BUILDING A HUMAN CELL ATLAS: KEY QUESTIONS FOR DISCUSSION

We expect the following considerations to be actively discussed over the course of our two-day meeting. We look forward to hearing your input!

What are the potential benefits of a Human Cell Atlas?
- What questions will it allow us to answer that have not been tractable before? What new questions might it raise?
- What types of new science might a Human Cell Atlas enable? What new technologies might it launch?

What should the scope of the Human Cell Atlas be?
- Should we plan only for a human cell atlas, or also consider analogous/similar efforts in other species (*e.g.*, key model systems)? If the latter, how could these projects interact for maximum scientific benefit?
- How can we leverage disease-relevant tissue and/or human genetic variation to assess the scope of differences in cell states within a reasonable scope?
- How can we best balance the needs for breadth and depth as we select tissues for inclusion?
- Which reference list of tissues would best suit the project's goals?
- Should we consider only adult tissues, or prenatal tissues as well? Should we consider *in vitro* samples from cell culture or organoids?
- What cell types would best suit the project (*e.g.*, iPSC-derived)?
- What sample sizes will be required (in terms of tissues, organs, cells)? What sampling frameworks will allow us to draw valid conclusions from a limited number of cells?
- Are there specific diseases that should be the subject of special focus?
- Should we first focus on a few tissues or organs to illustrate the consortium's coherence and potential?
- What might a successful pilot encompass?

Where and how should we source samples for the project?
- Should we consider only resected tissue, or also include transplant donors?
- Should we prioritize having a diverse swath of donors (in terms of age, genetics, *etc.*) or prioritize homogeneity?
- Should we pursue single, individual donors with key characteristics, or consider only groups?

Which technologies should we deploy to generate data?
- Will we need a single unified pipeline, or should we pursue multiple approaches to suit diverse needs throughout the project?
- Should we focus on high-throughput approaches, lower-throughput ones that offer greater depth, or a combination of the two?
- Should one technology (*e.g.*, single-cell RNA sequencing) be used for primary data generation? If so, what other technologies (*e.g.*, mass cytometry) will be best suited to meet additional needs (*e.g.*, increased scale, spatial resolution)?
- What considerations will be most important for deploying robotics?
- In the clinical community, how can we achieve unified approaches to processes that currently vary widely in execution (*e.g.*, tissue acquisition, disaggregation, etc.)?

Which computational strategies will best enable data analysis?

- As above with data generation, will we need a single unified pipeline for data analysis, or should we pursue multiple approaches?
- Should we release our data as soon as possible, as openly as possible? What ethical issues would such a strategy raise?
- How can we leverage the latest engineering advances -- from cloud computing to UI design -- to ensure both the most effective and open interfaces and efficient computation of massive datasets?
- How can we best serve and display data and processed results to the world?

How should we organize the activities of our scientific consortium?

- Which core pillars of activity and expertise should we organize around?
- What is the role and relationship of technological and computational groups to the larger consortium?
- How should we structure our activities to ensure quality assurance, compliance and data security?
- Which group(s) should take on critical questions and decisions around human anatomy and physiology, from clinical sample acquisition to expertise in normal physiology, disease biology and development?
- What should be the role of experts in model organisms?

How should we create and structure our scientific consortium?

- How can we best ensure participation from relevant communities? How can we best include members of the genomics, bioinformatics, and biological, and clinical communities? How should we ensure engagement from each of these groups?
- Who has standing to convene and represent the consortium? Should there be standing and rotating members?
- How should the scientific consortium engage and interact with funders?
- How should the scientific consortium engage and collaborate with pharmaceutical and biotech companies? With technology companies?
- How can we best ensure that we are intellectually inclusive, open-minded, and not overlap bounded, while maintaining focus and productivity?
- How can we structure our consortium in a way that suits its many diverse members (*e.g.*, loose affiliation, federation, centralization)?

What principles should underlie our activities?

- Should there be immediate and open access to all data? To all experimental protocols?
- If so, how can we ensure the compliance standards associated with human genetic information?
- What approaches should guide international cooperation and collaboration?
- How can we ensure that we have committed standards for quality and reproducibility?
- How can we ensure equity in the analyzed samples, by geographic, gender, and ethnicity considerations?
- What guidelines will exist for membership and standing?