Mapping the human

The Human Cell Atlas, driven by a collaborative spirit and rapid advances in single-cell methods, is poised to advance both biological understanding and technical development.

At *Nature Methods* we have been publishing methods to profile single cells for some time, but even we did not quite anticipate that they would drive a project as ambitious as the Human Cell Atlas (HCA), which released its first data sets in early April of this year.

The HCA intends to map the cell types of the human body. Though the project is at present largely driven by single-cell RNA sequencing, different types of maps are possible. Molecular mapping yields transcript, protein or epigenetic profiles of a given cell type. Functional and morphological mapping links these profiles to cellular behavior and appearance. Spatial mapping localizes cells within a tissue or organ. As the Human Genome Project is for genes, the HCA is intended as a reference for human cells.

It is by no means the only such project. The Allen Brain Atlas has for many years driven mapping projects in the brain, some at the single-cell level. The NIH Human BioMolecular Atlas Program (HuBMAP) funds the development of single-cell technologies and their application to tissue mapping. The Human Protein Atlas is a reference of staining patterns for over 10,000 gene products using immunohistochemistry, recently extended to the localization of proteins within cell lines. Finally, several initiatives profile single cells in model organisms, a needed complement to work in the human. Approaches both informal and formal are reportedly being taken to enable communication between the HCA and other projects. While it will not be possible to avoid some duplication of effort, it is important that such communication be actively pursued.

To join the HCA, scientists must do no more than commit to its values; this includes early sharing of data and methods, reflecting the open and collaborative nature of this scientist-driven project. The relatively loose, grassroots structure of the HCA undoubtedly brings with it both opportunities and challenges. Such a strategy distributes effort, maximizes scientific input, and most probably improves the chance of reaching the goal. But marshaling a large global coalition of scientists can be logistically challenging and could make it more difficult to reach consensus on questions both organizational and scientific.

Such questions abound, as they must for a project of this scale. How, for instance, does one even best define a cell type based on single-cell molecular profiles, as opposed to the more traditional approach of a few stable markers or a distinct morphology or location? Single cell-omes change not only with cell type, but also with cell state, subject to myriad influences. Indeed, the originators of the project propose that the HCA is likely to modify our very understanding of these categories.

But as data are produced, decisions have to be made about how cell types are annotated in the atlas, and some standards will need to be agreed upon for a coherent reference to be generated. Ideally, steps in data processing and interpretation should be transparent to users, and, where relevant, prior knowledge should be integrated. New ways to display such high-dimensional data meaningfully need to be worked out. These steps are made all the more challenging by the range of potential users for the atlas—from single-cell mavens to biologists and bioinformaticians with no such expertise, to possibly even clinicians.

Further complicating matters, methods for single-cell molecular studies are still in active development. No doubt the atlas will be refined as methods improve, and these early data sets are a much needed resource for the methods-development process itself. The HCA is also particularly well placed to periodically organize systematic methods comparisons, in the process establishing benchmarks against which future methods can be compared.

It is unlikely that a single method will meet all needs. Single-nucleus sequencing, for instance, is proving useful for tissues that are difficult to dissociate, such as the brain. Droplet-sequencing-based approaches are high throughput, allowing many cells to be profiled, but at lower sequencing depths. Microwell-based approaches, by contrast, typically have higher transcript capture rates and are useful for obtaining deeper profiles of certain cell populations. It will be important to develop several complementary methods to solve a range of problems.

On the computational side, it is critical that data generated by different approaches can be compared. In this issue a News and Views (p321) discusses two methods published in *Nature Methods*: one to project cells from a single-cell experiment onto a second reference data set, and the other to align time-ordered single cells and their gene expression patterns across experiments. Two recent papers in *Nature Biotechnology* describe methods to correct for batch effects in single-cell RNA-seq data and thus to integrate data from multiple experiments. These types of tools are needed to both construct an atlas and make full use of it as a standard for comparison.

The final shape of the HCA will probably not be apparent for many years, but there is little doubt that it will mark the way for explorers of both the biological and the methodological kind.