Summary of the NIH-HCA Meeting

1. Clinical Metadata

1.1: Establishing best Practices for defining core clinical metadata across groups/consortia

Problem we think we can solve by collaborating

- Establishing best practices that can be utilized across consortium and consortia or suggesting metadata types/levels available for consortia to adopt
- maximize the ability to utilize data across the research process (collection, aggregation, analysis)
- Maximizing the **context** of the sample (everything we know about it) so analysts have what they need and filling in the gaps of what was previously missing.

Recommendation(s)/ What we want to do -

- Develop a group to assess “meaningful” clinical data to define standards that would enable cross consortia analysis

Names (with Consortia) on the team:

- Moderator(s): Kristin Ardlie (GTeX), Sarah Mazilli (HTAN), Melissa Cook (HTAN/NCI)
- Notetaker(s): Sean Hanlon (NCI)
- Other Group Members and Consortia each work with: Marc Halushka, Jessica Langer, HTAN (NCI), Ian Fore (HTAN,NCI)

What additional expertise do you need? None

Our next meeting is? None

1.2: Identifying Consensus on Recording Provenance

Problem we think we can solve by collaborating

- Identifying Consensus on recording provenance - as part of the clinical metadata. This would include sample naming/identification

Recommendation(s)/ What we want to do -

- Ian’s suggestion: Given this occurs “everywhere” we should establish this as the home for the issue. I had talked with Jonathan Silverstein about collaboration on this..

Names (with Consortia) on the team:

- Moderator(s): Kristin Ardlie (GTeX), Sarah Mazilli (HTAN), Melissa Cook (HTAN/NCI)
- Notetaker(s): Sean Hanlon (NCI)
- Other Group Members and Consortia each works with: Ian Fore (HTAN/NCI)

At least one Email contact:

What additional expertise do you need?

- The architects of https://data.humancellatlas.org/metadata/design-principles/structure

Our next meeting is? None

---

2. Data Architecture and Integration

2.1: Defining Common Query Use Cases

Problem we think we can solve by collaborating –

- Lowering the barrier to query observation by feature matrices, and offering summary statistics.

Recommendation(s)/ What we want to do -

- Follow-up presentations of use cases and interfaces currently in the wild.
- Integrating efforts with the GA4GH RNAget initiative.

Names (with Consortia) on the team:

- Moderator(s): Laura Clarke
- Notetaker(s): ###, Tim Tickle
- Other Group Members and Consortia each works with:
- At least one Email contact:
- Tim Tickle

What additional expertise do you need? None

Our next meeting is?

- Carry on this conversation in the current HCA Portals Community call (that can be rebranded for this larger group).
    - Nils (HuBMAP)
    - Bruce Herr (HuBMAP)
    - Katy (HuBMAP)
    - Daniel Miller(BRAIN - program side)
    - (BRAIN BICCN - NeMO Archive)
    - Nikolay Markov (Lung seed atlas)

2.2: Data Storage and Data Movement

Problem we think we can solve by collaborating –

- Common processing pipelines to minimize data generation and movement
- Common standards for data inventories
    - Mapping in to terms, shared ontology resource
- Common APIs across data stores

Recommendation(s)/ What we want to do -

- Host a discussion on common processing pipelines (transcriptomics/epigenomics) BICCN with HUBMAP, HCA, other consortia. (Carol, Anup, Bill will find right hubmap contact).
- Follow up on cross-consortia data inventory standards (e.g., Datacite-like level of metadata); how to make to make this searchable across consortia (Satra, Anup, Bill+HubMap)

Names (with Consortia) on the team:

- Moderator(s): Owen White (BICCN), Carol Thompson (BICCN)
- Notetaker(s): Carol Thompson (BICCN), Satra Ghosh (BICCN/BRAIN Initiative)
- Other Group Members and Consortia each works with: Bill Shirey (HuBMAP), Anup Mahurkar (BICCN)

At least one Email contact: None

What additional expertise do you need? None

Our next meeting is? None

2.3: Interchanging data and metadata between consortia

Problem we think we can solve by collaborating –

- *Highest priority HMW. Clear preference for interchange data and metadata.*

Recommendation(s)/ What we want to do -
Action items:

1) Form a cross-consortium working group to examine the central inventory and determine how similar or different the data standards are across each consortium. Each consortium should identify individuals to perform an inventory of data formats across all modalities and deposit this information in a central location.
    a) Perform inventory of data types being supported for each modality
    b) Develop criteria describing what “good” data formats should be (e.g. open source). What does it mean to be a useful data format?
    c) Publish a report summarizing the current state and recommendations moving forward.
2) Make recommendations to funding agencies who can then officially support and push for these formats for all new and existing studies.
3) Repeat this process for metadata and data models

Names (with Consortia) on the team:

- Moderator(s): Alex Ropelewski (BRAIN/BIL, HuBMAP)

- Notetaker(s): Josh Campbell (HTAN), Kylee Degatano (HCA, BRAIN), Sharmi Ghosh-Janjigian (HTAN, NCI)
- Other Group Members and Consortia each works with: Maryann Martone (BRAIN-BICCN;), Jason Swedlow (OME), Tyler Best; Matt Wyczalkowski (HTAN, CPTAC3)
- Idan Gabdank (ENCODE;) Anna Maria Masci (LungMap phase1)

At least one Email contact: Alex Ropelewski (BRAIN/BIL, HuBMAP)

What additional expertise do you need?

- Each consortium needs to appoint a representative who will be familiar with the data types being used.

Our next meeting is? 2 weeks : Alex will organize a zoom meeting.

---

3. Multiplexed molecular imaging_spatial mapping tools

3.1: Biospecimen Preparation for high throughput, 3D Multi-modality imaging

Problem we think we can solve by collaborating –

- How to preserve the spatial and quantitative aspects of different molecular modalities during sample preparation, such as with hydro-gel embedding?
- Is the procedure compatible with RNA, DNA and protein detection?
- How to ensure robustness, consistency?
- How to improve reagent delivery?

Recommendation(s)/ What we want to do -

- Come up with some kind standard ground truth for evaluating different sample prep protocols.
- Establish channel for communication between programs, sharing of reagents / SOPs / definition of tech needs
    - Contact existing groups to establish a WG to flesh out concept, subsequently move forward with respective Steering Committees
    - Members should include interested parties (Bing Ren, Anup Sood , Peng Yin , Joe Gray and others). Will send out an email to the different consortia
- Develop new technologies to answer the overall question.

Names (with Consortia) on the team:

- Moderator(s): Peng Yin
- Notetaker(s): Philipp Oberdoerffer, Bing Ren,...
- Other Group Members and Consortia each works with:
- At least one Email contact:

Current members of the working Group:

- Bing Ren (UCSD)
- Peng Yin (Harvard Medical School)
- Anup Sood
- Joe Gray (OSHU)
- Hamda, Natnae (Astellas)

What additional expertise do you need? None

Our next meeting is? None

3.2: Metrics for Multi-modal assays

Problem we think we can solve by collaborating –

- How to account for noises (such as barcode collision) in the datasets?
- Many sources of noise - fragments, barcode collisions - can we computationally remove them?

Recommendation(s)/ What we want to do -

- Work with the Tools group

Names (with Consortia) on the team:

- Moderator(s): Bing Ren
- Notetaker(s): Dena Procaccini
- Other Group Members and Consortia each works with:

At least one Email contact:

What additional expertise do you need? None

Our next meeting is? None

4. Spatial Profiling Tools

4.1: Improving validation and standardization of antibody reagents for imaging and extending multiplex imaging to volumes and combinations with RNA methods

Problem we think we can solve by collaborating –

1. List of validated antibodies (clones, company) by platform, organ, tissue preservation method
2. Standardization of imaging methods using common tissue shared across different labs/imaging platforms
    a. Already happening with **Trans-Network-Project" at HTAN (SARDANA - more info: Prof. Peter Sorger).**
3. Standardization of cell segmentation, cellular identification, and spatial analysis
4. Creating a 3D atlas with volume imaging
5. Integrating protein/RNA methods

Recommendation(s)/ What we want to do -

- Form a Cross-consortia Working Group: Aim to Publish a Study in 12-15 months, tightly designed to use a few antibodies on a few targets using 2 platforms for spatial imaging (and link to Ab-validation procedures)
- Form several working groups who meet regularly to advance and, importantly, standardize the following fields/topics:
    - Data analysis: Coordinate with 2nd breakout session
    - Volume imaging 3D: Point-of-contact: Ronald Germain
    - Integrating protein/RNA methods: Point-of-contact: Emma Lundberg, others?

Action Items -

- Ron-Begin to put an initial group together (Stanford/CODEX, MIBI, MIT, HCA-Sarah, Elizabeth, Denis, Andrea (Germain), Jeannie Camilliro, Ajay Pillai, Julie Kim (HuBMAP); Fiona Ginty (HuBMAP), Jonathan Bock

Names (with Consortia) on the team:

- Moderator(s): Ronald Germain (NIH IRP), others?
- Notetaker(s): Dena Procaccini (NIH OD), Andrea Radtke (NIH IRP)
- Other Group Members and Consortia each works with: Elizabeth, Emma Lundberg Jonathan Bock

At least one Email contact: None

What additional expertise do you need?

- MIBI platform expertise (Sean Bendall lab?)
- Buy in from funders to support validation

Our next meeting is? TBD

4.2: Dissemination of image-based spatial transcriptomics standards and methods

Problem we think we can solve by collaborating –

- Broader dissemination of methods/technologies and standards

Recommendation(s)/ What we want to do -

1. Up to date comprehensive protocols on protocols.io (or other central location) from instrument setup to analysis.
    a. Leverage protocols.io and slack to develop a community to answer questions on these protocols (and methods)Q
2. Hands-on training workshops
    a. CSHL/Woods Hole type workshops for all imaging-based single-cell transcriptomics methods
    b. Workshops conducted by method developers for individual methods
    c. Pursue support for these efforts: NIH, CZI, HCA, others
3. Standards and references for larger community e.g. standard gene sets, standard tissues
    a. Identify standard tissues in model organism, e.g. mouse
    b. Identify standard tissue distributors, especially for human tissues
    c. Coordinate with KPMP - they have done tissues sharing benchmarking efforts
    d. Identify standard gene sets for different organs
    e. A common repository of validated probe sequences
    f. Pursue support for these efforts: NIH, CZI, HCA, others
4. Seed the development of next generation analysis approaches

Names (with Consortia) on the team:

- Moderator(s): Xiaowei Zhuang
- Notetaker(s): Norbert Tavares (HCA), Jeff Moffitt
- Other Group Members and Consortia each works with:
- At least one Email contact:
- Xiaowei Zhuang; Norbert Tavares; Alexandre Denadai-Souza; Jeff Moffitt

What additional expertise do you need? None

Our next meeting is? None

4.3: Benchmark tissue consortium

Problem we think we can solve by collaborating –

- Technologies are validated and deployed on a wide range of tissue types that make comparisons across technological platforms, and biological validation of those technologies, very challenging.

Recommendation(s)/ What we want to do -

- Develop an interdisciplinary working group to define a set of tissues, and preparation conditions, for benchmarking and validating spatial technologies. We recommend 3-5 tissues from across the body--including the brain (often used in tech dev in this space), as well as other somatic tissues with very disparate embryological origins.
- The important thing will be to define biologically driven metrics of performance for these tissues. These could include:
- Cell type localization. Many tissues have cell types whose distributions have been known for decades. Single cell data offers an often more-granular, and highly quantitative definition of cell types. Establishing “tiers of granularity” to evaluate technology performance is key.
- Define cross-tissue commonalities that can serve as anchor points: for example, vascular structures and immune cells are present in basically all somatic tissues. These could be very helpful for determining the relative performance of technologies on different tissues.
- Histological features: every tissue has histological features with wide-ranging shapes and sizes. Pre-defining these features using gold-standard measurements enables a clear benchmarking of technology resolution.
    - For example, many brain tissues are laminar. We can measure the laminar thickness of these features using histology, and compare the thickness in various spatial technologies.
- Common coordinate framework registration: the tissue types identified should have a common coordinate framework on which to register the samples taken, in order to be certain that measurements are being made in the same location in the tissue

Names (with Consortia) on the team:

- Moderator(s): Evan Macosko
- Notetaker(s): Andrew (Farmer) made some notes in the green boxes.
- Other Group Members and Consortia each works with:
- At least one Email contact: Andrew’s email:

What additional expertise do you need?

- This effort would require: 1) technologists inventing new measurement tools; 2) biologists making measurements on the selected tissues with the relevant knowledge to define credible biological benchmarks in each tissue.

Our next meeting is? None

4.4: Sharing and comparing of existing multiomic protocols (RNA + protein in same sample)

Problem we think we can solve by collaborating

- improve method sharing, begin standardization of approaches and commonly used workflows

Recommendation(s)/ What we want to do -

1. Pull together existing protocols from within this group;
2. Reach out to others in this space and solicit their protocols;
3. Share protocols in a accessible location (via protocols.io, other methods);
4. Establish where there is consensus across protocols/processing that apply to multiple tissue types (or are there unique protocols for different tissue types);
5. Develop a common set of control samples for benchmarking;
6. Develop standards allowing to normalize across methods
7. Controls materials;
8. Report to community

Names (with Consortia) on the team:

- Moderator(s): Fiona Ginty
- Notetaker(s): Jenny Rood
- Other Group Members and Consortia each works with: Zorina Galis, Sebastian Pott, Mike Snyder, Ken Lau,Suvarna Gandlur Nasreen Haque

At least one Email contact:

What additional expertise do you need?

- Representatives for each technology

Our next meeting is? TBD

4.5: Comparing clinical sampling/phenotypic information to contextualize multi-omic data

Problem we think we can solve by collaborating –

- Correct interpretation of multi-omics data would be facilitated by having more detailed meta data about the sample/patient origins (e.g. metabolites/RNA expression may be differently affected in different ages/disease states, therapies etc).

Recommendation(s)/ What we want to do -

- Develop guidelines for clinical collaborators (e.g. similar to REMARK biomarker guidelines); develop control samples and surrogate molecular markers for environmental exposures (smoking, etc.), damage, markers for tissue quality (i.e. hypoxia); spatial location in the original sample/organ. Make it easy to access and track with the omic data.

Names (with Consortia) on the team:

- Moderator(s): Fiona Ginty
- Notetaker(s): Jenny Rood
- Other Group Members and Consortia each works with: Sebastian Pott, Mike Snyder, Zorina Galis, Ken Lau, Nasreen Haque
- At least one Email contact:

What additional expertise do you need?

- Clinical data, EHR records expertise, understand the limits of data collection today, what could be done going forward for future studies?

Our next meeting is? TBD

5. Development and Pediatric

5.1: Interpreting the emerging development atlas

Problem we think we can solve by collaborating –

1. Defining CCF in development
2. How to best use of early embryos by consortium. Common mechanism to find samples. Each group has their own inventory systems. Adoption common metadata would help cross-search (e.g. HCA DCP metadata).

Recommendation(s)/ What we want to do -

- Trajectory analysis is key to understanding dev biology;
- Efforts need to be made to avoid sparsity, e.g. in terms of developmental time points, e.g better computational methods; fill in gaps and build dense temporal sampling with model organisms that allow mapping of rare human samples
- Very important to have CCF for the embryo because position with respect to embryo really matters, set up developmental biology CCF network or link to existing frameworks; linking to anatomists' knowledge (rich historical knowledge and nomenclature)
- Have a common repository for data sets (including model organisms) incl. Data generators. I.e. in the first instance just list of data sets, then step 2: large scale integration = "registry"
- Call to funders to support registry (may be general not just dev biology)
- Possible mechanisms are via central coordination or incentives
- Possibly identify 'champions' for registries; include trainees in efforts

Ideas for repositories:

- Who is doing what?
- What data sets are out there? (including model organisms)
- have a landing page portal and link to the different resource pages & existing data bases; could be done via HCA webpage ?
- For sample access: links to existing biobanks

Action Items:

- further methods development for multi-omics profiling of single samples.
- identify personnel in different labs who may curate existing data and push repositories forward
- pooling data to address questions that require larger number of individuals, e.g. male vs female analysis.

Names (with Consortia) on the team:

- Moderator(s): Muzz Hannifa, Alain Chedotal, Deanne Taylor, Jonah Cool
- Notetaker(s): Kerstin Meyer, Jonah Cool, Gary Bader
- Other Group Members and Consortia each works with: Gary Bader (UToronto), Marc Charette, Gloria Pryhuber (LungMAP)

At least one Email contact:

What additional expertise do you need?

- Developmental biology community for CCF
- Bioresources to coordinate sample collection
- Organ development anatomists (some clinicians are very good)

Our next meeting is? None

5.2: How can we develop a common coordinate framework (CCF) for development and pediatric atlas projects that change over time?

Problem we think we can solve by collaborating –

- Develop dynamic CCF

Recommendation(s)/ What we want to do -

- Learn about existing CCFs & ontology terms / standard vocabs (many exist) e.g. second heart field, neural crest
- Ensure the HCA CCF technology community is linked to projects that need dynamic CCFs

Names (with Consortia) on the team:

- Moderator(s):
- Notetaker(s):
- Other Group Members and Consortia each works with:
- At least one Email contact:

What additional expertise do you need?

- Developmental Biologists
- Developmental Geneticists (ex: Sharon Plon, Phil Lupo, Wendy Chung…)

Our next meeting is? None

5.3: How can we develop collaborations between HCA and developmental biology communities?

Problem we think we can solve by collaborating –

- Define biological questions to answer by setting up collaborations directly with developmental biologists

Recommendation(s)/ What we want to do -

- Coordinate with existing devbio and pediatric conferences to have an HCA/single cell session to help spark collaborations and identify important scientific questions from these communities that single cell data can help answer, such as:
    - Society for Developmental Biology (SDB) https://www.sdbonline.org/2020mtg_satellite_symposia
    - British Society for Developmental Biology (http://bsdb.org/) - Katrina Gold or Sheny can link to this.
    - American Society of Human Genetics meeting (https://www.ashg.org/meetings/2020meeting/sessions/ancillary-industry-events/) Application Deadline: May 22, 2020. Meeting: Oct 27-31 2020. organisers/committee to find common themes for HCA/HubMAP to work together on.
- There may be possibilities for combined training and annotation events where participants can gain analytical skill and help inform biological interpretation of existing data.
- HCA Development network leads can send an email to the organizers of these conferences to ask if it is possible to set up a session.
- Engage dev biologists to Human Development meeting

Names (with Consortia) on the team:

- Moderator(s): Muzz Hannifa, Alain Chedotal, Deanne Taylor, Jonah Cool
- Notetaker(s): Kerstin Meyer, Jonah Cool, Gary Bader
- Other Group Members and Consortia each works with: Gary Bader, Marc Charette, Gloria Pryhuber

At least one Email contact:

What additional expertise do you need? None

Our next meeting is? None

5.4: How can we access developmental and pediatric samples?

Problem we think we can solve by collaborating –

Identify international sources of biospecimens and properties of each (e.g. what types of samples are available and if international sample sharing is possible)

Recommendation(s)/ What we want to do -
Start off with a database of what is available e.g.

- HDBR, UK
- HUDECA, France
- Birth Defects Research Laboratory (BDRL) at University of Washington, Seattle. https://grantome.com/grant/NIH/R24-HD000836-49
- Toronto: https://biobank.lunenfeld.ca/?page=About%20Us - gestational ages from 8 to 21 weeks gestation, frozen and fresh (early trimester tissue samples) samples. Available on a cost-recovery basis.
- Neonatal organ donor network (link via Gloria Pryhuber, lungmap) BRINDL: /home?
- May want to explore additional connections to NICU tissue resources that exist and are potentially underutilized

Also check w/funders/organizers of postmortem pediatric tissue sourcing like Swifty: https://www.swiftyfoundation.org/initiatives/post-mortem-tissue-collection/giftfromachild/

Efforts to generate a common framework that is based on obtaining similar ethics so that they can be shared across studies.
Helpful to share protocols under which samples were collected and guidance on sharing

Names (with Consortia) on the team:

What additional expertise do you need?

Our next meeting is?

5.5: How to access existing (development/pediatric) data?

Problem we think we can solve by collaborating –

- Identify all sources of developmental and pediatric data and capture them in a shared registry

Recommendation(s)/ What we want to do -

- E.g. Coggle mind maps to capture papers and contact people associated with each paper and data
- Need to identify champions and people who really care about organizing the data.
    - Could we integrate trainees into this activity, appropriately acknowledged (e.g. review paper)?
        - Action item: ask within labs and local communities to find volunteers for this activity
- Should standardize the data that is captured.
- Would ideally need some central funding to actively collect and organize all the papers - to find curators.
- Funders: Could HCA central help with this?

Names (with Consortia) on the team:

- Moderator(s): Muzz Hannifa, Alain Chedotal, Deanne Taylor, Jonah Cool
- Notetaker(s): Kerstin Meyer, Jonah Cool
- Other Group Members and Consortia each works with: Gary Bader, Marc Charette, Gloria Pryhuber
- At least one Email contact:

What additional expertise do you need? None

Our next meeting is? None

---

6. Common Coordinate Frameworks

6.1: CCF User Interfaces & Anatomical Structures and Cell Types (ASCT) Tables

See introductory slides for this subtopic here.

Problem we think we can solve by collaborating –

1. We are interested to see if we can agree on Anatomical Structures and Cell Type (ASCT) tables for 5-10 organs, see initial set at https://tinyurl.com/ASCT10x10? In the UK, this table is also called "Pathology and Cell Ontology Table". Original table was published for kidney, see Table 5 in https://www.biorxiv.org/content/10.1101/828665v1.
2. What CCF approaches/solutions are being tested in various tissues/Consortia?
3. What maps of the vasculature exist in different organs? Which consortia are mapping vascular pathways within their organ(s)? Are others exploring using the Vasculature as a Coordinate System to Map All the Cells in the Human Body?

Recommendation(s)/ What we want to do -

- We will have a **ZOOM debrief** on or around April 9.

Names (with Consortia) on the team:

- Moderator(s): Jim Gee (BICCN), Katy Borner (HuBMAP)
- Notetaker(s): Lisel Record - Thank you!
- Other Group Members and Consortia each works with: Zorina Galis (HuBMAP), Marc Charette (HuBMAP, NHLBI), Sarah Teichmann (HCA: Helmsley, Wellcome, MRC, H2020 and CZI Seed Networks), **Please see listing at the end of this document.**

What additional expertise do you need?

- Anatomists, pathologists, radiologists.

Our next meeting is?

- We plan to have a 1h **ZOOM debrief** on April 9 between 10a-2p ET. If you would like to join, please share your name and email.

6.2: Common coordinate frameworks and computational physiology.

Problem we think we can solve by collaborating –

- Comparing data from multiple species. CCFs provide material coordinates that give a common fiducial point in different species.
- Dealing with the dynamic changes of an organ (beating hearts, breathing lungs, motile colon, etc).
- Dealing with tissue growth - relating changes at cell level (size, shape, orientation) to changes at the tissue level (changing shape as obtained from microCT for example).

Recommendation(s)/ What we want to do -

- Create a list of all of the (known) datasets available (ie. Vanderbilt, Children's Hospital Philadelphia, Penn too.) Healthy and abnormal organ images (Hearts, etc.)
    - List of institutions and their associated patient images (MRI, CT, etc.) and clinical metadata (age, sex, health status, important genetic mutations etc.)
    - Fiducial points - Brain locations in peds
- Make the tools and data (images) accessible for analyses.
- Agree on a standard CCF across the CFDE. Agree on standard metadata framework.
- Link Human Cell Atlas (HCA) to physiology via the CCF.

Names (with Consortia) on the team:

- Moderator(s): Peter Hunter
- Notetaker(s): Tim Tickle
- Other Group Members and Consortia each works with: Ellen Quardokus HuBMAP, Aaron Horning HuBMAP, HTAN, Deanne Taylor Kid's First, Mark Coles HCA, Lucy Hsu, NHLBI/NIH

At least one Email contact:

What additional expertise do you need? None

Our next meeting is?TBD

7. Metadata - Schemas & Ontologies

7.1: Defining and socializing meta-data standards and tools for cross-consortia collaboration

Problem we think we can solve by collaborating –

- *Raising awareness of existing resources and avoid reinventing incompatible wheels:*
- recommending best practices for data models and standards (re)use including
    - the various parts of the ecosystem surrounding metadata (e.g. submission, validation, querying, data portals etc.)
    - available data models and standards serving different use cases: both on "supply (data generator) and demand (end-user data consumer)"-sides; examples: provenance (supply-side) and cohort building (demand-side); across different data types (e.g. clinical data, biospecimen, popular assays like scRNA-seq)
    - existing tool implementations that are compatible with a) and b)
- incentivize consortia/projects to use and contribute to these best-practices when applicable (or indicate why these practices are not applicable to their use cases) - develop publication/funding requirements; introduce other "business drivers"
- *Lower the burden of working with metadata and data submission:*
    - develop/reuse tools and interfaces that make it **easier** for scientists to do the right thing (e.g. be ready, via suitable tools, to accept how their data is now w/o additional burden of reformatting; make metadata submission fun)
        - note: UIs and tools are not developed in vacuum separated from ontologies and schemas, but are 1) driven by the same use cases; 2) have inputs that are determined by ontologies and schemas; 3) can be generated based on data models. Hence common data models/standards may imply at least common features in tooling, if not right out implementations reuse
    - facilitate developing common semantics for data generators and consumers
    - develop process for gathering inputs from data generator and data consumer experts when developing new *data models* tailored to a specific consortium; the process should provide sufficient infrastructure and flexibility to enable reuse of existing standards and ontologies (e.g. mixing and matching/knitting together portions of ontologies; mechanisms referring to data dictionaries in existing standards; mapping across overlapping data dictionaries from different standards) as well as capture

the motivations of experts in relation to metadata (i.e. use cases addressed by the metadata/data model/standard); that would facilitate the tools-related work above

Recommendation(s)/ What we want to do -

- Work on best-practices resource for data models and standards (re)use; and metadata related tools
    - evaluate standards and related tools that already exist in the context of the problems above
    - decide which parts of them can be tailored for our specific communities (i.e. atlases); are amenable to interoperation and provide examples (e.g. CDISC has standards for study protocol, data collection, data aggregation, analysis; HL7 FHIR; Biospecimens data; CDISC Controlled Terminology; ISA; PFB; ISO11179, JSON-LD, etc.)
    - Engage in development of relevant HL7 FHIR "Resources" (Research Study (maturity level 0); Research Subject (level 0); Specimen (diagnostic) (level 2).
    - provide examples from organizations/consortia that have gone through the process of adopting standards, data models and tools in the context above, not only in NCI/HCA but other NIH institutes.

Names (with Consortia) on the team:

- Moderator(s): Milen Nikolov (HTAN)
- Notetaker(s): Jeremy Miller (BICCN), Melissa Cook (HTAN/NCI)
- Other Group Members and Consortia each works with: Matt Wyczalkowski (HTAN, CPTAC3), Ian Fore (NCI, HTAN), Jason Hilton (Lattice), Anna Maria Masci (Duke LungMap phase I), Enrique Sapena Ventura (HCA)

At least one Email contact:

What additional expertise do you need?

- Examples/connections from organizations/consortia that have gone through the process of adopting standards, data models and tools in the context above, not only in NCI/HCA but other NIH institutes

Our next meeting is?

- We have not coordinated that; however, we pointed out an existing recurring cross-consortia meeting that can be used as a venue for next steps: Cell Atlas Data Curation WG. Can get in touch via Milen Nikolov (HTAN/Sage Bionetworks), Laura Clark (HCA/EMBI), Nils Gehlenborg

7.2: Playbook for Setting, Finding, and Integrating with Standards

Problem we think we can solve by collaborating –

- HMW set a playbook for setting standards, finding them, integrating with them

Recommendation(s)/ What we want to do -

- NIH consortia join a workshop on finding and sharing standards
    - could bring people together with CEDAR and Carol at RTI, goFAIR initiative, whomever else, a few talks and a brainstorming on how to leverage some of these tools to search and share data standards
    - Could join with M4Ms - metadata for machines workshops - by goFAIR
    - Could do retrospectives on what's worked well and what hasn't in different consortia
    - Goal of Meeting: To articulate mechanisms by which consortia of investigators can learn from one anothers metadata contributions to understand what standards are in use and to work towards standards that could be applicable across a wider community.
    - When has a consortia found they needed to rework their metadata choices and how did you do it?
- Testing technology that makes it easy for users to adhere to standards that are created, help people submit to different consortia
    - Testing CEDAR with IDR and/or NeMO/BCDC for example
    - Schedule a CEDAR demo, invite audience across this conference
    - Follow up meeting on how using it/the testing goes would be an option

- **Names (with Consortia) on the team:**
- Moderator(s): Carol Thompson (BICCN)
- Notetaker(s): Kylee Degatano (HCA, BRAIN)
- Other Group Members and Consortia each works with: Heather Creasy, (NeMO, CFDE), Jason Swedlow (OME), Ben Hitz (ENCODE), Mark Musen , Ajay Pillai (HuBMAP), Ingrid Youngworth (ENCODE)

What additional expertise do you need?

- NIH program officer support for workshop.
- NIH support for online webinar(s) on CEDAR and other metadata standards tools.
- Representatives of other consortia to participate in testing using CEDAR or other tools.
- Need list of consortia from meeting organizers.

Our next meeting is? In one month.

7.3: Anatomical Ontology Reference Atlas Model

Problem we think we can solve by collaborating –

- Standardize means in a FAIR way of capturing different components of anatomical atlases, e.g., coordinate system, reference data, terminology and delineations in a machine readable form so that different versions of an atlas (spatial or cellular) can be compared.

Recommendation(s)/ What we want to do -

- Build on work done in INCF on brain and apply to other organs (see image below and slide deck to determine whether this model works for other organ systems: https://docs.google.com/presentation/d/1qc-y0klgdRrNrlzjjiAxsm7bF2FMoUa-i4h8leYirHg/edit#slide=id.p)
- Also on work in Uberon aligning to Allen atlases

Names (with Consortia) on the team:

- Moderator(s): Chris Mungall
- Notetaker(s): Ajay Pillai, Maryann Martone
- Other Group Members and Consortia each works with: Ajay Pillai (HuBMAP)

At least one Email contact: Maryann Martone (BRAIN)

What additional expertise do you need?

- Additional consortia members who are developing CCF's

Our next meeting is? 1 month

7.4: Deriving ontological relationships from CCFs

Problem we think we can solve by collaborating –

- Current anatomy ontologies have textual spatial descriptors (e.g. this region borders region X, is a part of region Y, surrounds region Z), but these are often not encoded in a computable way, and it is hard to reason over these or validate these.

Recommendation(s)/ What we want to do -

- Use CCFs, atlases, and spatial information to derive ontological relationships and validate these.
- Use BSPO to describe these spatial relations: http://obofoundry.org/ontology/bspo (e.g. anterior-to, superficial-to), as well as RO http://obofoundry.org/ontology/ro (tributary of, branches-from).

Names (with Consortia) on the team:

- Moderator(s): Chris Mungall
- Notetaker(s): Ajay Pillai, Maryann Martone
- Other Group Members and Consortia each works with: Ajay Pillai (HuBMAP)

At least one Email contact: Chris Mungall
What additional expertise do you need? None

Our next meeting is? None

7.5: Cell Ontologies

Problem we think we can solve by collaborating –

- Collaboratively building a cell ontology that is fit for purpose for annotating and querying across single cell transcriptomic data.

Recommendation(s)/ What we want to do -

- Core funding is needed for basic work on the cell ontologies: regular releases, basic housekeeping, training & outreach.
    - Possibility to provide money through individual organ focussed groups - in a time-frame suitable for ontology developing groups (e.g CL; Uberon) to get resources in place to respond.
- We need working groups to audit current ontologies (cell ontology - others?) for coverage in particular domain areas (organ systems). This should be facilitated by outreach. These should include pilot projects to quickly show return on investment. **Aim**: Set up organ-specific groups to review and contribute to the cell ontology.
- Reach out to existing annotation standardization efforts, e.g. Peter Kharchenko's cell-type annotation platform: https://github.com/hms-dbmi/cap-example/blob/master/model.md
- Outreach (and better doc) needed to train the community to contribute directly via pull requests to the cell ontology.
    - Open up cell ontology meetings and widely publicise them.
    - Set up open training sessions

    -
- Standards required for cell type annotation - clearly recording provenance and evidence - in a form that allows for easy sharing of annotated data (e.g. in Loom). Can this be folded into a broader standards working group? - bring this up in cell annotation breakout?
- Unmet need: Data driven cell type specification. Potential working group.

Names (with Consortia) on the team:

- Moderator(s): David Osumi-Sutherland

What additional expertise do you need? None

Our next meeting is? None

---

8. Multiplex Molecular Profiling Analysis

8.1: How to leverage multi-modality profiles to better define cell types and states?

Problem we think we can solve by collaborating –

- How to define cell types when you get high-resolution data, bringing it into biologically determinable domains?

- **Recommendation(s)/ What we want to do -**
- Form a team
- Need to make some operational definition of terms
- Reach out to consortia
- Need expertise in experimental, computational, biology,
- Potential members:
    - Meng Wang (Snyder), Anshul & Greenleaf (nominated by Snyder), Rob Tibshiruni, Joe Ecker will identify, Bing Ren will nominate a lab member to join too, Sebastian Preissl, Benedict Anchang, Dana Pe'er, Potential biological system, HSC

Names (with Consortia) on the team:

- Moderator(s): Bing Ren
- Notetaker(s): Dena Procaccini
- Other Group Members and Consortia each works with:
- At least one Email contact: Benedict Anchang

What additional expertise do you need? None

Our next meeting is? None

8.2: How to make your containers FAIR

Problem we think we can solve by collaborating –

- Find and search for containers and improve their standardization (inputs and outputs) description
- How to create standards and best practices for quality assurance, container construction

- How reproducible are the containers and how do they behave on different datasets

Recommendation(s)/ What we want to do -

- Containers and workflow coordination across consortia
- Approach other consortia groups on standardization (e.g., data, containers)
- Decide and build containers that include data

Names (with Consortia) on the team:

- Moderator(s): Richard Conroy
- Notetaker(s): Richard Conroy, Satra
- Other Group Members and Consortia each works with:
- At least one Email contact: Bill Shirey, Matt Ruffalo

What additional expertise do you need?

- UC SantaCruz
- Galaxy + CWL + GA4GH
- ReproNim

Our next meeting is? None

8.3: What pipelines exist across all consortia, and how can they feed into development of new multiplex analysis pipelines?

Problem we think we can solve by collaborating –

- Across consortia, there are many different pipelines for generating data from a single data type modality (e.g. scRNA-Seq, scATAC-Seq, in situ). How can we learn what pipelines exist across all consortia, how they are run, and what output data they create? Next, how can we create pipelines that read in and integrate those data types for a multiplex analysis?

Recommendation(s)/ What we want to do -

- We need representatives from each consortia who know the respective analysis pipelines and have time to contribute. They will add what their pipelines do and links to code in a shared google sheet organized by consortia. We expect to find pipelines for dealing with both single and multiple data modalities.
- We want pipelines that can integrate the different data modalities across consortia. We need the data model (how data is stored and metadata) for each data type across each consortia, and we need to converge on data formats and metadata. What common metadata across data types are needed for a multiplex analysis?

- We should write multiplex analysis pipelines using workflow languages (WDL, CWL). These are used currently by most members of this group when running pipelines.

Names (with Consortia) on the team:

- Moderator(s): Orr Ashenberg Notetaker(s): Rajasree Menon, Muzlifah Haniffa, Matthew Ruffalo
- Other Group Members and Consortia each works with: Anup Mahurkar (BRAIN/BICCN), Rajasree Menon (KPMP), Idan Gabdank (Encode consortia)

At least one Email contact:
What additional expertise do you need?

- We need email addresses for members with sufficient interest and representation from the different consortia in order to start this.

Our next meeting is? None

---

9. Spatial Profiling Analysis

9.1: Create a challenge for cell type and neighborhood identification across highly multiplexed image technologies

Problem we think we can solve by collaborating –

- Combine all available technologies and computational methods across dozens of research groups and find the most successful combination of computational approaches for specific biological questions.

Recommendation(s)/ What we want to do -

- Manual ground-truth annotation by domain experts (supported by H&E, IHC and highly multiplexed data). We need data sets from a variety of technologies. There are already several ongoing efforts at HTAN/HuBMAP to create those data sets.
    - What technologies? What antibody panels? Which tissues?
- We could start with a Sandbox for the community and move to a wider audience (to include pre-processing, processing, spatial analysis, neighbourhoods, cell type and cell state calling, cell morphology).
    - Emma Lundberg suggested using imjoy.io or/and bioimage.io as a template for a highly multiplexed imaging sandbox.
- Use a common database for data storage (and visualization)
    - Peter Sorger suggested to use HTAN/HuBMAP data set as an example for storage/visualization
- Run a Hackathon / Sprint to build a basic infrastructure (to be coordinated by HuBMAP HIVE+HTAN)
- HTAN and HuBMAP datasets can be used to kickstart the process.

Names (with Consortia) on the team:

- Moderator(s): Denis Schapiro (HTAN)
- Notetaker(s): Zoltan Maliga (HTAN) and Sinem Saka (HuBMAP)
- Other Group Members and Consortia each works with: Ron Germain (NIH), Anup Sood (HuBMAP, GE), Qin Ma (HuBMAP), Peter Sorger (HTAN), Neil Kelleher (HuBMAP), Emma Lundberg (HPA,Swedish Consortium)

At least one Email contact:

What additional expertise do you need? None

Our next meeting is? None

9.2: *Spatial description of cell type patterns can reveal biological function*

Problem we think we can solve by collaborating –

- Cell types exhibit a variety of spatial patterns (e.g. gradients, repeating units, layers) whose knowledge is essential to understand biological function. We propose to leverage spatial transcriptomics data to capture and classify spatial patterns of cell types. In particular, we agreed that we need to develop mathematical metrics to quantitatively describe the pattern properties, as most patterns might not be visible "at a naked eye". For instance: what is the granularity of a particular cell type? Does cell type X tend to be a neighbor of cell type Y? Such questions can be quantitatively addressed by mathematical indicators which we propose to collaboratively develop.

Recommendation(s)/ What we want to do -

- We agreed that the amount of available data is a substantial limitation, and we plan to compute the data amounts needed for each method, in order to facilitate methods comparison and evaluation. We agreed that different fields (such as physics or ecology) already partially tackled the problem of spatial characterization, so that the best initial course of action could be to test strategies already adopted in different fields. We agreed that histopathology should be used in conjunction with ST data to extract properties of patterns. We agreed that including cell morphology would be ideal whether possible.

- **Names (with Consortia) on the team:**
- Moderator(s): Tommaso Biancalani (BRAIN, HubMAP, CZI-HCA)
- Notetaker(s): Alex Ropelewski (BRAIN/BIL, HuBMAP)

At least one Email contact:
What additional expertise do you need?

- Expert in spatial characterization of patterns from different fields (e.g. physics, ecology, weather forecasting, geospatial atlases) would greatly help the proposed research.

Our next meeting is? We did not schedule a next meeting.

9.3: Generation of 3D reference volumes and frameworks

Problem we think we can solve by collaborating –

- Integration of spatial information into transcriptomics, research on position-specific effects (diseases), comparability of sample types (organ-specific),

- Development of a coordinate system, which acknowledges the individual specificity of volume, shape and architecture of organs and their subunits
- Building environmental maps of organs and its units
- Providing access to whole human organs and its subunits for developing workflows and for creating first results

Recommendation(s)/ What we want to do -

- standardized workflows, integration of transcriptomic data sets into imaging-based coordinate frameworks, use of standardized organ-spec. reference landmarks

Names (with Consortia) on the team:

- Moderator(s): Anna Nele Herdina (HCA, Liver Cell Atlas)
- Notetaker(s): Sabrina Summer (HCA, Liver Cell Atlas)
- Other Group Members and Consortia each works with: Wolfgang J Weninger (HCA, Liver Cell Atlas), Stefan H Geyer (HCA, Liver Cell Atlas), Lukas F Reissig (HCA, Liver Cell Atlas)

At least one Email contact:

What additional expertise do you need?

- data integration specialist(s), spatial transcriptomics (FISH, RNAscope)

Our next meeting is? None

9.4: Linking spatially targeted MS to specific cell types through IF and spatial transcriptomics.

Problem we think we can solve by collaborating –

- Identifying the proper test system to develop methods… what tissue, cell types, antibodies?
- Correlating transcript abundance with molecular abundance.

Recommendation(s)/ What we want to do -

- Determine tissue/cell types to start with.
- Collect multimodal molecular imaging data on serial kidney tissue sections.
- See if we can compare specificity/granularity in cell type definitions that are abel to be determined with various modalities/molecular classes.

Names (with Consortia) on the team:

- Moderator(s): Jeff Spraggins

- Notetaker(s): Elizabeth Neumann
- Other Group Members and Consortia each works with: HuBMAP and HCA

At least one Email contact:

What additional expertise do you need?

- Work with biologists to determine the most clinically relevant cell types to perform initial experiments. In this case, for the kidney, what are the most important 3 cell types to target.

Our next meeting is? April 13th.

---

10. Tissue Collection & Processing

10.1: The importance of tissue freshness for optimizing tissue interrogation

Problem we think we can solve by collaborating –

- Tissue freshness - single center versus shipped tissue. When and how is autopsy tissue useful either as a control or for studying disease. Sample collection: time spent until preservation and how this impacts assay quality. How does this affect tissue sharing? How to separate out tissue quality heterogeneity from assay quality and heterogeneity?

Recommendation(s)/ What we want to do -

- Guideline for standardization - tissue specific and disease specific. Type of biospecimen specific. Emphasize the importance of documenting clinical and biospecimen metadata. Metadata must take into account where, when, tissue coordinates, and under what circumstances tissue is collected and processed.
- Establish standardized tables across various Consortia to prioritize metadata elements. Make an app that others could use.
- Establish a hotline/chat channel for different researchers to communicate about common everyday problems. This can be connected to and included in the app. For example a dedicated Slack channel.
- Establish cross cutting training for post-docs and others working in this area.

Names (with Consortia) on the team:

- Moderator(s):Jonathan Himmelfarb (Kidney Precision Medicine Project)
- Notetaker(s): Jiyeon Choi (NCI, Lung cancer, Lung single-cell eQTL in Asian population)
- Other Group Members and Consortia each works with: Anna Nele Herdina (HCA, Liver Cell Atlas)Nasreen Haque (Atherosclerosis,NYMC), Lori Coburn (Vanderbilt, Gut Cell Atlas), Yu-Hua Tseng (Joslin Diabetes Center, CZI-HCA Adipose tissue network)

At least one Email contact:
What additional expertise do you need? None

Our next meeting is? None

10.2: Challenges with informed consent, and patient potentially identifiable data

Problem we think we can solve by collaborating –

- "Is single-cell transcriptome data, for example, de-identified non-human subject data?" This could pose a challenge in sharing individual-level raw data with the community and limit the data usage going forward. Planning ahead for putting things in consent beforehand is challenging.

Recommendation(s)/ What we want to do -

- Draft guidelines and/or templates for data sharing across centers applicable to most research studies and consortia that would be considered acceptable by IRBs in general.
- Create a chat room or forum for diverse investigators to address challenges that they are facing and to describe best practices.
- Educational materials for clarity on what is meant by consent and what are the implications of different languages in the consent process for downstream use of data.
    - Is it possible for the NIH to take the lead across all Institutes in designing a centralized effort to define what is identifiable data? Can the NIH work with the CTSA program and single IRBs to come up with universally acceptable templated language for data use agreements? Can the NIH work with the European Union and other stakeholders to harmonize data sharing efforts?

Names (with Consortia) on the team:

- Moderator(s): Jonathan Himmelfarb (Kidney Precision Medicine Project)
- Notetaker(s):Jiyeon Choi (NCI, Lung cancer, Lung single-cell eQTL in Asian population)
- Other Group Members and Consortia each works with:Lori Coburn (Vanderbilt, Gut Cell Atlas), Yu-Hua Tseng (Joslin Diabetes Center, CZI-HCA Adipose tissue network), Anna Nele Herdina (HCA, Liver Cell Atlas),Nasreen Haque (Atherosclerosis,NYMC)

At least one Email contact:
What additional expertise do you need? None

Our next meeting is? None

10.3: Impact of processing on tissue viability and integrity

Problem we think we can solve by collaborating –

- How is the time to preservation and processing impacts your technology (when the samples is fresh/ frozen/ fixed/ in organ transplant buffers)?
- HMW define how tissue sources (biopsies, resection, autopsy, liquid samples) impact processing?
- HMW select optimal temperature, time and conditions to preserve cell states at i*n vivo* states? HMW optimize time of freezing / preservation of material (i.e. tissue=> nuclei)?

Recommendation(s)/ What we want to do -

- Determine the difference between viability and integrity of tissue when assessing tissue transferring from the subject to the lab.
- Recommendation: benchmarking of tissue procurement protocol against reference data sets to identify procurement artefacts, i.e. stress response genes as one measure alterations in processing.
- Using databases from HCA and other networks for searching for methodological challenges addressed in different organ systems /cell types. Ideally, methods that were deposited were also indexed: annotated with metadata of protocols defining specific uses, publications, and data / performance characteristics of methods in standardized manner.

Names (with Consortia) on the team:

- Moderator(s): Asaf Rotem
- Notetaker(s): Matthias Kretzler
- Other Group Members and Consortia each works with:

At least one Email contact:

What additional expertise do you need?

- Meta-standard tagging processing elements in the protocol.io submitted experimental strategies. This might be a new feature in Protocols.io and could help find the best protocol for isolation of specific cell types from a specific tissue, with an expected outcome.

Our next meeting is?

- Check-in in a month to evaluate implementation progress.

10.4: Tissue acquisition and distribution strategies

Problem we think we can solve by collaborating –

- HMW balance complexity of sample processing at core sites versus on site analysis?
- HMW efficiently allocate tissues in procured material ?

Recommendation(s)/ What we want to do -

- Recognize tradeoff between tissue processing on site versus development of technologies towards a specific preservation approach.
- How to design your tissue procurement strategy? Coordinate strategy via round table strategy of all relevant shareholders, utilize iterative strategy by deploying close feedback loops of QC to indicate what approach delivers for what application.
- Use case studies in specific organ systems: then assess how this was implemented successfully in a disease - in an organ specific manner.
- Using Protocols.io to index emerging or established network protocols for tissue procurement as a resource for a specific project / organ system / disease.

Names (with Consortia) on the team:

- Moderator(s): Asaf Rotem
- Notetaker(s): Matthias Kretzler
- Other Group Members and Consortia each works with:

At least one Email contact:

What additional expertise do you need?

- Interdisciplinary team essential. Critical: pathologists, interventional radiologists.

Our next meeting is?

- Check-in in a month to evaluate implementation progress

10.5: Strategies to retrieve difficult cell types

Problem we think we can solve by collaborating –

- HMW process and preserve tissue to capture “difficult’ cell types?
- HMW monitor tissue processing in relationship to the tissue chain of custody and associated metadata?
- HMW processing targeting cell types: epithelium, mesenchymal, inflammatory cell types, etc.?
- HMW prioritize processing technologies to enable optimal preservation of a specific cell type?

Recommendation(s)/ What we want to do -

- Define the target tissue and the purpose of tissue acquisition for your study.
- Understand the interstitial compartment - cell matrix interactions which need to be overcome to release cells from tissue.
- Understand how your tissue geometry impacts your isolation procedures, solid tissue requiring mechanical dissociations versus biopsy tissue open to enzymatic dissociations.
- Cross-organ assessment of optimal strategy to release cells from tissue based on the degree of extracellular matrix - fibrosis.
- Address how to process difficult cells to capture: fat cells showing mechanical and biochemical properties impacting isolation and analytical pipelines, differentiated epithelial cells and syncytial cells (myofibroblasts): whole single cell analysis versus single nuclei strategy to consider for applications.
  Establish a metric to indicate unique challenges to specific cell types and optimal study protocols for cell based assays.

Names (with Consortia) on the team:

- Moderator(s):Asaf Rotem
- Notetaker(s):Matthias Kretzler
- Other Group Members and Consortia each works with:

At least one Email contact:

What additional expertise do you need?

- Domain expertise in cell biology of cell types of interest, consider cross organ lessons to be learned and shared: index this knowledge in protocol.io
- Matrix biologists expertise helpful in this context.

Our next meeting is?

- Check-in in a month to evaluate implementation progress

10.6: Metadata acquisition

Problem we think we can solve by collaborating –

- What is the minimum metadata needed for clinical parameters
- What is the minimum metadata needed for tissue procurement parameters
- For which tissues histopathology is informative and what are the features that are helpful?
- Tracking shipping and what is optimal storage temperature/condition?
- How do we recognize artifacts?
- How do we trace sources of artifacts?
- Communication?

Recommendation(s)/ What we want to do -

- Those with expertise in same organs should start comparing their lists of metadata elements to prioritize “must have” and “should have” type data.
- Initially need to document extensively how tissue is collected and preserved.
- Share same tissue across sites to check robustness of QC measures across technologies.
- Establish strong communication and organization of working groups, establish databases and distributing the knowledge gained

Names (with Consortia) on the team:

- Moderator(s):Sanjay Jain
- Notetaker(s): Jennifer, Sanjay, Gloria
- Other Group Members and Consortia each works with: Sanjay Jain (HuBMAP, KPMP)

At least one Email contact Alexandre Denadai-Souza
Jennifer Zamanian (CZI Seed Network DCC)
What additional expertise do you need?

- Communicator and organizer to bring the consortia together
- Communication tools

Our next meeting is? None

10.7: Sharing protocols

Problem we think we can solve by collaborating -

- What methods work across tissue and across consortia?
- What does not work?
- What are the common genes associated with artifacts across consortia?
- Report outliers during OMICs analysis that can be traced back to tissue metadata

Recommendation(s)/ What we want to do -

- Need a platform where consortia can interact and post success and failures (e.g. protocols.io, github…)
- Engage stakeholders to deposit the gene lists or metric that are used for identifying stress artifacts
- Establish strong communication and organization of working groups, establish databases and distributing the knowledge gained

Names (with Consortia) on the team:

- Moderator(s): Sanjay Jain
- Notetaker(s): Gloria, Jennifer, Sanjay
- Other Group Members and Consortia each works with:sanjay jain (HuBMAP, KPMP)

At least one Email contact:

What additional expertise do you need?
- Communicator and organizer to bring the consortia together
- Communication tools

Our next meeting is? None

---

11. Affinity Reagent Development and Standards

11.1: Raising Efficiency in Antibody-based Imaging: Pre-analytical Variables and Validation

Problem we think we can solve by collaborating –

- Create a conceptual framework and prioritized battle plan to raise the operational baseline of Ab-based imaging technologies.

Recommendation(s)/ What we want to do -

Create a Working Group with the goal of publishing a White Paper* with this Outline

1. Target Selection (Neil / Kwanghun) & Variability (Stephen Hewitt) (Andrea Radtke, Elizabeth Neumann, Jeff Spraggins)
2. Renewable Reagents (Ananda Roy, Pothur Srinivas)
3. Modified Reagents (Conjugation, etc.) (Sinem)
4. Reagent Validation (Anup / Neil - ask Emma Lundberg)
5. Reproducibility (Stephen, riff on Ab-validation, experimental replication; closed platforms)

*An invitation from *Nature Methods has* been issued.

Names (with Consortia) on the team:

- Moderator(s): Neil Kelleher (HuBMAP), Stephen Hewitt (HuBMAP), Kyunghun Chung (BICCN)
- Notetaker(s): Elizabeth Neumann and Jeff Spraggins
- Other Group Members and Consortia each works with: HuBMAP, BRAIN Initiative, and HCA; Anup Sood, Andrea Radtke (Ronald Germain/NIH)

At least one Email contact: (point-of-contact)

What additional expertise do you need?

- Emma Lundberg from HPA: Peter Sorger (with Sinem)
- Garry Nolan (CODEX) or Sean Bendall (MIBI)
- Someone from Ron Germain's Group (Andrea Radtke)
- Vendors AbCAM (rep =?), CST (Jonathan Bock), Thermo Pierce (rep=?).
- Neil to ask Jeanne Camarillo (HuBMAP Ab-guru) to help:

Our next meeting is?

- Neil to set two dates for two ZOOM calls in next 6 weeks. Goal: super solid paper outline by ~May 20th, 2020. Standing ZOOM Meeting ID will be: https://northwestern.zoom.us/j/4484282346
- Admin. Assistant for these ZOOM Meetings will be Ms. Shari Bratanch:

12. Atlas Integration

12.1: Collaboration to develop data-driven quantitative ontologies across cell types and tissue space

Problem we think we can solve by collaborating –

- Develop data-driven quantitative ontologies that include cell type and spatial information
- Organ specific Jamborees to improve existing multiscale anatomy ontologies (Uberon) and make tailored views that are accessible for working biologists.
- Impact of technology used to obtain information
    - Can we develop a uniform framework across modalities

Recommendation(s)/ What we want to do -

- organ site-specific jamborees to improve existing multiscale anatomy ontologies (uberon) and make them accessible to biologists in the organ specific groups.
    - Focus on deviation from normal, e.g. in cancer
- Start with introductory event, then split out into organ specific sessions with report backs, can be tied in with existing working groups?
    - Identify most advanced / defined ontologies and representatives
    - Need a common framework ontology such as Uberon as basis for discussions
    - Led by those furthest along, both healthy and tumor
    - **Planning Committee: Andrew Adey, Maryann Martone, David Osumi-Sutherland,**
        - Involve but go beyond efforts from CCF, e.g. HuBMAP, Katy Borner, HCA, Allen
        - Involve funding agencies, develop concepts for funding needs
        - Code based system, e.g. William Mondy
- Series of meetings? Start of with case studies, lessons from Drosophila field, Allen brain atlas
    - Talking points?:
        - Use known cell cell associations, signatures that may change during malignant transformation
        - Cell type definition and disease-associated changes - ensuring framework for incorporating “deviation from normal”
        - Hierarchical infrastructure to define cell type-specific digression from norm
            - Top level should be definable across platforms and modalities

        - Challenge of data incorporation: directly data driven, vs using data as evidence to back up existing or new ontology structure
    - Deviation from normal in terms of location, how does location define (cancer) cell behavior, response heterogeneity
    - Consider how to best measure “distance” from the norm
    - How to integrate across scales
    - Computational formalisms rather than visual
    - Data-driven quantitative ontologies that include cell type and spatial information

Names (with Consortia) on the team:

- Moderator(s): Joe Gray
- Notetaker(s): Philipp Oberdoerffer
- Other Group Members and Consortia each works with: Maryann Martone (BICCN-BRAIN-SPARC) Ajay Pillai (HuBMAP) Aaron Horning (HTAN, HuBMAP) ; David Osumi-Sutherland (EMBL/EBI) Jessica Langer (GCA); Andrew Adey (OHSU, HTAN, BRAIN) Joe Gray (HTAN) William L. Mondy, Zorina Galis (HuBMAP,) Katy Borner (HuBMAP)

What additional expertise do you need?

- Pathologists, anatomists, organ experts

Our next meeting is? None

12.2: Cross-atlas cell state curation and mapping to tissue ecosystems

Problem we think we can solve by collaborating –

- Develop approaches for curating cell **state** (proliferative, inflammatory, fibrotic, EMT, etc.) across atlases and to **map** cells within their ecosystem / tissue architectures

Recommendation(s)/ What we want to do -

- Curation of cell states across tissue contexts (beginning with inflammatory cell state); must capture terminology and sample/cell provenance
- Delineate cell topography within distinct tissue architectures based on multiplex, in situ proteomics/transcriptomics coupled with expert pathological annotation (incorporate histology/vasculature)
- Long term goal: Cross-tissue/atlas database of cell states will facilitate standardization, integration with multi-modal omic measurements and computational modeling

Names (with Consortia) on the team:

- Moderator(s): Christina Curtis
- Notetaker(s): Sean Hanlon, Fiona Ginty
- Other Group Members and Consortia each works with:
- At least one Email contact:

What additional expertise do you need?

- Engagement with anatomic pathologists; this is happening within atlas efforts (by tissue site) but cross-atlas/tissues annotations and dedicated curation efforts are needed
- Ensure representation from immunology, developmental biology, cancer biology, organ site experts etc. so that views are represented (cellular context is key)
- Engage modelers early on; methods span machine learning to mechanistic models; considerations for defining data structures
- Interaction with Cell Annotation working group is essential; others that have developed ontologies

Our next meeting is?

- ~April/May 2020 (Peter Sorger / Christina Curtis to coordinate in collaboration with related subtopic groups emerging from this meeting)

12.3: Atlas related education and outreach activities

Problem we think we can solve by collaborating –

- Collaborate on producing short videos and presentations that highlight the various Atlas activities with the goal to disseminate knowledge, cross-pollinate and fuel collaborations

Recommendation(s)/ What we want to do -

- Develop 30 min didactic presentations followed by discussions
- Learn about topics from a multidisciplinary perspective (example: inflammatory cell states; deep learning for digital pathology)
- Flesh out further with the Outreach working group

Names (with Consortia) on the team:

- Moderator(s): Christina Curtis
- Notetaker(s): Sean Hanlon, Fiona Ginty

At least one Email contact:

What additional expertise do you need?

- Support from Atlas initiatives

Our next meeting is? None

12.4: Kidney Atlas Projects

Problem we think we can solve by collaborating –

- HMW ensure the different consortia (and different organs) are using the same vocabulary? Ensuring individuals from all networks are working together.
- HMW ensure data are interoperable? Using the same data formats?
- HMW emphasizes the need to make data accessible to larger research community members (not just data power users)?

Recommendation(s)/ What we want to do -

Cross Consortia activities:

- Sharing tissue across consortia (DUA/MTA/CDA master agreement from NIDDK)
- Nomenclature workshops KPMP, HCA, GUDMAP, RBK, HuBMAP => ongoing working group (contact)
- Significant challenge towards interoperability for data sharing data: processes to data generation and data releases and framework: coordination inside consortia and between consortia
- Correlating of data sets generated from reference tissue to map independent technologies on multiscalar data generation.
- Aspirational goal: Reference tissue utilized between consortia for calibration and standardization of data sets generated

Names (with Consortia) on the team:

- Attendees: Matthias Kretzler, Joshua Levin, Jonathan Himmelfarb, Bruce Herr, Becky Steck, Abhijit Naik, Ellen Quardokus, Deborah Hoshizaki, Sara Lin

At least one Email contact: Becky Steck

What additional expertise do you need?

- Interaction with FAIR WG (meeting this afternoon)

Our next meeting is? None

13. Data Modeling & Integration

13.1: How to deconvolute bulk omic data using single-cell data

Problem we think we can solve by collaborating –

- How to get single-cell information from bulk and vice versa.

Recommendation(s)/ What we want to do -

- Controls e.g. gold standards
- We could use the consortium to build reference data sets to test different algorithms.
- Understand define limitations
- Get recommendations on parameters where algorithms work

Names (with Consortia) on the team:

- Moderator(s): Mike Snyder, Bob Murphy
- Notetaker(s): Anup Mahurkar (BRAIN/BICCN)
- Other Group Members and Consortia each works with: François Aguet (GTEx/TOPMed/CPTAC)

What additional expertise do you need?

- Members from each consortium

Our next meeting is? Unknown.

13.2: Longitudinal Inference Studies

Problem we think we can solve by collaborating –

1) Challenges:
    a) Control experiments were missing in MoTrPAC. For instance, the circadian rhythm was a factor that was not considered
    b) Pseudo-time trajectories and what is the minimal data necessary for creating the trajectory
    c) Data visualization of this type of data
2) Algorithms:
    a) C means clustering
    b) Monocle can be used for building pseudo-time and use spatial data
    c) Use of cell shape for reverse interpolation

3) Is interpolation data-driven?
4) Data visualization
    a) Using averaging might lead to misleading trajectories. So it might be useful to use more granular trajectory analysis instead of dimensionality reduction to build these trajectories
    b) Can we build developmental time courses as trajectories based on these methods? We need to have enough samples to build these models.

Recommendation(s)/ What we want to do -

- Establish working committee
- Define minimal data that is released with the data
- Make sure multiple omics are collected

Names (with Consortia) on the team:

- Moderator(s): Mike Snyder, Bob Murphy
- Notetaker(s): (BRAIN/BICCN)
- Other Group Members and Consortia each works with: Kavya Sharman (HuBMAP), Denis Schapiro (HTAN)

What additional expertise do you need?

- Members from each consortium; outside experts

Our next meeting is? Unknown.

13.3: Multi-omic data integration

Problem we think we can solve by collaborating –

1) Integrate spatial data (imaging) with other Omics data--context of cells matters!
2) Two different single-cell assays (ATACseq vs. RNAseq and proteomics, metabolomics, microbiome)
3) Visualization

Recommendation(s)/ What we want to do -

- One general working group that might break into sub groups working groups to evaluate:
    1) Methods that work across modalities instead of pair-wise methods. Particularly across spatial and omic assays
    2) Outline the limitations of methods and assays
    3) Setup a separate visualization group
- Having gold standard datasets is very useful and need to be built, but how do you build gold standard datasets across modalities as it may not be possible.

Names (with Consortia) on the team:

- Moderator(s): Mike Snyder, Bob Murphy
- Notetaker(s): Anup Mahurkar (BRAIN/BICCN)
- Other Group Members and Consortia each works with: Kavya Sharman (HuBMAP), Denis Schapiro (HTAN)

What additional expertise do you need?

- Members from each consortium; outside experts

Our next meeting is? Unknown.

13.4: Expanding cell-cell interaction models to include long range signaling

Problem we think we can solve by collaborating –

- Predict what long range signals affect network behaviors in development, homeostasis and disease.
    - Can we model/predict the effect of systemic inflammatory factors (e.g. cytokines) on local signalling circuits in immune-mediated inflammatory diseases (such as inflammatory arthritis/IBD)?
    - Can we understand the effects of hormones on signalling networks within tissues, such as those that drive circadian rhythms (melatonin), metabolism and stress (cortisol)?
    - What can we learn about the ability of the CNS to modulate cell networks in peripheral tissues (and vice versa), e.g. via the vagus nerve in the gut/brain axis?
- Single-cell genomics may bring a molecular mechanism point of view to a physiological process.

Recommendation(s)/ What we want to do -

- Start a discussion with folks in the physiology/immunology communities to learn more about how they model these systems.
- Link to physiology modeling community e.g. ICSB http://systems-biology.org/conference/announcement/001258.html
- Identify people who are interested, maybe a satellite session at meetings
- Review article - compile a list of papers; please contribute!

Names (with Consortia) on the team:

- Moderator(s): Shannon Hughes (HTAN)
- Champion(s): Gary Bader (HCA);
- Notetaker(s): Dena Procaccini, Shannon Hughes

- Other Group Members and Consortia each works with: Steve Sansom (HCA/UK RACE)

What additional expertise do you need?

- Systems biologists, physiologists and immunologists!

Our next meeting is?

- A new Slack Channel on the HCA Slack: Cell-Ecosystems -- we'll start here
- To join HCA Slack: https://join-hca-slack.data.humancellatlas.org/ then join the cell-ecosystems channel.

13.5: How would we create an ontology describing the function of a group of cells e.g. the function of a cell/tissue gradient?

Problem we think we can solve by collaborating –

- How would one create a Gene Ontology for multi-cellular systems, how would you construct that?
    - How does one account for time in such a construct?
    - Or is the idea more about a structural model (i.e. hair follicle layers giving rise to a whole hair)
- How to create a organizational descriptive graph (DAG) that represents cells or "modules" that go together to describe tissue level structures;
    - A set of labels that can be applied to cell states that could be combined to describe a tissue
    - Cells types can be in multiple states concurrently that combine to create tissue level function
    - State = (signaling activation + cell fate/type)
    - The ontology is an extension of a network motif, but still need the model and a definition of the system

Recommendation(s)/ What we want to do -

- None

Names (with Consortia) on the team:

- Moderator(s): Shannon Hughes/Gary Bader
- Notetaker(s): Dena Procaccini/Shannon Hughes
- Other Group Members and Consortia each works with:

At least one Email contact: (Lung seed network)

What additional expertise do you need?

- Chris Mungal

Our next meeting is? None

13.6: Genetic basis of cellular identity

Problem we think we can solve by collaborating –

- Repository of existing genetic diversity studies and augmenting existing studies by a genetic diversity component.
- "Genetic meta analysis" within and across tissues, to redefine regulatory landscapes in health disease
- Establishment of reference single-cell genetic resources beyond RNA (ATAC/proteomics/spatial RNA?), ideally using reusable resources (cell lines, iPS, etc.)
- Using genetics as a tool to define cell identity vs. observational assays
- To model rare diseases and understand how genetic perturbation in organoids or animal models could affect cell state/type

Recommendation(s)/ What we want to do -

- Genetic variation component in data repositories
    - Recommendation to metadata working group to incorporate race & ethnicity and genetic data into ingest processes, or via record linkage with specialized repositories for controlled access genetic data (e.g. dbgap)
    - Foster the creation of genetic data resources by pooling data, starting by surveying existing efforts
    - Support structures to facilitate genetic studies, e.g. define protocols, process (including IRB approval), integration of open access (single-cell) and controlled access (genetics)
- Genetic meta analysis
    - Establish best practices and benchmarks for genotyping (germline & somatic) from single-cell assays as part of the HCA
    - Incorporate best practice genotype calling into DCP workflows, with and without bulk genotype reference, to deliver added value to existing studies
    - Initiate HCA/NIH working group to derive regulatory maps, within large tissues (e.g. blood), later across tissues
    - Define somatic mutation profiles & characteristics across human cell types & tissues
- Reference cell bank of single-cell multi omics profiling
    - Working group to define most suitable system for establishing standard reference(s) to benchmark single-cell genetics assays, e.g. cell line, iPS-derived, organoid

    - Create first open access community datasets with genotype, RNA, ATAC(?) to foster methods development and community uptake
    - Distribution center for access to samples & material, similar to HapMap in the ages of LCLs
- Genetics for cell identity.
    - Reachout to cell identity working groups to consider genetic regulation to define cell identify in the comparison to other modalities
- Single-cell technologies and models for genetic perturbations in disease
    - Community portal and protocol exchange
    - Normal variation panels for canonical systems that are used in rare disease genetics (e.g. iPS derived) to reduce entry barrier and increase power.

Names (with Consortia) on the team:

- Moderator(s): Oliver Stegle
- Notetaker(s): Jiyeon Choi

At least one Email contact:

- (single-cell genetics computational methodology)
- Deanne Taylor, (Pediatric Genetics Analyses, single-cell analyses)
- Ayellet Segre, (computational methods for genetic regulation and integration with GWAS)
- (NCI, Lung cancer)

What additional expertise do you need?

- Community support & coordination via Slack (#single-cell-genetics on HCA slack)

Our next meeting is? None

13.7: Cellular dynamics, plasticity, perturbations

Problem we think we can solve by collaborating –

- Meta-analysis of aging based on HCA data -> discover shared and specific signatures and pathways (requires adequate metadata).
- Incorporate single-cell “intelligence” into existing cohort studies of aging and age-related diseases (requires proactive collaboration/interaction/funding).
- Develop and evaluate better analysis methods tools for single-cell time series analysis (requires annotation and repositories of time series datasets)

Recommendation(s)/ What we want to do -

- Aging:

        - Recommendations to metadata working group for capturing aging-relevant metadata/information throughout HCA.
            - Laura Clarke- contact person for metadata community; HuBMAP: Ajay Pillai:
            - Recommendation to collect peripheral blood from HCA sample donors (where possible) and subject to DNA methylation profiling for epigenetic clock / biological age analysis.
        - Recommendation that aging studies should involve (where possible) a systemic perturbation / dynamic response. For instance, examining a cohort of individuals of varying age recovering from flu.
        - Recommendation to build upon ongoing large longitudinal cohorts (such as CARDIA) that capture aging-related metadata (age, cardiovascular health, lung health trajectories) and add single-cell assays to upcoming sampling.
- Time-series:
    - Recommendations to metadata working group for capturing time-series and perturbation-relevant metadata/information throughout HCA.
        - Laura Clarke - contact person for metadata community; metadata-community@data.humancellatlas.org
        - HuBMAP: Ajay Pillai:Review/perspectives paper on single-cell time series modeling (no volunteer identified to lead such an initiative, but several people would support/contribute)

Names (with Consortia) on the team:

- Moderator(s): Christoph Bock
- Notetaker(s): Peter Kharchenko, Alexander Misharin
- Other Group Members and Consortia each works with: Benedict Anchang NIH, Ziv Bar-Joseph, CMU, HuBMAP, Jim Hagood, HuBMAP and Normal Aging Lung Cell Atlas (NALCA), Ed Lein, Fabian Theis, Martijn Nawijn

What additional expertise do you need? None

Our next meeting is? None

14. Data QA/QC

14.1: Best practices and recommendations for Quality Assurance (QA) and Quality Control (QC)

Problem we think we can solve by collaborating –

- Enable researchers to have a uniform view of quality across consortiums for each data modality:
    - Facilitating integration studies across datasets/consortiums
    - Identify benchmarking datasets that can be used for algorithm development

Recommendation(s)/ What we want to do -

A) Immediate recommendations:
    a) scRNA/ATAC-seq
        i) Determine consensus on reference files used in preprocessing pipelines (Gene annotation, Genome build)
        ii) Determine “basic” QC metrics (e.g. total UMIs).
B) Develop consensus for each data modality across consortiums:
    a) Identify key person(s) or subworking groups in each consortium with knowledge of that data modality and associated metadata
    b) These people can inventory of QA/QC metrics used for that data type in that consortium. If inventories already exist, then it is simply a matter of sharing current working documents.
    c) Form a working group to develop a consensus:
        i) Which QC metrics should be considered “standard”
        ii) Document standards for how those QC metrics should be calculated
    d) Have each consortium adjust processing standards to include recommended metrics

Action items:

1) Get leads of multiplex immunofluorescence (e.g. CODEX) in contact with each other to determine if there is a path forward for developing common QC metrics
2) Organize a meeting of “key” people for each data type to set priorities for inventorying and settling on QC metrics
3) Make recommendations to funding agencies to incentivize the work needed for determining consensus and implementing changes

Names (with Consortia) on the team:

- Moderator(s): Josh Campbell
- Notetaker(s): Richard Conroy, Josh Campbell,
- Other Group Members and Consortia each works with: Jason Hilton (Lattice), Alex Ropelewski (BRAIN, HuBMAP), William Sullivan (HCA/DCP)

What additional expertise do you need?

- Names of leads for multiplexed immunofluorescence / CODEX imaging working groups

Our next meeting is? None

14.2: Designing the FAIR Pipeline

Problem we think we can solve by collaborating:

- None

Recommendation(s)/ What we want to do

- Findability:
    - Explore schema.org and dockstore approach
- Accessibility:
    - Common elements for documentation and instructions on how to use a pipeline, aimed at different user personas
- Interoperability:
    - Standards around reference datasets for benchmarking
    - Interfaces
        - Agree to choosing standard formats. Link up with data format standards group or GA4GH
        - Documenting vignettes for examples
- Reproducibility/Reusability:
    - Demoing frameworks for testing of pipelines
    - Container and containerization standards
    - Draft an SLA around reproducibility/reusability
- Common space / Cave entrance
    - Site or Github with a wiki that collects this information
    - Could be data biosphere GitHub Org but will be evaluated by group.
    - Set up levels or layers of adherence and a checklist to evaluate adherence to FAIR pipeline development

Names (with Consortia) on the team:

- Moderator(s): Kylee Degatano

- Notetaker(s): Timothy Tickle (HCA, BICCN, LungMap)
- Other Group Members and Consortia each works with: Satra Ghosh (BICCN, ReproNim), Bill Shirey (HuBMAP), Matt Wyczalkowski (HTAN, CPTAC3,), Heather Creasy (NeMO BICCN, CFDE), Gabdank Idan (ENCODE), Jeremy Miller (BICCN, SpaceTx)

At least one Email contact: Kylee Degatano

What additional expertise do you need?

- Reach out to DREAM Challenge folks about testing and benchmarking subtopic
- Dockstore representative for the findability discussion
- IHEC consortia representative (? Martin Hirst)
- Representative from 3/30 Data Storage and Data Movement subtopic
- https://kistorm.com/-M2JMzvUUT892QjweQUQ/-M3hxgmlVjCSzdF-uQij
- Rep from 3/30 Profiling and Analysis: How to make your containers FAIR

https://kistorm.com/-M2JMzvUUT892QjweQUQ/-M3iPTdNnSUHwFc7-OhE

Our next meeting is?

- Kylee will reach out to set up a meeting to discuss these topics (few weeks to a month from now)

15. Sharing & Standardizing Biospecimens & Experimental Methods

15.1: Overcoming the Legal and Institutional Barriers to Sharing Tissues and Potentially Identifying Data (Omics)

Problem we think we can solve by collaborating –

- Sharing of consent forms?
- Know what the legal barriers are for sharing tissues?
- Sharing best protocols for collecting and storing tissues.
- Sharing ideas on consents/MTA/DUAs to facilitate the sharing of biobanked tissues.
- Facilitate conversations with ethicists and regulatory agencies to get the regulatory structure more amenable to allow science to move forward and allow reasonable sharing of tissues.
- Have NIH involved -- fund core facilities?

Recommendation(s)/ What we want to do -

- Have more cross-talk across consortia and large programs to share best practices for collecting and storing tissues.
- Take advantage of this consortia of consortia to work out the challenges in the regulatory concerns around tissue biobanking and sharing of tissues (ideas such as common consent form - that can be shared with the regulatory agencies)
- Seek NIH help
- Determine if a cross-consortia group can leverage their needs across a variety of studies to a more unified/streamlined approach to biobanking.
- When you have rare diseases, working cross-consortia can increase N of the studies.

Names (with Consortia) on the team:

- Moderator(s): G Pryhuber (LungMAP and HuBMAP)
- Notetaker(s): Marc Halushka , Sharmistha “Sharmi” Ghosh-Janjigian
- Other Group Members and Consortia each works with: Lori Coburn (Vanderbilt, Gut Cell Atlas)

At least one Email contact:
Mauricio rojas contact email:
What additional expertise do you need? None

Our next meeting is?

- To discuss how to positively affect the regulatory environment surrounding biobanking and sharing of tissues.

15.2: Cross-consortia sharing of Protocols and Feedback on Protocols

Problem we think we can solve by collaborating –

- Sharing of protocols?
- Feedback on protocols?
- QC and validation of protocols?

Recommendation(s)/ What we want to do -

- Each consortia develop an updated site for all protocols. Protocols.io could be a resource where these are all centralized as SOPs for biobanking.
- Identify key people who can lead this effort from each consortium -- moving SOPs and protocols through to a common place.
- Encourage sharing. Team science is better. Open sharing of ideas, samples etc.-- Add this?

Names (with Consortia) on the team:

- Moderator(s): Gloria Pryhuber
- Notetaker(s):Notetaker(s): Marc Halushka, Sharmistha “Sharmi” Ghosh-Janjigian
- Other Group Members and Consortia each works with:

At least one Email contact:

What additional expertise do you need? None

Our next meeting is? None

15.3: Establishment of Virtual cross-consortium biobank

Problem we think we can solve by collaborating –

- Establishment of a virtual resource for sharing studies & samples from across consortia
- Establish metadata that would be needed for the biospecimen within the biobanking
- Facilitate collections of samples for collaborative studies
- Find a means to establish material/data sharing opportunities and agreements
- Need to invest in common API or reusable URI for a sample to support data

modeling, and then for a central registry to facilitate sample sharing
    - (ex. NCBI accession their samples, consider adopting their approach). Accession ID vs. DOI
- Establish a pilot using existing studies (HCA) to have a use case
    - e.g., how to make the best use of a benchmarking sample

Recommendation(s)/ What we want to do -

- **Establishment of Virtual cross-consortium biobank**
  It is increasingly clear there are many useful biobanked samples that have been generated across various consortiums. Many samples are plentiful enough to be useful for other analyses outside the consortium for which they were collected. Thus, discovery of what samples are available is needed. We propose the cross-consortium Virtual BioBank (VBB) be established to aid discovery and push forth a metadata standard for biospecimems that could be adapted across consortiums. The plan would be to provide existing metadata working groups with use cases to guide the further refinement of metadata work already underway, then use the final product in the virtual biobank. Consortial biobanks with tissues they want to make more broadly available would register samples for sharing in the VBB along with the metadata necessary for searching and evaluating samples by users.

Names (with Consortia) on the team:

- Moderator(s): M. Todd Valerius (GUDMAP/RBK)
- Notetaker(s): Sarah Mazzilli (HTAN), Maryann Martone (), Melissa Cook (HTAN/NCI)
- Other Group Members and Consortia each works with: Kristin Ardlie (GTeX), Laura Clarke, Bruce Aronow

What additional expertise do you need? None

Our next meeting is? None

15.4: Identification of biobanking effort that can be utilized for "benchmarking" studies for cross network assay QC (i.e. benchmarking set of samples available cross-consortia)

Problem we think we can solve by collaborating –

- Identify studies/biobanks for benchmarking samples (where samples are plentiful)
- Identify what are the types of cases and samples would be beneficial for benchmarking (standard tissues, diversity of donors)
- Confirming the standardization of protocols for collection

- Enable cross and within consortia benchmarking.
- Define type of sample storage (FFPE and/or frozen)

Recommendation(s)/ What we want to do -

- **Establish "benchmarking" set of samples available cross-consortia**
  Using common tissue sources is a way to benchmark emerging technologies in comparison to existing ones, in particular the quality and sensitivity of experimental modalities. To maximize the availability of a shared, openly shared tissue source that includes high volume and broad tissue/organ sampling, we recommend a large sample collection from a limited set of cadaveric donors. This would enable abundant reference benchmarking tissue across tissues from the same donor that may be used across many consortia and non-consortial research laboratories. This would require funding of a collaborative group that involves metadata-computational expertise, and tissue collection and biobanking expertise.

Names (with Consortia) on the team:

- Moderator(s): M. Todd Valerius (GUDMAP/RBK)
- Notetaker(s): Sarah Mazzilli (HTAN), Maryann Martone (), Melissa Cook (HTAN/NCI)
- Other Group Members and Consortia each works with: Kristin Ardlie (GTeX), Laura Clarke, Bruce Aronow, Chris Briggs

What additional expertise do you need? None

Our next meeting is? None

16. Cell Type Annotation

16.1: Automated cell type annotation

Problem we think we can solve by collaborating –

1. Reconciliation of types of evidence for annotations

Recommendation(s)/ What we want to do -

1. Goal: Nucleation of a training dataset for annotations
    a. Identify and aggregate data (perhaps by community?)
    b. Construct and share SOP for iterative annotation from the lung group
    c. Construct an iterative approach to refine those annotations (council of experts? call for participation across consortia?)
    d. Release data to community

What additional expertise do you need? None

Our next meeting is? None

16.2: Expert annotations_ tools and initiatives

Problem we think we can solve by collaborating –

- How do we agree to a standard annotation format?
- How will we name annotations in a consistent fashion (naming and strategy to name)?

Recommendation(s)/ What we want to do -

- How do we agree to a consensus format?
    - We would like associated consortia to agree to release annotations with standard annotation formats
    - Would like to have a focused effort funded to develop tooling around this.
    - Look into existing technologies and schema that can support this.
    - Develop standard naming and indexing systems focused on transcriptomic clusters.
- How will we name annotations in a consistent fashion (naming and strategy to name)?
    - Establish a joint working group between different organs/consortia to make generalized annotations. Aiming for simple, usable systems.

- This should be coordinated with ontology efforts.

Moderator(s): Peter Kharchenko
Notetaker(s): Timothy Tickle (HCA, BICC, LungMap)
Names (with Consortia) on the team:

- Chris Briggs, Sr Data Curator HuBMAP:
- Carol Thompson, BCDC/BICCN:
- Alexandre Denadai-Souza, Gut Cell Atlas KU Leuven:
- David Osumi-Sutherland (EBI)
- Ed Lein, BCDC/BICCN:
- Dave Rogers UCSC

At least one Email contact: Peter Kharchenko
What additional expertise do you need? None

Our next meeting is?

- **Contact** Peter Kharchenko

16.3: Share Biosamples

Problem we think we can solve by collaborating –

- Improve the ability to share tissue so that we can collect data from different modalities. Identifying groups that have different techniques to apply to the same tissue

Recommendation(s)/ What we want to do -

- KPMP can share information on how to how to align paper work to make this possible

Names (with Consortia) on the team:

- Moderator(s): Kerstin Meyer
- Notetaker(s): Kathy Reinold

What additional expertise do you need? None

Our next meeting is? None

16.4: Common Nomenclature

Problem we think we can solve by collaborating –

- Identifying ontologies that we can use cross-species, cross-modality

- Need for common nomenclature -- common upper ontology
- Consider using a separate field to allow organ-specific ontology development.

Recommendation(s)/ What we want to do -

- Umbrella that can combine these; funding mechanisms
- High level interactions between HCA, hubmap, etc
- Training for labs to annotate using common nomenclature
- Additional expertise needed: ontologists + software tool designers for imaging and RNA-seq to combine these tools together.
- Organ-specific jamborees as noted by other break-outs

Names (with Consortia) on the team:

- Moderator(s): Kerstin Meyer
- Notetaker(s): Kathy Reinold
- Other Group Members and Consortia each works with: Kathy Reinold (Broad)

What additional expertise do you need?

- Reach out to leaders of consortia to lobby funders to support this.

Our next meeting is? None

---

17. Data Search and Visualization

17.1: Cross-platform search: challenges, opportunities and requirements for stronger technical collaboration,

Problem we think we can solve by collaborating –

- Search across all atlases for key metadata such as tissue type, experimental approach (protocol) to find related information
- Advance from free text search to controlled vocabularies and entity names, most likely in a minimum-information framework
- Spatial search against common coordinate frameworks, at least within individual tissue types or disease areas.
- More sophisticated search (e.g. with query levels) needs use cases so it can be correctly specified

Recommendation(s)/ What we want to do -

- Need to achieve integration across atlas efforts with respect to key technical issues such as search, indexes, ontologies; we have widely dispersed expertise.
- Need to have a realistic assessment of what search could accomplish in a realistic timeframe given available resources
- Need (simple) ontologies of controlled vocabularies to better understand the methods used to collect specific data sets.

Names (with Consortia) on the team:

- Moderator(s): Peter Sorger (HTAN/HuBMAP)
- Notetaker(s): Sarah Mazzilli (HTAN)
- Other Group Members and Consortia each works with:

At least one Email contact: Ben Hitz (ENCODE DCC), Jonathan Silverstein (HuMAP),

What additional expertise do you need? None

- **Our next meeting is?**
  Meeting around search technologies does this need to be a sandbox of some sort?

17.2: Proposal for NIH CFDE (initial target OTA-20-005)

Problem we think we can solve by collaborating –

- Develop Data Formats, Exchange Mechanisms, and Tools to Enable Cross-Consortia or Cross-Project Registration and Querying of Single-Cell and Spatial Profiling of [Heart, Kidney, Lung, Liver, Brain], focused on addressing specific medical questions. Test the implementation of a CCF across these datasets.

Recommendation(s)/ What we want to do -

- Develop and demonstrate data formats that capture representations of data from independent profiling experiments for one or more organs. Prove querying, cross-dataset integration, and visualization, e.g., via the CCF Exploration User Interface, https://hubmapconsortium.github.io/ccf-ui/
- Build efforts based on concrete, specific use cases with both experimental, and biomedical questions.

Names (with Consortia) on the team:

- Moderator(s): Jason Swedlow (OME)
- Notetaker(s): Alex Ropelewski (BRAIN/BIL, HuBMAP)
- Other Group Members and Consortia each works with Sabrina Summer MUW (HCA), Katy Borner, IU (HuBMAP), Jim Hagood, UNC (HuBMAP and Normal Aging Lung Atlas)

What additional expertise do you need?

- Need to develop based on inspection of the survey of organs, datasets from CCF breakout.

Our next meeting is? None

18. Ethics and Diversity

18.1: Ethics - sharing tools, approaches and best practices

Problem we think we can solve by collaborating –

- Identifying mechanisms and ways to share ethics tools (e.g. consent form templates, approaches to data tiering, etc.) across consortia. This will be helpful to harmonize language, consent approaches and best practices across consortia.

Recommendation(s)/ What we want to do -

- Identify existing ways/portals to share information and documents:
    - Can we put this on protocols.io?
        - Currently used, for example, to share sample collection protocols. Allows versioning, which is useful.
    - Or a consent ontology? http://www.obofoundry.org/ontology/ico.html
    - Potentially convened a group to decide what repository we use for protocols, or something we create. We could create a place for our own protocols, but if we don't get feedback, not very useful.
- Reach out to metadata teams to determine degree of interaction on ethics guidance? (HuBMAP is working with HCA on a cross consortia metadata group - maybe contact Laura Clarke?)
- How do other consortia approach data tiering (e.g. raw data, metadata)? Discussing approaches, models would be helpful.
- Finally, for future collaboration, it would be useful to identify key contacts for ethics for all the different consortia and inviting them to join the discussion

Names (with Consortia) on the team:

- Moderator(s): Emily Kirby, Alex Shalek, Orit Rozenblatt-Rosen
- Notetaker(s): Dena Procaccini, Kristin Ardlie
- Other Group Members and Consortia each works with: To be identified

At least one Email contact: Emily Kirby, HCA Ethics Working Group

What additional expertise do you need? None

Our next meeting is?

- HCA Ethics Working Group meeting via bi-monthly teleconferences

18.2: Equity

Problem we think we can solve by collaborating –

- Engage global scientific community in the analysis and interpretation of data
- Connectivity between the community
- Support structures - elements from grants to help support the interactions

Recommendation(s)/ What we want to do -

- Making people aware of where they can collect information
- Identify existing gaps - lack of support or interactions we need
- How do we engage with more communities? Other regions of the world to make more connections.
- Who are the right people to connect with around the world? Identifying them is important.
- Listening to the community.

Names (with Consortia) on the team:

- Moderator(s): Alex Shalek
- Notetaker(s): Dena Procaccini/Kristin Ardlie
- Other Group Members and Consortia each works with:

At least one Email contact: Alex Shalek, Norbert Tavares (CZI/HCA)
What additional expertise do you need?

- Greater representation to understand the needs of the global community

Our next meeting is?

- Brazil or Vietnam in late summer/fall
- We should schedule call to engage others who are interested in participating in the topic

19. FAIRness

19.1: Exploring the complexity of FAIR in practice

Problem we think we can solve by collaborating –

- Identify key areas where the consortia together could go deep on the minimum information for specific technical platforms
    - **Focus on the F in FAIR for single cell RNASeq across the various NIH consortia**
- Other discussion points
    - Suggestion: Define some use cases to illustrate need for FAIR between consortia data (start with existing general use cases that have been developed)
    - Suggestion: Focus on building out shims between consortia and how to come up with simple bridges focused on F-in-FAIR that span technical differences
    - Suggestion: Agree on and implement common API layer upfront in each Consortium that supports full interoperable
    - Suggestion: Have a support service for implementing / harmonizing data elements?
    - Suggestion: Pilot to map / link most commonly used ontologies?
    - Suggestion: Each consortium aligns what it has with a common high level data model for "dataset"? (e.g. DATS, DCAT, schema.org, bioschemas.org)
    - Suggestion: Best practices guide for new consortia getting off the ground so they don't need to reinvent the wheel?

Recommendation(s)/ What we want to do -

- Come up with several use cases that scientists are looking for and use that illustrate the complexities of FAIR?

Names (with Consortia) on the team:

- Moderator(s): Matt Wyczalkowski
- Notetaker(s): Anonymous Alligators, Penguins, Foxes, Crows
- Other Group Members and Consortia each works with: Satra (BICCN, ReproNim, BRAIN Initiative), Fore (HTAN, CRDC, NCI), Matt Wyczalkowski (HTAN), Ian Fore (NCI), Satra Ghosh (BRAIN), Bruce Herr (HuBMAP), Chuck McCallum (HuBMAP), Richard Conroy (HuBMAP)

At least one Email contact: Matt Wyczalkowski (HTAN, CPTAC3)

What additional expertise do you need?

- Someone with big carrots and sticks

Our next meeting is?

- Check to see if there is synergy with other groups as a way to move forward with this idea

19.2: How might we use Data Citation as an entry into education on FAIR and why they should care about establishing FAIR data

Problem we think we can solve by collaborating –

- Improve adherence to the FAIR principles across the consortia in a clear and consistent manner

Recommendation(s)/ What we want to do -

1. Assess current capabilities (https://www.nature.com/articles/s41597-019-0031-8)
2. Educational meetings or a workshop on supporting data citation
    a. Repositories/data centers
    b. Publisher/journal centered view
    c. Researcher centered view
3. Develop best practices guidelines. Set up GitHub resource for documents.

Names (with Consortia) on the team:

- Moderator(s): Maryann Martone (BICCN, SPARC)
- Notetaker(s): Kylee Degatano (HCA, BRAIN/Broad), Melissa Cook (HTAN/NCI)
- Other Group Members and Consortia each works with: Bill Shirey (HuBMAP), Idan Gabdank (ENCODE), Heather Creasy (BICCN NeMO, CFDE)

At least one Email contact: Maryann Martone

What additional expertise do you need?

- Representatives of groups that are working similar things or have tools; members of the consortium

Our next meeting is?

- About a month for this subgroup but we are hoping that the organizers of this meeting will bring together groups with common interest in FAIR

19.3: Developing a persistent structure for on-going education around FAIR

Problem we think we can solve by collaborating –

- Improve adherence to the FAIR principles across the consortia in a clear and consistent manner. Develop a larger plan around further education; having workshops on a regular basis (perhaps as webinars on special topics). Will keep people engaged and involved. Coordinate with outreach group.

Recommendation(s)/ What we want to do -

- Create a "customer success" group (look to what the Broad Data Sciences Platform has done)
    - Contact group at NIH charged with wrangling all these consortia; perhaps the Office of Data Science Strategy. But would have to involve HCA so not NIH.
    - Create a small dedicated team to be involved in education about, promotion of the value proposition of FAIR and promote adoption of FAIR.
    - Develop best practice guidelines for implementation
- Lead the development of the framework for layers of adoption of FAIR principles
- Forum/Community for implementation expertise (ex. Github resource, Slack channel)
    - Central resources for information sharing, e.g., Stack overflow (guidance, FAQs) with moderators. INCF NeuroStars may be an option
    - Also provide 1:1 support
- Build on existing solutions across the consortia and broader community

Names (with Consortia) on the team:

- Moderator(s): Maryann Martone (BICCN, SPARC)
- Notetaker(s): Kylee Degatano (HCA, BRAIN/Broad), Melissa Cook (HTAN/NCI)
- Other Group Members and Consortia each works with: Bill Shirey (HuBMAP), Idan Gabdank (ENCODE), Heather Creasy (BICCN NeMO, CFDE)
-

At least one Email contact: Kylee Degatano

What additional expertise do you need?

- Representatives of groups that are working similar things or have tools; members of the consortium

Our next meeting is?

- About a month for this subgroup but we are hoping that the organizers of this meeting will bring together groups with common interest in FAIR

20. Outreach

20.1: How do we coordinate outreach efforts?

Problem we think we can solve by collaborating –

- How do we coordinate outreach efforts?

Recommendation(s)/ What we want to do

- Mailing lists for areas of interest
- HCA mailing list can be created for outreach
- Designate 1-3 co-leaders for this mailing list
    - These folks would be responsible for following up with emails/action items that come into this mailing list

Names (with Consortia) on the team:

- Moderator(s):
- Notetaker(s): Kristine Schwenck, Broad, HCA
- Other Group Members and Consortia each works with:

At least one Email contact: Kristine Schwenck

What additional expertise do you need?
Emails to add

Our next meeting is? None

20.2: Publish a meeting report about a joint effort of all the various stakeholders of this NIH/HubMap/HCA... meeting

Problem we think we can solve by collaborating –

- Work to coordinate efforts and tell the world about the broad set of efforts discussed at this conference.
- Include links to all the consortia and related outreach efforts

Recommendation(s)/ What we want to do -

- None

Names (with Consortia) on the team:

- Moderator(s):
- Notetaker(s):
- Other Group Members and Consortia each works with:

At least one Email contact:

What additional expertise do you need?
- The organizers of the meeting. Need to ask them what plans are in place already.

Action item:
- Robin will email Donnalyn

Our next meeting is? None
