu SW, Shin DS (2012). The Visible Human Projects in Korea and China with improved images and diverse applications. Surg Radiol Anat 34, 527–534.

Eberwine J (2017). Down the rabbit hole of single-cell genome analysis. Mol Cell 66, 304–305.

Ecker JR, Geschwind DH, Kriegstein AR, Ngai J, Osten P, Polioudakis D, Regev A, Sestan N, Wickersham IR, Zeng H (2017). The BRAIN Initiative Cell Census Consortium: lessons learned toward generating a comprehensive brain cell atlas. Neuron 96, 542–557.

Edgar R, Mazor Y, Rinon A, Blumenthal J, Golan Y, Buzhor E, Livnat I, Ben-Ari S, Lieder I, Shitrit A, et al. (2013). LifeMap Discovery: the embryonic development, stem cells, and regenerative medicine research portal. PLoS One 8, e66629.

Eng CL, Shah S, Thomassie J, Cai L (2017). Profiling the transcriptome with RNA SPOTs. Nat Methods 14, 1153–1155.

Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, Bordwell A, Can A, Corwin A, Dinn S, Filkins RJ, et al. (2013). Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proc Natl Acad Sci USA 110, 11982–11987.

Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vasquez G, Black S, Nolan G (2018). Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. bioRxiv doi: https://doi.org/10.1101/203166.

Haque A, Engel J, Teichmann SA, Lonnberg T (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med 9, 75.

Harding SD, Armit C, Armstrong J, Brennan J, Cheng Y, Haggarty B, Houghton D, Lloyd-MacGilp S, Pi X, Roochun Y, et al. (2011). The GUDMAP database—an online resource for genitourinary research. Development 138, 2845–2853.

Hatano A, Chiba H, Moesa HA, Taniguchi T, Nagaie S, Yamanegi K, Takai-Igarashi T, Tanaka H, Fujibuchi W (2011). CELLPEDIA: a repository for human cell information for cell studies and differentiation analyses. Database (Oxford) 2011, bar046.

Hill AJ, McFaline-Figueroa JL, Starita LM, Gasperini MJ, Matreyek KA, Packer J, Jackson D, Shendure J, Trapnell C (2018). On the design of CRISPR-based single-cell molecular screens. Nat Methods 15, 271–274.

Huang L, Ma F, Chapman A, Lu S, Xie XS (2015). Single-cell whole-genome amplification and sequencing: methodology and applications. Annu Rev Genomics Hum Genet 16, 79–102.

Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, Wan E, Wong S, Byrnes L, Lanata CM, et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat Biotechnol 36, 89–94.

Krutzik PO, Nolan GP (2006). Fluorescent cell barcoding in flow cytometry allows high-throughput drug screening and signaling profiling. Nat Methods 3, 361–368.

La Manno G, Soldatov R, Hochgerner H, Zeisel A, Petukhov V, Kastriti M, Lonnerberg P, Furlan A, Fan J, Liu Z, et al. (2017). RNA velocity in single cells. bioRxiv doi: https://doi.org/10.1101/206052.

Lacar B, Linker SB, Jaeger BN, Krishnaswami S, Barron J, Kelder M, Parylak S, Paquola A, Venepally P, Novotny M, et al. (2016). Nuclear RNA-seq of single neurons reveals molecular signatures of activation. Nat Commun 7, 11022.

Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, Wildberg A, Gao D, Fung HL, Chen S, et al. (2016). Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. Science 352, 1586–1590.

Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, Turczyk BM, Yang JL, Lee HS, Aach J, et al. (2015). Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nat Protoc 10, 442–458.

Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, Terry R, Jeanty SSF, Li C, Amamoto R, et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. Science 343, 1360–1363.

Lein E, Borm LE, Linnarsson S (2017). The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. Science 358, 64–69.

Levesque MJ, Ginart P, Wei Y, Raj A (2013). Visualizing SNVs to quantify allele-specific expression in single cells. Nat Methods 10, 865–867.

Liu Y, Beyer A, Aebersold R (2016). On the dependency of cellular protein levels on mRNA abundance. Cell 165, 535–550.

Lizio M, Harshbarger J, Abugessaisa I, Noguchi S, Kondo A, Severin J, Mungall C, Arenillas D, Mathelier A, Medvedeva YA, et al. (2017). Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals. Nucleic Acids Res 45, D737–D743.

Luo C, Keown CL, Kurihara L, Zhou J, He Y, Li J, Castanon R, Lucero J, Nery JR, Sandoval JP, et al. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science 357, 600–604.

Moffitt JR, Zhuang X (2016). RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH). Methods Enzymol 572, 1–49.

Musselmann K, Green JA, Sone K, Hsu JC, Bothwell IR, Johnson SA, Harunaga JS, Wei Z, Yamada KM (2011). Salivary gland gene expression atlas identifies a new regulator of branching morphogenesis. J Dent Res 90, 1078–1084.

National Institutes of Health (2018). The Human BioMolecular Atlas Program, vol. 2018.

Preissl S, Fang R, Huang H, Zhao Y, Raviram R, Gorkin DU, Zhang Y, Sos BC, Afzal V, Dickel DE, et al. (2018). Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat Neurosci 21, 432–439.

Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, Noble WS, Duan Z, Shendure J (2017). Massively multiplex single-cell Hi-C. Nat Methods 14, 263–266.

Rodriguez-Muela N, Litterman NK, Norabuena EM, Mull JL, Galazo MJ, Sun C, Ng SY, Makhortova NR, White A, Lynes MM, et al. (2017). Single-cell analysis of SMN reveals its broader role in neuromuscular disease. Cell Rep 18, 1484–1498.

Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA (2017). The Human Cell Atlas: from vision to reality. Nature 550, 451–453.

Schubert W, Bonnekoh B, Pommer AJ, Philipsen L, Bockelmann R, Malykh Y, Gollnick H, Friedenberger M, Bode M, Dress AW (2006). Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. Nat Biotechnol 24, 1270–1278.

Sender R, Fuchs S, Milo R (2016). Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. Cell 164, 337–340.

Shah S, Lubeck E, Zhou W, Cai L (2016). In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. Neuron 92, 342–357.

Shay T, Kang J (2013). Immunological Genome Project and systems immunology. Trends Immunol 34, 602–609.

Stachelscheid H, Seltmann S, Lekschas F, Fontaine JF, Mah N, Neves M, Andrade-Navarro MA, Leser U, Kurtz A (2014). CellFinder: a cell data repository. Nucleic Acids Res 42, D950–D958.

Stunnenberg HG, International Human Epigenome C, Hirst M (2016). The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. Cell 167, 1897.

Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S, et al. (2010). Towards a knowledge-based Human Protein Atlas. Nat Biotechnol 28, 1248–1250.

Valentine JW, Collins AG, Meyer CP (1994). Morphological complexity increase in metazoans. Paleobiology 20, 131–142.

Vickaryous MK, Hall BK (2006). Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. Biol Rev 81, 425–455.

Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, Liang Y, Rivkin E, Wang J, Whitty B, et al. (2011). International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. Database (Oxford) 2011, bar026.

Zheng S, Papalexi E, Butler A, Stephenson W, Satija R (2018). Molecular transitions in early progenitors during human cord blood hematopoiesis. Mol Syst Biol 14, e8041.

Zheng Y, Huang X, Kelleher NL (2016). Epiproteomics: quantitative analysis of histone marks and codes by mass spectrometry. Curr Opin Chem Biol 33, 142–150.

Ziegenhain C, Vieth B, Parekh S, Reinius B, Smets M, Leonhardt H, Hellmann I, Enard W (2017). Comparative analysis of single-cell RNA sequencing methods. Mol Cell 65, 631–643.