Adipose

Status

The Adipose Network is still at a very early stage. We have had several discussions with selected colleagues, but we have not yet established a true working network.

Adipose tissue provides technical challenges that are somewhat unique, related to the huge amount of lipid stored in adipocytes. The large lipid droplets make adipocytes fragile and incompatible with most single cell approaches including spatial transcriptomics. However, several groups are beginning to have success applying single nucleus methods to characterize isolated adipocytes as well as whole adipose tissue from both mice and humans.

Strategy moving forward

At the meeting we agreed that postdocs from the Rosen and Mandrup groups would take the lead on establishing a network of junior researchers to drive the initial comparison and integration of data across labs, methods, depots and physiological conditions for both humans and mice. This effort will begin in the next few months. A first milestone will be to define and publish a common nomenclature based on data integration. In addition, we will establish a home page to stimulate network identity and disseminate our work. Furthermore, we plan to organize informal workshops that will be open to all. Finally, in order to generate comprehensive reference maps of 'all' human adipose tissues in healthy and diseased patients, we need to recruit additional funding and personnel. How to do that will be the topic of future discussions.

Eye

Update of the current status.

Single cell or single nuclei transcriptome profile has been conducted for all parts of the eye, including over 2 million cells from the retina, 57K from the Iris, 51K from corneoscleral wedge, 34K from ciliary body, 37K from cornea, and 4K from lens. These data will form the basis to build cell atlas v1.0 for the human eye.

Challenges

Several challenges have been identified, including

1. Compilation and Integration of datasets generated from multiple groups with different technologies.

2. Proper annotation of the large number of cell types in the tissue.
3. How to distinguish cell type vs cell state.
4. Mechanism for making processed reference datasets easily accessible for researchers who are not bioinformaticians.

Needs

We identify there are needs to:

1. Datasets to cover variation in human population, such as age groups, gender, ethnic groups.
2. More thorough profiling of regions of the eye in addition to the retina.
3. Generation of single cell spatial transcriptomics data and epigenomics data.

Future plan

Our plan for the next 12 months is to:

1. Complete the eye cell atlas v1.0 and make it available to the community.
2. Grow the eye network and organize regular meetings and other activities.
3. Form working groups with focused tasks.
4. Organize community annotation effort for the eye cell atlas v1.0.
5. Extend cell atlas by including development single cell dataset.

Genetic Diversity

Points that did not translate into concrete action items:

- We need to demonstrate the scientific/biomedical value of generating and analyzing data from diverse cohorts. Blood-based projects have the greatest diversity - how can we use them to show value?
-
- Important to address ethics, logistics, local resource limitations and health disparities that may exist across groups in one single location.
- Technology keeps moving. For any GDN project, it's important to decide which technology to use, whether we should wait for it to get cheaper or go now, how deeply to sequence, how many cells to sequence, etc. (experimental design).
- What is the role of statistical power analysis in designing the study?
- Vendors have an interest in population-scale single cell – engage them early to see how they can contribute

- In any large-scale study, there is a tradeoff between "home-brewed" data generation technologies that could be substantially cheaper vs off-the-shelf commercial technologies that are likely to be more reliable and less technically variable.
- What data/summary statistics can GDN researchers share without compromising privacy?
- How to coordinate genetic diversity efforts across HCA, HuBMAP, ICDA, sc eQTLgen, other programs and consortia?

Action items (tentative):

- Draft a white paper on the mission and scope of the HCA Genetic Diversity Network (GDN)
- Quantify diversity in our datasets, identify gaps that need to be addressed (this aim is shared with the Equity Working Group)
- Define GDN data/metadata standards for sc eQTLs and other relevant data types
- For ease of comparison across studies: standardize the cell types we will analyze across studies of any one sample type (initially: blood).
- Work out the logistics of generating single cell data from a geographically distributed healthy cohort.
- Then reach out to ICDA (for example) to apply the same methods to case-control studies on geographically diverse cohorts.
- Initiate regular GDN meetings, perhaps every second month, to showcase science, share techniques, seed new collaborations. For coherence of purpose, the meetings will have an HCA focus. But can include interested non-HCA researchers and administrators (HuBMAP, ICDA, ...).

Heart

Here we are at the Heart of the HCA!!

Our BioNetwork is the most ‘up-beat’ of all facing the challenge to study the human heart which beats over 100,000 times each day to continuously pump about 7 litres of blood through the intricate and complex system of blood vessels. The four chambers of the heart withstand remarkably different pressures, reflected in their different architecture, and sustained by cells need that work synchronously to enable the coordinated cardiac pump function. From

the heart blood in pumped directly in the large arteries and subsequently into medium, small e very tiny calibre capillaries that with their ramifications reach each organ, tissue, and cell of the body. The return trip of the blood to the heart via the venous system completes the complex and continues work of the cardiovascular system.

The complexity of the cardiovascular system is reflected by an exponentially growing network of international contributors with 189 participants from all 5 continents. So far data from 38 individuals 259 samples resulting in 844441 cells/nuclei have been recorded within the HCA. Multiple studies have been published in the last 18months with a focus on creating a healthy cardiovascular reference atlas. The reference data sets have been proved crucial in mapping the transcripts encoding for SARS-CoV-2 entry receptors. Further, exploratory studies on human disease are emerging with the healthy reference atlases proving essential as control cohorts.

To enable thorough comparisons such as male versus female hearts and age dependent analysis, there is a need to increase the number of heathy donor samples collected and analysed. In this perspective access and record of metadata is crucial and an individual effort from each team is required. Indeed, in view of the dynamic cardiovascular network and the rich datasets being produced, to enable optimal integration of data sets from diverse sources, it is crucial to work closely with clinicians to access and collect comprehensive donors and patients' metadata. Further, inclusion of histopathology reports as well as macroscopic and more importantly microscopic images (H&E staining) of the tissue profiled has been reiterated. Indeed, synergy with pathologists is proposed as an essential step to provide confirmation of the tissue sub-anatomical components profiled, correlate to any histopathological finding and, importantly, to ensure continuity for future implementation of single cell and spatial 'omics technologies in clinical diagnostic settings.

Most of the current studies covered multiple anatomical regions, including left and right ventricles, left and right atria, interventricular septum, and apex. The next phase will aim to profile more specialized areas of the heart and expand the type of blood vessels studied. The availability of whole donor hearts to select regions of interests is restricted and in the case of diseased samples there are limitations determined by the need for tissue by pathologists on one side, even more by the accessibility of certain areas during surgery to ensure patients

safety. Technological advances such as spatial transcriptomic on FFPE samples will allow to study retrospective samples providing a broader source of tissue.

The heart undergoes substantial morphologic and hemodynamic changes at birth and throughout life until adulthood. Thus, to study congenital cardiac disease (a leading cause of death in children which affects nearly 1% of births) and provide adequate references for each step of post-natal development, the network recognises the urgency of building a longitudinal heart atlas from infant, childhood, and adolescent tissues. Notably, a gap also needs to be filled for young adults (20-40 years old) as the existing healthy hearts datasets are generally obtained from donors > 40years old.

The group has discussed the need to broaden the ancestral diversity of healthy tissue studied especially in view of the increased cardiovascular risk of specific ethnicities underrepresented in the datasets. In this regard, there is a clear need to intensify community engagement to promote participation of underserved groups in Western countries as well as developing ones. Notably, controlled data access is the crux for some groups of patients and donors who would agree to donate their tissues/organs for research, but they don't feel comfortable with open data access. Flexibility in this regard has the potential to expand the number of samples available.

Outreach activities should also involve tissue procurement agencies and clinicians across the globe. Funding agencies should be aware of the need to invest in projects enabling the establishment of mutually beneficial international research agreements with underrepresented countries for inclusiveness and global improvement of health. To achieve a truly global involvement, there is a need to initiate conversations with governments and policy markers in countries where fear of exploitation led to the establishment of strict laws aimed to protect genetic and tissue resources, ultimately slowing international collaborations.

The establishment of a first reference heart atlas provide the platform for the phase studying subtleties of disease. For many cardiovascular diseases there are limitations in accessing multiple samples from the same patients to follow the disease from early to late stages. To overcome these limitations a close collaboration with clinicians will be required to design clinical studies targeting the most urgent clinically relevant questions.

Model systems are required for functional validation and need to be selected based appropriately with larg(er) animal models being required beyond mouse models for specific pathologies such as cardiac rhythm diseases. The mouse has a heart frequency much lower than human and is not a good model for these types of pathologies but is a valid system in other instances such as ischemic heart disease. The analysis of pig hearts has been proposed as relevant preclinical validation is done in pig models before clinical trials can commence but limitations of these model remain where genetic manipulation and costs can be prohibitive. Functional validation of human data requires further optimization of iPS-derived cells to obtain mature and specific phenotypes representing the heterogeneity observed in the hearts and in blood vessels. Investment of resources in developing organoids and potentially using live myocardial slices are required. One system won't work for everything and appropriate systems need to be selected to address specific questions.

In summary, the network has made major progresses in establishing a version 1 of the Human Healthy Adult Heart Atlas and starting to map the blood vessel network. Data collected have already proven useful in mapping viral entry receptors during COVID-19 pandemic as part of collaborative studies from the HCA. More work is needed to increase the number of donors studied, cover ancestral diversity, increase the anatomical regions profiled prioritizing disease relevant districts. Several ongoing projects have started tackling human cardiovascular pathology and the next phase will see an increasing number of studies defining subtleties of disease.

Immune

Immune Bionetwork Community Events – open to the entire community

1- Monthly Immune Bionetwork Seminar Series

• We will invite labs from the immune network to present their ongoing work

• Meetings will be done monthly, with 2 labs presenting over 1hr (25min/lab)

• Scope of the meeting will be to present ongoing immune cell atlas work,

including individual projects and integration/harmonization across projects

2- Immune Bionetwork Annual Meeting

• Organize a first in-person meeting (and could have virtual component) in 2022

• Primary goal of the first meeting may be centered around cell type annotation

through a jamboree event

3- Inventory/survey of ongoing Immune Bionetwork projects

• Generate a detailed inventory of the different datasets being generated by member

of the HCA community

• Gather information about published datasets

• Google spreadsheet will be circulated to gather information

• Information will help prioritize the first immune cell atlas draft

Immune Bionetwork Coordination Meetings

Meetings 1-4: Invite 6-10 groups to present their approach to immune cell type

annotation.

• Participants: open to entire Immune Bionetwork and will invite a representative

per other network interested in immune cells; will be requesting volunteers to

participate

• Scope: Every meeting will be about 1hr; each group will have 20min to present a

• Primary goal: Have different groups present their cell type annotation strategy

through the lenses of their project o This includes discussing upstream data

processing, which features need to be prioritized that are more relevant to immune

cell annotation (e.g. TCR, antibody-derived tags (ADT) to prioritize markers for

splitting lineages and mapping cell states)

o Note: the goal is NOT to present on the specific science/biology studied

nor the discoveries made.

• Outcome: The summary of approaches undertaken will be posted in a Google doc

after each presentation and will be opened for comments/questions.

Meetings 5-6: Summarizing & discussing main frameworks for data processing

and cell types/states discovery and annotation

• Participants: open to the entire Immune Bionetwork and representatives of other

bionetworks interested by immune cells. Will want active participation from

computational experts.

• Scope: Will be up to 1h30 meeting (additional time to be allocated as needed).
•

Primary goal:

o Present a summary all the cell clustering & key annotation frameworks

presented (meetings 1-4) We are not looking for a one size fits all solution;

o Discuss what could be reasonable parameters to standardize clustering

analysis and cell annotations

o Consider leveraging several annotation strategies in parallel (some could

be specialized to specific modalities of measurements) and implement

ranking/metrics/consensus frameworks with cross-cutting terms. Manual curation

could focused on disagreements between parallel annotation frameworks

o Consider semi-automated approach, involving experts helping with manual curation.

o Until we have a solid first manually curated immune cell atlas draft, cannot rely just on machine learning

o Discuss optimal platform to empower crowd-sourcing annotation

• Outcomes: A set of general recommendations for annotation will be summarized in a Google document that will be available for comments for the broader community over the course of a 2-3-week period.

Meeting 7 – Jamboree on immune cell type annotation

• Participants: Open to all Immune Network; will seek to have groups made of computational biologists and immunologists to participate to the event

• Scope: In-person 2 to 3 day meeting (could be virtual)

• Primary goals: Test & optimize the annotation framework proposed by meeting 5

o Dataset: Combined COVID-19 PBMC datasets that will be made available to all the groups. Will involve a team of computational experts upfront that will help with data integration and initial clustering of main immune cell lineage. The dataset will need to be sizable as annotations become challenging with larger datasets

o Exercise: Team will be given 1 day and a half to annotate the different cell types. Depending on the number of participants, we may. divide the different cell lineages or ask them to analyze the full dataset. Each team will give a presentation of their annotation by day 2. By the end of the

meeting, we will compare annotation solutions between groups and discuss the source of discrepancies. These results may help in refining the annotation framework. Could consider comparing resulting manual annotation to automated annotation

• Outcome: Our refined recommendations for analysis & annotations will be summarized in Google doc & reported in a white paper. Results will be available on the data portal.

Meeting 8: Present annotation frameworks to the broader immunology community across different cell lineages to get their feedback

• Participant: Selected immunologist experts across different cell lineages

• Scope: 1h30 meeting per cell lineage group as well as 2-week window to review and comment on written recommendations (allocated time to be adjusted)

• Primary goals:

o Present the summary of our initial annotation framework recommendations to gather feedback from immunology experts across cell lineages. Need the buy-in of the broader immunology community. Need strategies to properly integrates legacy knowledge in our annotation

o Will make the cell types/states annotation results from jamboree available on platform to empower crowd-sourcing feedback on our annotation by the broader immunology community

• Outcome: Refined recommendations for analysis & cell type annotation frameworks

Meetings 9-11: Developing a roadmap for the Immune Cell Atlas (ICA)

• Participant: Open to all Immune Bionetwork and representatives from other bionetwork

• Scope: 3-4 meetings together with offline work on drafting the ICA roadmap. Will divide the different sections of the roadmap across volunteers involved. (allocated time to be adjusted)

• Primary goal: Draft the Immune Bionetwork roadmap, to be submitted as a white paper, including guidelines & plan for integrating different datasets & efforts o Roadmap will summarize analysis, annotation and nomenclature recommendations, including outcomes of the jamboree

o Roadmap will describe and discuss recommendation on different critical experimental and analytical topics, such as: What sampling & measurement strategies per tissues are needed to capture all existing immune cell states with >1% frequency? How many individuals should we profile across gender/ age range/ ethnicities/ geographical locations? Immune cells are defined by their functional response to immune challenges – How much effort should we place on non-challenged versus challenged states? What is the minimal set of modalities of measurements (e.g. GEX, CITEseq, ATACseq, spatial omics, functional studies) that are needed to define new immune cell states? How do we know if predicted

state differences through scRNAseq are biologically meaningful? Can we develop

scalable functional tests? How do we validate new immune cell states? How can

we work together to harmonize analysis & annotation across studies & tissues?

How to integrate immune cell data from other bionetworks and annotate across

organs How do we know if we found the same cell types/states between different

studies/tissues? How do we name predicted new cell types/ states? How do we

integrate historical nomenclatures? How can we best empower every tissue

network to contribute to the Immune network? How can we develop a user-friendly

online interface to empower immunologists around the World to explore the data &

contribute to annotation? What will be the criteria needed to conceive first

reference immune draft per organ systems and across the human body?

• Outcome: White paper publication Meeting 12 and onward (will start in parallel

to other meetings; more details to be provided upon receiving feedback)

• Participant: All key stakeholders (e.g., immunologists, genomicists,

computational biologists, technologists) interested in putting together a first

immune reference atlas. Will be open to all Bionetworks.

• Primary Goals: o Identify groups of immune cell datasets that would benefit from

being integrated/merged per tissue/organ system (based on survey/inventory data)

o Proceed with merging relevant dataset together and harmonizing data

annotation; will divide the work/dataset across interested groups. Could further subdivide merged datasets based on cell lineage

o One of the efforts could be an extension of the jamboree on COVID-19 PBMC data o More to be added upon receiving feedback.

• Outcomes: o Datasets uploaded on web-portal to empower community-based feedback on annotation and to enable broader community to navigate/query the data o Manuscripts detailing how datasets were merged and annotation harmonized

List of other general topics during the Immune bionetwork breakout session:

• Some immune cell atlas effort may go beyond immune cells when analyzing specimens such as lymphoid tissue.

• Need to actively bridge with all biological network, since all of them are profiling immune cells

• Second tier analysis: looking at immune transcriptional programs across all cells/tissues. Developing analysis frameworks that would be helpful to all Bionetworks

• Need to develop the annotation framework that can work for all human and mice

→ essential for mouse-human mapping. Need to actively engage the mice community • Initial immune cell atlas draft will continue evolving as more data and modalities will be integrated. Focus of first draft will depend on survey results of ongoing efforts

• Need to integrate disease/inflammation data in initial draft to empower mapping

cell states at higher granularity (involve members for ongoing

disease/inflammation efforts)

• Meta-data: what is the minimal set of meta-data to consider for normal vs. disease

datasets. Information on meta-data gathered will be collected through Immune

Bionetwork survey

Liver

The HCA liver atlas project has collected single cell data on over 100 human samples and is collecting a range of single cell, single nuclei and spatial transcriptomics and imaging data to help map the human liver. The liver breakout session identified new collaborators who we invited to join our monthly liver project planning and science sharing meeting series and trainee-led working groups in spatial transcriptomics, computational biology and tissue processing protocols. We also worked on a roadmap for the liver atlas project, covering topics, including: fundamental scientific challenges in liver biology, mapping liver cell types, mapping liver heterogeneity, spatial organization of the liver, supporting equity, diversity and inclusion to build a representative atlas, cross-cutting themes connecting to other tissues (e.g. endothelial cell, macrophages, wound response, and inflammation), data integration challenges, liver technical challenges and project organization & planning.

Our main action items are to publish a roadmap article this year and continue to expand the liver biological network.

Lung

Current state and use of HCA-Lung Atlas

What have we accomplished, what do we want to achieve, and when are we done?

The feeling in the breakout was that we are most certainly not done. First of all, we are only looking at static processes – and there was some discussion on how we could look at dynamics, what would be informative 'functional datasets'? Clearly, there are lots of aspects missing, including transitional/activated immune subsets, but the same holds true for normal variation of lung structural cells associated with specific demographics, environmental factors, etc. We are currently missing the following aspects of diversity (not complete):

- Altitude
- BMI
- Ethnicity
- Ages: under 20, centenarians
- Smoking (this is often not recorded with sufficient granularity)
- Vaping (often not recorded at all)
- Medication use (also very difficult)

In part, this boils down to lack of or difficulty with metadata recording (in part due to sample acquisition infrastructure). At the same time, we also need to know what type and degree of variability we do see in the current accumulated lung datasets. This information can then guide a statistical approach to estimate when we would have captured most of the variability, using the decreasing incremental fit to the model based on specific covariates such as for instance BMI or ethnicity. In an undersampled situation, however, covariates are not independent, and the source of the observed variance might remain elusive. Can we use other metrics to gauge completeness? Cell type composition for instance is a very difficult metric to use to evaluate the completeness of the atlas as this is so highly variable, dependent on sampling method.
Another aspect of completeness of the atlas involves the number of the locations sampled and the coverage of the common coordinate framework. Spatial methods need to be taken into account when evaluating the completeness of the atlas, and this does involve careful sampling of specific locations, including microdissections on discrete regions such as shown by the Tata lab in the plenary lecture. This needs to be tied in with 'classical' pathology and histological imaging to be able to inform on the exact structures we're sampling.

We might also need to accept that establishing the atlas will be an iterative process, at least for the coming years. Then a radically different approach to the question 'when are we done' could stem from the projected use of the atlas, depending on the goals we want to achieve as a Lung network. How useful is the atlas in it's current form for the community, including those without extensive bioinformatics expertise? The actual, practical use of the atlas and the feedback from the users could also guide further development: where are datasets of sufficient depth and scope, and where do we need more? References are used, even when imperfect, and consensus on the 80-90% of the most common cell types and states can be reached quite easily, this could be a good start for a reference atlas v1.0.

When we look at disease, how much additional variation does this bring? Does diseased lung involve entirely new cell states and cell types, or is it more of a variation of the healthy states? This would need a better definition of normal: how variable is normal? There was consensus on the difficulty to define normal. Normal represents a distribution, and most basic aspects of normality are known, the basic cell types, some of the major sources of variation. But to define this in great detail will be an iterative process that will take quite some time, still. Adding more modalities might actually help describe the stable states of a cell type, while the transcriptome data might well be inherently noisy and identify multiple states of a population of cells that is highly homogeneous at the level of the chromatin landscape. This also needs further developments in computational techniques and accepting some limitations of each method as well as some uncertainty in the outcomes.

There was consensus that detailed and accurate recording of metadata is key to guide these analyses and help estimate the various sources of variation in an integrated atlas. However, GDPR poses a threat to metadata collection, which is currently a real problem in mainland Europe. Another way to identify which are the essential metadata that need to be collected is to define a number of important biological questions that could be answered using the atlas, and then assess which metadata would be required to answer these questions. One example is the response to injury: a lot of environmental and demographic metadata could evolve around insults that injure the lung – the lung will also have only so many ways to repair this injury: we initially need to try and make these questions and the required metadata as generic as possible, and only focus on detailed, specific questions later.

In general, linking organ physiological (such as lung function), clinical, (histo)pathological or even (non-invasive) imaging (meta)data to the single-cell data could be a very strong use case of the atlas. One limitation, however, is that this would require consistent sampling of the structures/cells that would be meaningful for these functional parameters. To allow these sorts of analyses, well-phenotyped clinical cohorts might be of use, and targeted analyses in such cohorts could be a valuable addition to the Lung Cell Atlas.

The spatial aspect of the Lung Cell Atlas clearly requires more data. Visium is great in guiding annotation of cell types to specific locations within the lung, but lacks single-cell resolution. High resolution techniques with sufficient sensitivity such as MERfish, In situ sequencing or SCRNSHOT (not a complete list) will likely be more informative for mapping cells onto the lung tissue architecture and reveal location-dependent gradients of gene expression. Problems to be solved in this area are the high auto-fluorescence and the cell segmentation, so this might be resolved by combining protein and RNA detection and performing membrane

stainings. For 3D imaging, tissue expansion helps, as well as partial tissue digestion to allow better penetration of probes and antibodies. If we agree on what genes to use for probe design within the Lung network, that would be a great help for generating data across multiple labs that can then be more accurately compared. Both LungMAP and discovAIR have already made probe lists for spatial techniques, these will be shared/harmonized and distributed within the Lung network.

Finally, we discussed whether an atlas should use individual genes (as measured in most single-cell approaches) or gene programs, defined by the covariate structure of the genes for the atlas. The gene program also defines a cell type, and integration (or any follow-up analysis) is performed on the gene program, or the covariate structure of the genes, rather than on the individual genes. The current integrated datasets might well sacrifice some of the gene programs by forcing data together. These gene programs could be recovered by spatial but also using non-integrated scRNA-seq data. Several leading computational groups are working on methods to define covariance and gene programs that don't lose information by integration. This also establishes the concept of meta-cell by aggregating across cells - using both expression value and covariance (to capture missing data) – which helps to define very robust gene programs. Covariance of gene programs can also be a better measure of cell state than the mean expression of a single gene.

Musculoskeletal

The Musculoskeletal Biological Network aims to coordinate clinical and non-clinical researchers to map the musculoskeletal system across development and adulthood. This is an ambitious task given the diversity of musculoskeletal tissues and the challenges of working with hard-to-access tissues with abundant extracellular matrix.

Our short-term goals are to:

1. Build our membership of interdisciplinary researchers with an interest in the musculoskeletal system.
2. Survey our network and potential members to identify current progress, challenges and commonly used methodologies.
3. Initiate quarterly Musculoskeletal BioNetwork meetings to present research, build collaborations and discuss challenges
4. Continue to explore methodologies and generate "Sky Dive" data to build atlases

5. Outline our initial roadmap towards an atlas. Analysis of certain musculoskeletal tissues, for example synovium and tendon is underway within network groups, and our roadmap will need to reflect the different challenges and stages of research of each tissue type.
6. Integrate and interact with other BioNetworks that share common tissues (development) or challenges (genetic diversity, kidney)

Our discussions at the HCA Annual meeting centred around metadata requirements and how this can be harmonised within the constraints of tissue collection ethics. The requirement for cell annotation was discussed and we identified the need for jamborees or equivalent to assist annotation efforts, alongside interaction with other BioNetworks to identify common cell types. discussions centred around a cohesive inventory of atlased cells that allows tracking of progress without duplication of reporting once data is formally deposited. Aligned with this, we also discussed tissue cartography approaches, with emphasis placed on how best to document and then deposit data so the location of each tissue sample transcriptomically analysed is recorded appropriately. It was suggested that we should also reach out to other biological networks to understand how each different network is approaching this topic.

Nervous System

The structure of the biological network: dominated by the large and well-funded NIH BICCN consortia, but with several additional funded projects also studying brain tissues (e.g. the Swedish Human Development Cell Atlas, the EU-funded HCA network).

We discussed the challenge of counting cells in 3D, e.g. to set scale for cell census or to detect cell pruning during development. The need for automating 3D image annotation at scale, e.g. with neural network.

We talked about what is currently missing from ongoing projects: e.g. peripheral nervous system, early and late development, pediatric. Non-human primates can be a great resource to fill in gaps in time or tissues that are difficult to sample. Evolutionary conservation is also an excellent metric for defining likely functional

cell types and genes, e.g. to identify those genes that really support the function of each distinct cell type.

Human diversity: how many samples needed to cover e.g. ethnic diversity. Ed Lein thinks thousands of brains will be needed, and this implies focusing on a small number of regions that can be reliably identified. Sten Linnarsson agreed but proposed that sex/gender differences are a good start where differences may be more robust in specific regions of the brain. Could use mouse to get a first handle on where most of the genetic diversity is located in the brain. It will be important to start without prejudice, since we know so little about where genetic diversity affects the brain, how, and to what extent.

Re detail of atlas: there will be Lumpers and Splitters but sometimes no need to go to maximum detail - e.g. therapy may not need the finest level, but rather focus on targeting broader classes of cells.

Finally, we had an interesting chat about all the intriguing and exciting hypotheses that are generated while building atlases, and how we don't have time to pursue them all. We proposed at some point to write a review covering exciting questions and observations, with no answers, to stimulate people following up interesting hypotheses.

Oral and Craniofacial

Our network's history of assembly and collaboration is relatively recent; however, it has grown from 4 to 25 investigators over the last 9 months. We roughly balanced between sexes/genders and are >70% early career researchers. While this is exciting to see, we are aware that we are heavily reliant on North American and European investigators. Part of our conversations at this HCA meeting were focused on more intentional recruitment to our network from the global South, Africa, Asia, and Oceania. With current and future funding opportunities, we believe this is an exciting opportunity to grow our network with these partners and true community engagement.

Further, since establishing our bionetwork in September 2020, there have been a number of first oral niche-specific publications in the adult oral cavity including for buccal mucosa, gingiva, tooth pulp, and minor salivary glands. We also believe that there are additional unpublished datasets of a major salivary gland (parotid) and

dorsal tongue. There are two separate, integrated datasets (buccal+gingiva as well as salivary gland+gingiva) that have also been published. In total, we think there are about 250,000 published and unpublished cells that are in need of integration, annotation, and spatial validation. Our current conservative estimation is that among these 250,000 cells, there are at least 10 novel cell types for the human body. **We believe that this integrated oral and craniofacial cell atlas is the most essential task for v1 of the HCA.** Regarding this, much of our discussion this week focused on the need for more cell sequencing within already sequenced niches and also in a few additional sites (tonsils and dorsal tongue were primary sites of interest). Additional needs were for funding (ie. our own oral seed network) and computational scientists to join our network to participate in dataset integration and annotation.

In rank order, we walked away from this meeting committed to 1) find computational partners to work towards sequencing/integrating all relevant niches for v1, 2) establishing annotation working groups within our bonetwork for these integrated niches, 3) joining other bionetworks to establish the potential for meta-integrated analyses between orals/tissues/systems, and 4) continuing--at minimum--our annual symposium (next one in June 2022). Once we have contributed the more important niches to v1, we will establish a greater vision for how to comprehensively define the needs for v2.

Ines and I will be meeting over the next few weeks to further discuss strategies for further collaboration and discovery within and among bionetworks. **Please reach out to us for collaboration to achieve these goals.** Our network is currently connected to the skin, gut, lung, immune, genetic diversity, developmental, and organoids bionetworks. We look forward to establishing relationships with the adipose, musculoskeletal, and nervous system networks as well considering our unique niche houses these tissue types as well.

Organoid

1. Accessing and interacting with organoid data within the Human Cell Atlas
-> The Organoid Cell Atlas Portal for which Oliver Stegle's lab is currently preparing a first version as part of the HCA|Organoid project is going to be central for utilizing organoid data within the Human Cell Atlas. We talked about use cases that are specific to organoids and will facilitate the integration with data from organ-centric atlases.

2. Handling data protection / GDPR for patient-derived organoids and matched primary samples
-> The planned extensions of the HCA Data Portal that will support controlled access data will make it possible to include sequencing data of living patients. This is very important and in fact essential for most patient-derived organoids.

3. Broadening the organoid data that are available within the Human Cell Atlas infrastructure
-> We encourage everyone to consider submitting their papers about single-cell sequencing of organoids to the HCA Publications Committee and the data to the HCA DCP. There is a lot of single-cell organoid data that is produced by individual labs, and it would be good to channel more of those data into the HCA infrastructure to facilitate access.

4. Interaction with other HCA Biological Networks
-> As a cross-cutting Biological Network, there is lots of complementarity and scope for interaction with the organ-specific biological networks, and of course with the Developmental Cell Atlas community. These synergies will be pursued through the HCA Biological Networks seminar series and other channels.

Pancreas

Discussed themes:

- Technical platform for data sharing
  - **Ulrike Taron (Charitè-Universitätsmedizin Berlin)** – introduced the cloud workspace recently developed by the data management team of the ESPACE consortium. This cloud is deployed as a Kubernetes container and has four goals:
- Make data accessible and interoperable and in agreement with FAIR principles
- Act as a repository of code and pipelines
- Provide applications for interactive visualization of data
- Provide a smooth interface for the future integration of data in the HCA Data Coordination Platform.
  - **JP Cartailler (Vanderbilt University, Nashville TN)** – presented the underlying structure of Pancreatlas platform, human pancreas-specific biological imaging resource capable of displaying 150+ different image formats and metadata.
- Pancreatlas with its Flexible Framework for Integrating and Navigating Data (FFIND) was developed with the Javascript-based React framework and can

connect with other resources and support countless imaging or other structured data management needs (built with FAIR principles).

- The FFIND application programming interface (API) retrieves and displays images in PathViewer, the web client associated with the OMERO Plus server (Glencoe Software).
- FFIND (https://github.com/Powers-Brissova-Research-Group/FFIND) is open source technology solution used for an organ-specific implementation (Pancreatlas). Perhaps other HCA bionetworks might find it useful to integrate to their needs.
  - In the follow-up discussion, there was a general agreement to connect the two portals to integrate omic and imaging datasets. The first step will be to establish common vocabularies and ontologies and the definition of APIs to share and exchange data.

- Procurement of islets and tissues
  - **Wilko Weichert (Technical University Munich)** – gave an overview of the strategy adopted by the ESPACE consortium to procure fetal and adult pancreas biopsies, coordinating partners from the Leiden University Medical Center (Leiden) and the San Raffaele Hospital (Milan, Italy).
- Biopsies from different regions of the pancreas (body, head, tail, processes uncinatus) are used for single-nucleus RNA and ATAC sequencing and for proteomic approaches (CODEX).
- Cryosections and FFPE samples are also prepared for histological analyses.
  - **Marcela Brissova (Vanderbilt University, Nashville TN)** – provided an overview of US-based resources and initiatives currently collecting human pancreas/islet samples & datasets**.** This included:
- Infrastructure for organ procurement
- Procedures for multimodal analyses of pancreas and isolated islets
- Connecting/integrating imaging dataset from these programs by Pancreatlas platform.
  - The follow-up discussion focused on redacted donor clinical data, donor age, and overall definition of "normal" healthy pancreas. In normal donors >50 years old, it is common to detect lipomatosis, fibrosis, cystic dilated ducts and pre-cancerous lesions. Samples procured in the ESPACE consortium are currently kept in a centralized tissue bank in Munich and could be available for further analyses and validations.

- Developmental aspects of pancreas
  - **Françoise Carlotti (Leiden University Medical Center)** – gave an overview of fetal pancreatic tissue procurement form elective abortions (up to 22-weeks gestational age) along with ethical and regulatory items.
- Analysis of these precious tissues will integrate single nucleus RNA-seq, ATAC-seq and CODEX multiplexed imaging

- The is to procure tissues from 4 different stages of fetal development
  - **Diane Saunders (Vanderbilt University, Nashville TN)** – covered the procurement and analysis of pancreas from pediatric donors (up to 10 years of age) including
- Rapid organ expansion after birth
- Rearrangement of islets and pancreas architecture
- Analysis pipelines at Vanderbilt and integration with islet functional studies
  - Common interests of the groups will favor and speed up closer collaborations in the near future.

Next steps:

- Have a meeting scheduled on 7/15/21 to discuss Pancreas Bionetwork and collaborations between ESPACE and representative of pancreas and islet mapping efforts in the U.S.

Bionetworks to connect with:

- Developmental Atlas
- Pediatric Atlas

Reproductive

The spatiotemporal dynamics of the tissue, the influence of hormones and differences to their rodent counterparts, makes the study of the reproductive system challenging. During this session, we had the opportunity to articulate the challenges we are all facing when studying reproductive tissues, including how we liaise with cell annotation or how we can functionally validate our results. We also discussed how we can share datasets within the teams and establish connections between us. Finally, we talked about future work for our reproductive network, from mapping tissue architecture to studying cellular function.

Cellular annotation. We know very little about reproductive tissues, and that makes the annotation of cells complex. For example, a lot of us obtain multiple clusters of stromal cells that are challenging to annotate. We agreed that the annotation of cells using specific genes expressed and location of cells in tissue is the best way to go. Transcription factors are also helpful to annotate cell states, and scATAC-seq facilitates its detection. We recognise the value of proteins to define function, but the dynamics of proteins and RNA is different in this tissue. Thus, it may be

confusing to annotate clusters using proteins without taking into account their dynamics. Using multiple approaches to validate our findings is also helpful.

Spatial data. Novel technologies are emerging, and it is often difficult to choose between high-throughput and sensitivity. We thought it is relevant to define our question before choosing the desired technology. For the exploratory phase, we may exploit high throughput technologies such as spatial transcriptomics or nanostring. For the validation phase, we may utilise more targeted approaches such as ISS or multiplexed smFISH. Some partners are doing 3D maps of reproductive tissues, and highlighted the relevance of them to understand tissue architecture. We also discussed the importance of considering proteins, and studying the spatiotemporal dynamics of RNA and protein in this tissue. In line with that, some partners mentioned that monoclonal antibodies on the market are great and tend to work great for them. We thought it would be interesting to have a shared list of "working" antibodies between us in the future.

Numbers & ethnicity. Numbers matter as they allow us to obtain greater granularity on the clusters we define. While some partners were more optimistic about computational methods to integrate datasets, some others were more hesitant. However, we all agreed it is important to process samples using optimised protocols for tissue digestion, as otherwise, low quality data may introduce some bias. Increasing numbers mean we should also be more inclusive. Genetic differences in reproduction have been observed for some reproductive disorders (eg. preeclampsia) but not for others (eg. ovarian cancer). We should make a collective effort to include samples from multiple genetic backgrounds and study the function of this genetic diversity.

Animal models. Organism models have proven useful to functionally validate our findings. Reproductive tissues are different between species, and we acknowledge the importance of performing cross-species comparisons. When doing so, it is important to look for the general program and not the expression of specific genes, which may differ between species. Some partners have a lot of expertise in that area and have shared beautiful computational and experimental approaches to resolve such a challenge. While we may not find an ideal animal model for reproductive tissues, we may be able to choose specific ones depending on the question we have in mind.

Perturbations & in vitro models. In the last decade, there has been an exponential growth in the development of *in vitro* models to study reproduction. We are all excited to leverage our reproductive maps to improve the *in vitro* conditions by, for example, mapping ligand/receptor pairs. *In vitro* models allow us to perform perturbations to better understand functionality of cells and define new cell states. There is a need to develop *in vitro* models for some reproductive tissues, and those can be benchmarked against our atlases. Another way of perturbing cells is looking at disease states. Some states that we find in disease may not be found in healthy conditions, and we think it is important to also consider them. There was also enthusiasm for profiling environmental perturbations, and how this could affect cell function. The other perturbation considered by some members was the microbiome and its effect on cell function and phenotype.

Future directions. We think there is lots to be done in the field. We should start looking at tissue architecture more closely, and consider epigenetic changes. We should increase the size of our cohorts but also take into consideration multiple genetic backgrounds. Comparing *in vivo* with *in vitro* data as well as performing cross-species comparison will allow us to select specific species to functionally validate our results. Finally, we think there is room for improvement of current *in vitro* models. *In vitro* models will allow us to profile perturbed states, not accessible so far in our dataset.

Skin

The skin breakout session had a global and interdisciplinary reach, with scientists joining from e.g. Australia, Europe, Japan and US, and expertise in skin biology, single-cell technology and computational analysis. After a short introduction about the goal of generating the roadmap skin, its purpose and importance, we invited discussions on:

(1) What resources and information would be useful for the skin network to support single-cell biology of individual groups and/or teams.
(2) How to form an even more inclusive and international network in order to inform, communicate and collaborate easily and efficiently.
(3) What are specific challenges for the organ skin; we started to discuss how to reach a first consensus nomenclature.

Discussion ideas and outcome to these three topics:

(1) Resources and Information:
1a. It would be important to have a representative set of markers for cell types that can be used, to save time and avoid confusion (see point 3 below).
1b. A more efficient way to communicate and share resources for all of us such as protocols for cell dissociation (e.g.: sharing via depositing on *Protocols.io).*
1c. Especially for “newcomers" in the field, it would be helpful to have access to critical evaluation of the available data and methods. How to do this best, no clear consensus has been reached; one suggestion was to start a forum.
1d. Better access to metadata and original cell annotation; currently much of this information is not easily available and significantly slows down data-reanalysis and/or integration.
(2) Inform, Communicate, Collaborate:
2a. Regular (non-formal) web meetings to discuss specific topics such as - Computational issues (integration, spatial alignments, cell annotation) - Biological issues (sampling and processing of tissues, protocols) - Technology (up-to-date with single-cell genomics, proteomics, spatial methods).
2b. Provide an easy tool to communicate, such as SLACK. One such channel already exists at HCA (#skin).
2c. Outreach via our newly launched https://skincommunity.org to inform about webinars, tools, etc.
(3) Provisional consensus nomenclature.
3a. We discussed strategies to generate a first provisional consensus nomenclature of skin cell types and states. Important notes here were that this first nomenclature should not be rigid, i.e. rather modular to be updated when new datasets, spatial and/or protein information etc. become available. It would be important to be able to “track” the evolving changes in nomenclature in order to refer back and connect upcoming with earlier studies. A further wish/suggestion was that it would be very helpful to have a web tool to query data against an existing reference.

In sum, this first global HCA skin bionetwork discussion was very fruitful and lively, and we will continue our discussion on 8th of July 2021 at the HCA Bionetwork Seminar featuring Oral/Craniofacial and Skin Networks in the *Skin breakout room.*