HUMAN CELL ATLAS

2023 Human Cell Atlas General Meeting

Day 1 - 10 July 2023

8:30 AM - 1:00 PM EDT (4 Hours, 30 Min)

Day 1 - Part 1

9:00 AM - 9:10 AM EDT (10 Min)

Welcoming Remarks

Main Hall

HCA Science and where we are with integration efforts

Gary Bader
Principal Investigator
University of Toronto
Speaker

9:10 AM - 9:40 AM EDT (30 Min)

State of HCA

Main Hall

Dana Pe'er
Principal Investigator
Sloan Kettering Institute
Moderator

Aviv Regev
Principal Investigator
Genentech
Speaker

Sarah Teichmann
Head of Cellular Genetics and Se...
Sanger Institute
Speaker

9:40 AM - 10:25 AM EDT (45 Min)

Session 1: Human Cell Atlas: Progress and applications

Main Hall

1. **Philip L. De Jager:** Uncovering the causal chain of cellular events leading to Alzheimer's disease
2. **Neil Henderson:** Multimodal decoding of human liver regeneration
3. **Tuuli Lappalainen:** Functional variation in the human genome: Lessons from the transcriptome

Dana Pe'er
Principal Investigator
Sloan Kettering Institute
Moderator

Tuuli Lappalainen
Principal Investigator
New York Genome Center & Colu...
Speaker

Philip De Jager
Principal Investigator
Columbia University Medical Cen...
Speaker

Neil Henderson
Principal Investigator
University of Edinburgh
Speaker

10:25 AM - 10:55 AM EDT (30 Min)

Refreshment Break

10:55 AM - 12:30 PM EDT (1 Hour, 35 Min)

Session 2: Evaluating and Enhancing our Approach to Building Atlases

Main Hall

1. **Malte Luecken:** Update on Atlas integration pilots
2. **Ed Lein:** From draft to comprehensive human and comparative brain cell atlases
3. **Alexandra-Chloe Villani** and **Shyam Prabhakar**: Insights from integration efforts
4. **Panel discussion:** How do we define Atlas 1.0?
5. **John Randell:** Updates from the HCA Data Ecosystem

John Marioni
Group Leader
EMBL-EBI & University of Cambri...
Moderator

Alexandra-Chloe Villani
Principal Investigator
Broad Institute, Massachusetts G...
Panelist

Aviv Regev
Principal Investigator
Genentech
Panelist

Malte Luecken
Principal Investigator
Helmholtz Munich
Speaker

Shyam Prabhakar
Senior Group Leader
Genome Institute of Singapore
Panelist

Sarah Teichmann
Head of Cellular Genetics and Se...
Sanger Institute
Panelist

Edward Lein
Principal Investigator
Allen Institute for Brain Science
Speaker

Dana Pe'er
Principal Investigator
Sloan Kettering Institute
Panelist

John Randell
Chief Alliance Officer
Human Cell Atlas
Speaker

12:30 PM - 2:00 PM EDT (1 Hour, 30 Min)

HCA Ethics Working Group (EWG) Meeting (observers welcome - registration required)

CR2

12:30 PM - 2:00 PM EDT (1 Hour, 30 Min)

Lunch

1:30 PM - 3:00 PM EDT (1 Hour, 30 Min)

Day 1 - Part 2

WATCH RECORDING

2:00 PM - 2:45 PM EDT (45 Min)

Session 3: Keynote #1

Main Hall

What makes a stem cell a stem cell and how does it go bad

Sagi Abelson
Early-Career Investigator
Ontario Institute for Cancer Rese...
Moderator

John Dick
Principal Investigator
Princess Margaret Cancer Centre,...
Keynote

2:44 PM - 4:01 PM EDT (1 Hour, 17 Min)

Session 4: Breakout Session #1: Toward better integrations

Topic #1: Increasing accessibility/building great and diverse scientific communities

Topic #2: Strengthening integration pipelines

Topic #3: Ancestral diversity

2:45 PM - 4:00 PM EDT (1 Hour, 15 Min)

(Breakout Session 1) Topic #1: Increasing accessibility/building great scientific communities

CR3

WATCH RECORDING

Topic Description:

The field of single-cell biology is experiencing a surge in both the number of researchers and the heterogeneity of disciplines involved. This rapid growth is accompanied by the increasing availability of data across different modalities, technologies, organs, pathologies, and organisms. This growth has led to the development of innovative computational tools to aid analysts in gaining insights from what may otherwise be very large, complex, and noisy data. However, as the field expands, it becomes increasingly challenging to navigate the vast array of data and computational tools and to prioritize and design new effective tools. Furthermore, while new single-cell technologies have made significant advancements, the cost of data generation still poses limitations. To address these challenges, it is crucial for the scientific community to come together and collaborate. In this session, we aim to highlight several key topics where the community can make valuable contributions and discuss the obstacles we face. By engaging in constructive discussions, we can explore alternative approaches and find new avenues for collaboration and progress.

Key points to discuss:

1. Technological breakthroughs related to data analysis: More than 1,400 tools are currently available to analyze scRNA-seq data. Just a few are user-friendly and don't require bioinformatic skills, such as APIs. Lack of benchmarking and best-practice workflows makes navigating the broad field of novel tools challenging.

2. Data generation and access are still limited: high cost associated with single-cell data generation and access, which restricts the accessibility of single-cell technologies to a broader scientific community and hinders widespread adoption and collaboration. Another challenge lies in the technical complexity of data generation and analysis. Single-cell technologies often involve intricate experimental procedures and protocols, requiring expertise and specialized training.

Here are some questions to foster discussion:

1. What are the main obstacles you see preventing the wider availability of data access and analysis in our community? (QUIZ - https://www.polleverywhere.com/word-cloud)

2. How can we bridge the gap between experimentalists and computational biologists to facilitate effective data analysis and interpretation?

3. What steps can be taken to incentivize the development of open-source and user-friendly software tools? Ex.: Foster interdisciplinary collaborations between researchers in biology, computer science, and industry; Bioconductor packages that can be used directly from python; Consortium of foundational tools such as scverse

4. Are there any specific challenges or considerations when developing user-friendly tools for non-experts? How can we address them effectively?

5. What strategies can be implemented to promote collaborations and knowledge sharing within the scientific community? Ex.: Coordinating tool development to reduce duplication of effort

1. https://openproblems.bio/events/2022-08_neurips/
2. https://www.kaggle.com/competitions/open-problems-multimodal/

6. How can we establish benchmarking workgroups to assess the performance and compare the accuracy and reproducibility of different computational methods and algorithms?

7. What criteria should be considered when defining best practices in experimental design, sample preparation, data acquisition, and data analysis?

8. What strategies can be implemented to standardize data formats and metadata annotations in order to enhance data sharing, integration, and reproducibility?

9. Are there any existing initiatives or resources that can be leveraged to create comprehensive guides or tutorials for researchers new to the field? Ex.: https://www.sc-best-practices.org/preamble.html

10. What initiatives can be undertaken to enhance the training and education of researchers interested in single-cell and spatial transcriptomic?

11. Are there any specific challenges faced by researchers in resource-limited settings, and how can we address these challenges to promote inclusivity in the field? Ex.: Cheaper experiments

References:

Heumos, L., Schaar, A.C., Lance, C. et al. Best practices for single-cell analysis across modalities. Nat Rev Genet (2023). https://doi.org/10.1038/s41576-023-00586-w

Baysoy, A., Bai, Z., Satija, R. et al. The technological landscape and applications of single-cell multi-omics. Nat Rev Mol Cell Biol (2023). https://doi.org/10.1038/s41580-023-00615-w

Mariana Boroni
Principal Investigator
INCA - National Cancer Institute ...
Moderator

Nir Yosef
Principal Investigator
University of California Berkeley
Moderator

2:45 PM - 4:00 PM EDT (1 Hour, 15 Min)

(Breakout Session 1) Topic #2: Strengthening integration pipelines

Main Hall

WATCH RECORDING

This session will focus on integration pipelines that aim to convert raw data collected by 17+ consortia into a multiscale atlas of the human body. We will start with an overview of clinical and basic research use cases that drive 3D atlas construction and usage and a Q&A of desirable features for a multiscale human atlas--including coverage, quality, stability, extendibility, utility, licensing. Next, there will be brief presentations of data and code integration pipelines developed by different teams that use the evolving atlas to prioritize tissue acquisition, improve data analysis, expand existing ontologies to cover healthy humans, and develop user interfaces that support exploration and usage of multiscale atlas data. Questions might include:

- For which cells do we need which additional molecular modalities to add critical information that is not already encoded in gene expression?
- What data structures ensure the atlas is robust and expandable?
- How can modeling of disease-associated cells and their niches drive atlas construction?
- What computational methods can be used to identify undersampled regions of phenotype space?
- How can we synergize and learn from other resources (incl. HuBMAP, Human Protein Atlas, Open Targets) to build interoperable resources?

We conclude with a discussion of major challenges and opportunities to speed up data and code integration in support of human atlas construction and usage.

Emma Dann
PhD Student
Wellcome Sanger Institute
Moderator

Katy Borner
Principal Investigator
IU
Moderator

2:45 PM - 4:00 PM EDT (1 Hour, 15 Min)

(Breakout Session 1) Topic #3: Ancestral diversity

CR2

WATCH RECORDING

Topic description

Humans are diverse, and this diversity influences the properties of our cells, as well as our developmental processes, and disease mechanisms, diagnoses, and treatments. An atlas of human cells must therefore account for all forms of diversity, including variation in genetic ancestry, sex, age, geography, environment, and lifestyle. In addition to the clear scientific and medical needs, this is also essential to the equity goals and principles of the HCA. The benefits of biomedical and genomics research must be broadly shared, rather than restricted to specific population groups for which research is well-funded, for example individuals of European ancestry living in the West.

The mission of the HCA Genetic Diversity Network (GDN) is to investigate the impact of genes and environment (i.e., non-genetic influences) on molecular traits in cells from diverse individuals across the globe. Specifically, we seek to understand how key factors such as genetics, ancestry, sex, age, geography, and lifestyle influence the properties of our cells and tissues.

We are now ready to start work on v1.0 of the HCA Human Diversity Atlas. The overall strategy has been discussed in HCA Genetic Diversity Network Roadmap meetings and summarized here in our GDN roadmap: https://docs.google.com/document/d/1jVKAQhu4icv_zcQSSjjHeNPntyLlSFfJtW7Uqdcprrs. In this Breakout, we will discuss concrete steps to put this strategy into practice.

Example questions for the breakout session:

Which datasets are ready for inclusion in v1.0 of the Human Diversity Atlas?

What donor, sample, protocol and data quality descriptors can we include (donor, sample, and technical metadata)?

How should we describe genetic ancestry, ethnicity, race and other forms of identity? Which genotype profiling assay could we use? Self-reported vs data-derived descriptors? Discrete categories vs continuous coordinates?

Methodology, data analytics, QTLs: how to associate genetic variants and donor/sample metadata with single cell phenotypes?

Methodology, data analytics: how to address batch effects/lab-specific technical variation and biases in integrating across datasets (unique challenges faced by GDN)?

How to measure the degree of diversity in HCA datasets?

How can we leverage insights from past population genetics studies (based on DNA polymorphisms)?

Partha Majumder
Director
J.C. Martin Centre for Liver Resea...
Moderator

Shyam Prabhakar
Senior Group Leader
Genome Institute of Singapore
Moderator

Sophia George
Principal Investigator
University of Miami, Sylvester Co...
Moderator

4:00 PM - 4:45 PM EDT (45 Min)

Refreshment Break

4:15 PM - 6:00 PM EDT (1 Hour, 45 Min)

Day 1 - Part 3

WATCH RECORDING

4:45 PM - 5:25 PM EDT (40 Min)

Session 5: Lightning Talks

Main Hall

1. **Kian Hong Kock:** The Asian Immune Diversity Atlas (AIDA): Determinants of diversity in circulating immune cell states across Asia

2. **Collins Morang'a:** scRNA-Seq reveals elevated interferon responses and TNF-α signaling via NFkB in monocytes in children with clinical malaria

3. **Daniela Senra:** Stem cell determination from scRNA-seq data using a protein network-based approach

Arathi Mani
Engineering Manager
Chan Zuckerberg Initiative
Moderator

Kian Hong Kock
Research Fellow
Agency For Science, Technology a...
Speaker

Collins Morang'a
Postdoctoral Fellow
WACCBIP, University of Ghana
Speaker

Daniela Senra
Graduate Student
Universidad Nacional de La Plata
Speaker

5:25 PM - 5:30 PM EDT (5 Min)

Day 1 Closing Remarks

Main Hall

John Randell
Chief Alliance Officer
Human Cell Atlas
Speaker

5:30 PM - 7:00 PM EDT (1 Hour, 30 Min)

Reception & Poster Session

Poster Presenters

Day 2 - **11 July 2024**

8:30 AM - 11:30 AM EDT (3 Hours)

Day 2 - Part 1

WATCH RECORDING

9:00 AM - 9:05 AM EDT (5 Min)

Welcome Day 2

Main Hall

Muzlifah Haniffa
Principal Investigator
Wellcome Sanger Institute
Speaker

9:05 AM - 9:50 AM EDT (45 Min)

Session 6: Spatial methods

Main Hall

1. **Jian Ma:** Integrative single-cell spatial modeling of cell identity
2. **Fei Chen:** A scalable platform for single-nucleus spatial genomics
3. **Nadav Yayon:** A spatial thymus human cell atlas projected to a unidimensional tissue axis

Jian Ma
Principal Investigator
Carnegie Mellon University
Speaker

Fei Chen
Principal Invesgistaor
Broad Institute
Speaker

Nadav Yayon
Postdoctoral fellow
European Bioinformatics Institut...
Speaker

9:50 AM - 10:35 AM EDT (45 Min)

Session 7: New dimensions to enable new questions

Main Hall

1. **Christina Leslie:** Machine learning for single-cell regulatory genomics
2. **Paolo Cadinu:** Colonic Tissue Remodeling In A Mouse Model Of Inflammatory Bowel Disorder
3. **Caleb Lareau:** Elucidating clonal mosaicism in human lymphocytes

Kerstin Meyer
Wellcome Sanger Institute
Sanger Institute
Moderator

Caleb Lareau
Instructor
Stanford University
Speaker

Christina Leslie
Principal Investigator
Memorial Sloan Kettering Cancer ...
Speaker

Paolo Cadinu
Postdoctoral Research Fellow
Children's Boston Hospital and H...
Speaker

10:35 AM - 11:20 AM EDT (45 Min)

Refreshment Break

11:19 AM - 12:35 PM EDT (1 Hour, 16 Min)

Session 8: Breakout Session #2: New Technologies

Topic #1: Foundation models

Topic #2: Spatial methods (experimental)

Topic #3: New dimensions to enable new questions

11:20 AM - 12:35 PM EDT (1 Hour, 15 Min)

(Breakout Session 2) Topic #1: Machine Learning: Foundation models

CR2

WATCH RECORDING

Fabian Theis
Director
Helmholtz Zentrum München
Moderator

Bo Wang
Faculty Member
Vector
Moderator

11:20 AM - 12:35 PM EDT (1 Hour, 15 Min)

(Breakout Session 2) Topic #2: Spatial methods (experimental)

CR3

WATCH RECORDING

Session summary:

Spatial technologies allow us to restore structural information lost in single-cell techniques and gain insight into collective function. To study spatial patterns of cellular phenotypes, various spatial technologies with multi-modalities (genome, epigenome, chromatin accessibility, transcriptome, translatome, proteome,

metabolome, cellular function, etc) have been developed. These technologies involve trade-offs among several factors, including high detection efficiency, transcriptome-wide profiling, high resolution, histological imaging, and the large size of the tissue area that can be captured. This breakout session will explore ways to better select the most appropriate technology for various biological questions.

Example questions:

- What are example biological questions/mechanisms we could reveal using spatial technologies? Intercellular interactions, spatial trajectory/axis, subcellular molecular kinetics, etc.
- For spatial transcriptomics, what plotform is available and what are the pros and cons of each technology regarding their sensitivity, resolution, scalability, cost? (array- or image-based)
- Emerging modalities? (chromatin accessibility, histone modifications, TCR/BCR, slide-DNA-seq, mRNA translation, mRNA transcription & degradation, protein, metabolites, cellular function, etc)
- What are the major computational data analysis needs for spatial technologies? For example, how can we utilise/integrate suspension data with spatial technologies, how can we integrate spatial data from diffferent spatial technologies
- How many biological replicates and how many cells are needed for disease studies when using spatial technogies? How to conduct comparative analysis to dissect disease mechanisms?
- What are the benefits of multi-modality or cross-modality analysis? For example, having histologigal imaging data and other imaging modalitie for annotation of histological structure and cell segmentation? Cross-modality analysis from molecular features (spatial technologies) to functional features (metabolism, cell mechanics, electrophysiology, MRI, etc.)
- How to achieve direct 3D measurements or 3D reconstruction for thick-tissue blocks or whole-organ, and what is the advantage? Can we achieve common coordinate framework measured by spatial technologies for human organs?
- Any other tasks to tackle in the future?

Kazumasa Kanemaru
Postdoctoral Associate
Sanger Institute
Moderator

Xiao Wang
Principal Investigator
Broad Institute & MIT
Moderator

11:20 AM - 12:35 PM EDT (1 Hour, 15 Min)

(Breakout Session 2) Topic #3: New dimensions to enable new questions

Main Hall

WATCH RECORDING

Topic description

Single-cell atlases promise to provide a 'missing link' between genes, diseases, cells and therapies. By identifying the specific cell types, states, programs and contexts where disease-implicated genes act, we will understand the mechanisms of disease at the cellular and tissue levels and can use this understanding to develop powerful disease diagnostics; identify promising new drug targets; predict their efficacy, toxicity and resistance mechanisms; and empower new kinds of therapies, from cancer therapies to regenerative medicine. What new dimensions of data, computation and thinking must we develop to build a healthy reference human cell atlas that enables this vision? And what types of important questions will these new dimensions enable us to answer?

Example questions for the breakout session:

What will the ultimate Human Cell Atlas look like? Will we need a 3D virtual model of the entire human body that can be explored in virtual reality that we can zoom in from the whole body to organs, tissue structures, cells and molecules?

Will we need computational models that can simulate biologically and clinically relevant human body systems?

How will we integrate information across dimensions and scales of the body and link molecular biology and physiology?

What important types of questions will such a multiscale human model be useful to answer?

What new technologies will be needed to answer these questions?

What technologies are available now that help us work on such questions and what are the gaps where new methods are needed?

Aviv Regev
Principal Investigator
Genentech
Moderator

Dana Pe'er
Principal Investigator
Sloan Kettering Institute
Moderator

John Marioni
Group Leader
EMBL-EBI & University of Cambri...
Moderator

12:30 PM - 2:00 PM EDT (1 Hour, 30 Min)

Lunch

1:30 PM - 4:43 PM EDT (3 Hours, 13 Min)

Day 2 - Part 2

WATCH RECORDING

2:00 PM - 2:50 PM EDT (50 Min)

Session 9: Keynote #2

Main Hall

Stem cell-derived human embryo models- are they close to becoming functional embryos?

Philip Awadalla
Director
Ontario Institute for Cancer Rese...
Moderator

Janet Rossant
Chief of Research Emeritus
Sick Kids Research Institute
Keynote

2:50 PM - 3:20 PM EDT (30 Min)

Session 10: Lightning Talks

Main Hall

1. **Alex Tong:** Perspective Learning Continuous Dynamics from Time-lapsed Single-cell Data
2. **Romain Lopez:** Towards Causal Modeling of Single-cell Perturbation Data
3. **Charlotte Bunne:** Neural Optimal Transport for Single-Cell Biology

Sergio Triana
Postdoctoral Associate
Harvard University
Moderator

Charlotte Bunne
PhD Student
ETH Zurich & Broad Institute
Speaker

Alexander Tong
Postdoctoral Fellow
Mila & University of Montreal
Speaker

Romain Lopez
Postdoctoral Scholar
Stanford University
Speaker

3:20 PM - 3:35 PM EDT (15 Min)

Refreshment Break

3:35 PM - 4:25 PM EDT (50 Min)

Session 11: Frontiers in Deep Learning and the Human Cell Atlas

Main Hall

1. **Bo Wang:** scGPT: towards building a foundation model for single-cell multi-omics using generative AI

2. **Danilo Bzdok:** Leveraging machine-learning paradigms towards single-subject prediction

3. **Maria Brbic:** Deep learning for cell type discovery

Jinmiao Chen
Principal Investigator
Singapore Immunology Network ...
Moderator

Maria Brbic
Principal Investigator
Swiss Federal Institute of Techno...
Speaker

Bo Wang
Faculty Member
Vector
Speaker

Danilo Bzdok
Professor
Mila and McGill University
Speaker

4:29 PM - 5:45 PM EDT (1 Hour, 16 Min)

Session 12: Breakout Session #3: Building the whole human

Topic #1: Spatial methods (computational)

Topic #2: Deep learning and machine learning for single cell mapping

Topic #3: Cross tissue integration/analysis

4:30 PM - 5:45 PM EDT (1 Hour, 15 Min)

(Breakout Session 3) Topic #1: Spatial methods (computational)

CR2

Topic Description

Over the past five years, spatial omics methods have become an integral part of our toolbox for gaining a better understanding of tissue biology. Advances in experimental spatial omics techniques have increased their accessibility, scalability, and robustness. Consequently, we are now able to generate a greater number of large-scale, high-quality datasets of spatial omics data, enabling us to explore a more diverse range of questions. The surge in data availability has highlighted the need for computational methods specifically tailored for spatial omics data. Therefore, the main objective of the session will be to delineate the computational problem space associated with spatial omics data and subsequently identify the most urgent and impactful challenges in the field, particularly within the HCA community.

Session Outline

The flowchart below by and large summarizes our session, where blue indicates work that I and David will do before the session starts and red will constitute the session. That is, we'll try to define the current problem space for spatial omics methods in the context of HCA and then ask people to discuss the following: i) whether there are existing methods addressing these problems that are "ready to use", or ii) whether new and/or improved methods need to be developed. We hope to be able to reach out to the participants before asking them to contribute with ideas/questions/problems they consider important to discuss. We will ignite discussion with the following topic areas and will extend this list during the session:

Discussion A: Experiment design.

- Probe design and selection. How does fine-grained cell state discovery from sc/nRNA-seq relate to deployment of probe-based methods in individual bionetworks? What aspects of probe design are limiting at the level of granularity currently relevant for spatial investigation?
- Power analysis. Which statistics on spatial data pose issues in terms of data set size or power analysis methodology (neighborhood, sample metadata, ...)?

Discussion B: Deriving insights.

- Transferring information from the sc/snRNA-seq atlas. Is cell type deconvolution used, do people observe problems with these methods? Are single cell to spatial location mapping tools used, problems? Are notions of CCFs used to contextualize spatial information when mapping?
- Integration and batch correction. Experiences with integration or correction in the presence of biologically relevant sample covariates, like disease state of a tissue? Do tools developed for scRNA-seq work well for *FISH or spot transcriptomics tools?
- Cell-cell interactions: Are there aspects of cell-cell communication beyond neighborhood analyses that require further attention? What questions on cell-cell interactions remain after initial analyses on sc/nRNA-seq data to be addressed by spatial data?
- Cell niche composition: Are statistical and visualization techniques sufficient? Can hypotheses that relate to differential composition with respect to sample metadata be sufficiently addressed?

Deliverables

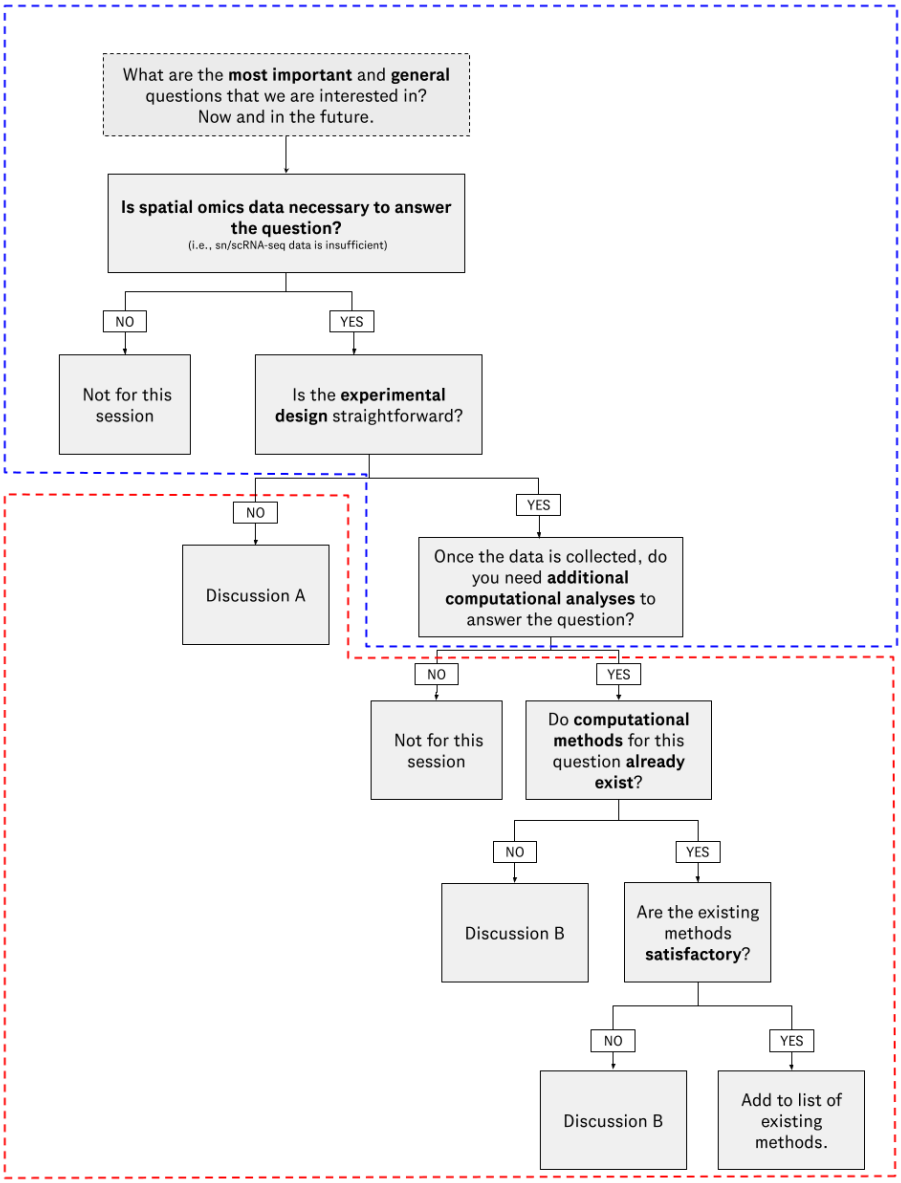
For the discussion we were thinking of smaller working groups that could brainstorm together, mainly to make people more comfortable in speaking up. Each group would then present their work/conclusions to the rest of the group where others can give feedback. As deliverables we currently see the following two items:

1. A map of the current and anticipated problem space. For example, a table listing:

-
- Whether the problem requires more advanced computational methods
- Whether existing methods exist or if new methods should be developed.
- If methods exist, whether they are satisfactory or if there's a need to expand on them.
- Stakeholder for each questions (e.g., bionetworks)
- Priority - how urgent is it that we address this problem

Question	Requires SO	Requires Computational Analysis	Need for new methods	If exists, satisfactory?	Stakeholders	Priority
⋮	⋮	⋮	⋮	⋮	⋮	⋮

2. Following 1), suggestions of the most relevant areas of focus for computational methods for the near and long term. That is, what do we think are the most important tasks to focus on right now (e.g., bottlenecks etc.), and what do we expect to become important in the future.



Alma Andersson
Sr. AI Scientist
Genentech
Moderator

David Fischer
Postdoctoral Associate
Broad Institute
Moderator

4:30 PM - 5:45 PM EDT (1 Hour, 15 Min)

(Breakout Session 3) Topic #2: Deep learning and machine learning for single cell mapping

CR3

WATCH RECORDING

Topic Description:

Machine learning and deep learning algorithms have found various applications in analysis of molecular data and particularly single cell and spatially resolved transcriptomics. These applications range from cell type annotation, dimensionality reduction, visualization, data simulation, gene regulatory network reconstruction, etc. However, there are still many challenges that need to be addressed in robustly and accurately applying these methods. This breakout session will explore these challenges and ways to overcome them in the context of cell atlases.

List of questions:

- Label quality: Many DL algorithms for single cell mapping require labelled data, but annotations may be inconsistent/wrong/conflicting. What role can ML/DL play in label correction and consensus?
- Measuring performance of unsupervised analysis: Visualization and clustering are often performed via unsupervised ML algorithms, but benchmarking these is difficult given the lack of ground truth. What new tools and/or data are required to benchmark atlas level dimensionality reduction methods?
- Annotation granularity: The granularity of cell type annotation is variable, from cell types to cell states. What is the role of ML/DL in helping set the annotation granularity of single cell atlases?
- Overfitting: Deep learning cell type mapping may be prone to overfit, which is of special concern across technologies/batches. Are new tools/approaches necessary to minimize annotation overfitting?
- Domain shift in annotation: Batch effects and domain discrepancies result in differences between the labeled reference data and target unlabeled data (a probably particularly important in cell type annotation of spatial transcriptomics using scRNAseq references). How can we address these issues using more robust DL/ML models?
- Simulation: Several simulators of scRNAseq data have been recently developed using ML/DL methods that enable data augmentation, imposing GRNs in generated data, gene knockout prediction, etc. What role can realistic simulators play in HCA and what shortcomings exist that ML/DL methods can overcome?

Kieran Campbell
Investigator
Lunenfeld-Tanenbaum Research ...
Moderator

Amin Emad
Principal Investigator
McGill University
Moderator

4:30 PM - 5:45 PM EDT (1 Hour, 15 Min)

(Breakout Session 3) Topic #3: Cross tissue integration/analysis

Main Hall

WATCH RECORDING

Topic description

Throughout the initial phase of the Human Cell Atlas, most of the studies have been limited to constructing single-cell atlases for individual organs and individual donor's cohort. As we enter a new phase of the HCA, different communities and HCA bionetworks will need to work together to build cross-tissue atlases to define cellular compartments shared across tissues, characterize the molecular features of broad cell types, and uncover rare and poorly characterize cell subsets within specific human tissues. Such cross tissue analysis will enable revealing tissue-shared and tissue-specific cell features and provide unprecedented insights into cell types and states underlying health and disease pathogenesis.

Cross tissue integration analysis is particularly critical for (i) distributed cellular networks (e.g. immune system, vascular system) where cell subsets and programs are present across multiple tissue (e.g., immune system); (ii) establishing how cell subsets and programs change overtime within an organ and across organ system (e.g. development atlas, immune system); (iii) characterizing how the impact of genetic and environment factors on cellular characteristics from ethnically diverse cohort shapes cellular programs defining health and disease states across organ systems (e.g., genetic diversity bionetwork).

During this session, we will discuss the challenges related to building cross tissue atlases to identify potential solutions to empower our community in this process. In addition, we will discuss the data types that will be required and the progress so far to prioritize data integration for cross tissue atlas generation.

Example questions for the breakout session:

Big picture question to frame the discussion:

- What does an integrated atlas look like to you within an organ system and across organ systems? What data modalities should it include and what are priority data and areas?

- What are other case scenarios where a cross tissue analysis is needed? E.g. other distributed networks, rapid temporal changes (e.g., pediatric atlases).

Challenges for cross tissue integration:

- How do you decide which computational methods and benchmark to use for the integration? What types of modalities and datasets should we prioritize first?
- How do you decide on which computational parameters to use and optimize?
- What is ground truth? How do you benchmark the ground truth?
- How can we identify tissue-specific cell states when sample/library preparation protocols vary across tissues, i.e., when tissue type and assay are confounded?

Validation: What technologies and readouts will be needed to validate prediction of shared and distinct cellular programs across tissues?

How to use the atlas: What types of queries would you like to be able to make in an integrated atlas? How can the integrated atlas be used by the community?

Alexandra-Chloe Villani
Principal Investigator
Broad Institute, Massachusetts G...
Moderator

Shyam Prabhakar
Senior Group Leader
Genome Institute of Singapore
Moderator

Kerstin Meyer
Wellcome Sanger Institute
Sanger Institute
Moderator

5:50 PM - 7:00 PM EDT (1 Hour, 10 Min)

Reception & Poster Session

Day 3 - 12 July 2023

8:00 AM - 11:30 AM EDT (3 Hours, 30 Min)

Day 3 - Part 1

WATCH RECORDING

9:00 AM - 9:05 AM EDT (5 Min)

Welcome Day 3

Main Hall

Sophia George
Principal Investigator
University of Miami, Sylvester Comprehensive Cancer Center
Speaker

9:05 AM - 9:50 AM EDT (45 Min)

Session 13: Keynote #3

Main Hall

Making the most of large cohorts: tools, thoughts and applications

Nir Yosef
Principal Investigator
University of California Berkeley
Keynote

Sophia George
Principal Investigator
University of Miami, Sylvester Co...
Moderator

9:50 AM - 10:50 AM EDT (1 Hour)

Session 14: How should we think about cell function?

Main Hall

1. **Katy Borner & David Osumi-Sutherland:** Multiscale Human Reference Atlas: Functional Tissue Units

2. **Gray Camp:** Towards Integrative Human Organoid Atlases

3. **Brian Brown:** Investigating cancer cell control of their local cell environment

4. **Jay Shin:** Atlas of Transcribed Cis-Regulatory Elements Reveals Disease-Associated Regulatory Modules in Distinct Cell Populations

Kristin Ardlie
Principal Investigator
Broad Institute
Moderator

Katy Borner
Principal Investigator
IU
Speaker

David Osumi-Sutherland
Group Leader
EMBL/EBI
Speaker

Gray Camp
Principal Investigator, Group Lea...
Institute of Molecular & Clinical O...
Speaker

Brian Brown
Principal Investigator
Icahn School of Medicine
Speaker

Jay Shin
Group Leader
GIS A*STAR
Speaker

10:50 AM - 11:15 AM EDT (25 Min)

Refreshment Break

11:14 AM - 12:16 PM EDT (1 Hour, 2 Min)

Session 15: Breakout Session #4: The Path Ahead

Topic #1: How should we think about, discover and validate cell function?

Topic #2: Emerging Ethical Issues

Topic #3: How will the atlas be used? How do we enable maximum scientific use and impact?

11:15 AM - 12:15 PM EDT (1 Hour)

(Breakout Session 4) Topic #1: How should we think about, discover and validate cell function?

CR3

WATCH RECORDING

Topic Description:

Cell replication, differentiation, migration, and tissue morphogenesis are influenced by classical molecular mechanisms, such as regulatory networks (e.g., immune-parenchymal cell interactions, gene-gene interactions). These mechanisms involve various critical molecular processes at the cellular level, including co-expressed gene modules, epigenomic regulation, cis- and trans-gene regulation, signal transduction, and cell-cell communication. These processes collectively contribute to the development of structural, biochemical, and physiological functionality in cells and tissues. The purpose of this session is to explore integrative approaches that can improve our understanding and validation of cell function.

II. Example Questions for the Breakout Session:

1. What exemplary processes can serve as models for unraveling the regulatory mechanisms underlying cell function? For example, how does stage-specific stem cell maturation contribute to alternative differentiation potential?

2. How do we validate and explore cell-cell interactions, and integrate these into our understanding of complex organ functions? What technologies (e.g., spatial) and models (e.g., co-culture, organoids) are needed to move beyond cell-autonomous function?

3. Broad question: What do we mean by "function"? (evolved, biochemical activity, other) Degree of function vs type of function? Normal vs altered function?

4. How can we enhance our understanding of cell function by integrating diverse molecular mechanisms and processes? What is there to be learned by comparing across processes?

5. What key regulatory mechanisms play a role in cell development, and how can we characterize them? What modalities (beside snRNA-seq, ATAC-seq) might we need to prioritize?

6. How do we think about and validate local versus distal immune regulatory mechanisms?- E.g., resident vs circulating immune cell interactions.

7. What technological advancements are necessary to accurately measure and assess the regulatory mechanisms associated with cell function? Should we focus on developing new technologies for this purpose?

8. How can computational methods and tools aid in mapping and analyzing regulatory mechanisms to gain insights into cell function? What validation approaches are necessary to ensure the reliability of these computational findings?

9. What other organized efforts besides the HCA are providing complementary data, analyses, models?

III. Conclusion:

The goal of this breakout session is to foster discussions and explore innovative approaches to understand, discover, and validate cell function. By addressing the questions provided, we can collectively advance our knowledge and develop strategies to enhance the integration efforts of the Human Cell Atlas. Furthermore, expanding the scope of the Human Cell Atlas to incorporate spatial and multi-omics data will contribute to our understanding of cell function. Identifying the most critical needs for developing new computational approaches will drive advancements in the field and align with the overall objectives of the annual meeting.

Sarah Crome
Principal Investigator
University of Toronto & Universit...
Moderator

Jesse Gillis
Professor
University of Toronto
Moderator

11:15 AM - 12:15 PM EDT (1 Hour)

(Breakout Session 4) Topic #2: Emerging Ethical Issues

CR2

WATCH RECORDING

Summary:

Building data governance frameworks, ethics policies, consent templates and other guidance has hitherto been the hallmark of the work of the Ethics Working Group but international discussions such as those surrounding the GDPR , the European Data Governance Act , WHO's Pandemic Treaty (currently under development) as well as the decisions on the Nagoya Protocol (plant/ animal digital sequence information) and their possible spillover effect onto human data sharing merit open discussion

Questions to be addressed:

As we turn to the exciting phase where HCA research results will be shared in both open and controlled access databases, what new challenges accompany this next phase? What are the ethico-social and legal needs (and wishes?) of HCA contributors? How to foster and ensure equitable participation and access? Are there emerging interpretations of data protection norms or of international and national legal norms governing biomedical research that merit discussion and input?

Emily Kirby
Academic Associate
McGill University
Moderator

Bartha Maria Knoppers
Principal Investigator
McGill University
Moderator

11:15 AM - 12:15 PM EDT (1 Hour)

(Breakout Session 4) Topic #3: How will the atlas be used?

Main Hall

WATCH RECORDING

A reference map of what healthy human tissues look like is the final aim of the HCA consortium. Have we gotten closer to that? Several biological networks have put together map drafts representing a global and integrated view of their group efforts. This breakout session aims to contribute to defining the next phase of the Human Cell Atlas and to identify what capabilities are currently lacking to answer the next generation of questions.

During this breakout session we will discuss:

1. Views on a representative map of human cells: are we close to a reference for

healthy human tissues?

2. The experience of the community with resources and results: questions, tools, and data.

3. Integrative data analysis across the different biological networks: what are the bottlenecks?

4. What new technologies and modalities our community should we thinking about in order to build Atlas 1.0 and beyond.

Graham Heimberg
Principal Scientist
Genentech
Moderator

Patricia Severino
Principal Investigator Principal In...
Hospital Israelita Albert Einstein
Moderator

12:15 PM - 12:30 PM EDT (15 Min)

Refreshment Break

12:29 PM - 2:00 PM EDT (1 Hour, 31 Min)

Day 3 - Part 2

WATCH RECORDING

12:30 PM - 1:00 PM EDT (30 Min)

Session 16: Closing Panel Discussion: What comes next?

Main Hall

Alexandra-Chloe Villani
Principal Investigator
Broad Institute, Massachusetts G...
Moderator

Mariana Boroni
Principal Investigator
INCA - National Cancer Institute ...
Panelist

Maria Brbic
Principal Investigator
Swiss Federal Institute of Techno...
Panelist

Fei Chen
Principal Invesgistaor
Broad Institute
Panelist

Aviv Regev
Principal Investigator
Genentech
Panelist

Jay Shin
Group Leader
GIS A*STAR
Panelist

Sarah Teichmann
Head of Cellular Genetics and Se...
Sanger Institute
Panelist

1:00 PM - 1:10 PM EDT (10 Min)

Closing Remarks

Main Hall

Gary Bader
Principal Investigator
University of Toronto
Speaker

Aviv Regev
Principal Investigator
Genentech
Speaker

Sarah Teichmann
Head of Cellular Genetics and Se...
Sanger Institute
Speaker