Multi-Level, Multi-Modal CCF UI for Data Providers and Users within the Human BioMolecular Atlas Program (HuBMAP)

MC-IU Team:

Katy Börner, Lisel Record, Bruce Herr II, Leonard Cross, Paul Macklin, Randy Heiland, Jim Sluka & Ellen Quardokus

Intelligent Systems Engineering, SICE, Indiana University, Bloomington, IN

Griffin Weber, Harvard Medical School, Boston, MA

Samuel Friedman, Opto-Knowledge Systems, Inc.

HuBMAP HIVE Q3 Demo Day
Harvard Medical School, Boston, MA

June 27, 2019
Overview

• Introduction

• Y1 Q3 Progress Report
  • CCF Ontology
  • CCF Meta-Data Review
  • CCF Data Wiki
  • CCF User Interface (UI)

• Y2 Plans (June 21, 2019 - June 20, 2020)
  • CCF Mapping and Ontologies
  • CCF UI and Tissue Registration UI
  • Visual Human Massive Open Online Course (VHMOOC)
  • User Studies
Introduction

- Provenance
- Patient
- Sample
- Sample Processing
- Technology (MS, IH, ...)
- Analysis
- Etc.

Propagate needs back to TMC's

- Only the data needed for the GUI

TMC: Tissue Mapping Center
PSC: Pittsburgh Supercomputing Center
Propagate needs back to TMC’s

- Provenance
- Patient
- Sample
- Sample Processing
- Technology (MS, IH, …)
- Analysis
- Etc.

Only the data needed for the GUI

TMC: Tissue Mapping Center
PSC: Pittsburgh Supercomputing Center
Y1 Q3 Progress - Ontology for Kidney
CCF Ontology: some guiding principles

• Reuse existing ontologies and data formats developed for projects similar to HuBMAP to the greatest extent possible
  • GUDMAP / RBK
  • Human Cell Atlas
  • …
• Reuse domain-specific ontologies and data formats
  • OME-Tiff (Open Microcopy Community advanced image format)
  • MIAME (Minimum Information About a Microarray Experiment)
  • …
• Leverage HuBMAP domain expertise!
  • Each TMC is an expert in its organ. Capture this in the organ-specific ontologies.
• Use a standard Ontology format and development tools
  • We will use OWL
  • Include test cases in the ontology itself (e.g. both A-box and T-box) for testing, validation and demonstration purposes.
• Cross-link with existing ontologies as much as possible
• Need partOf trees for simplified navigation in GUI.
CCF: Source Ontologies

Anatomic/Phenotypic
- Uberon
- Foundational Model of Anatomy (FMA) (has anatomical terms NOT in Uberon)
- Human Phenotype Ontology (HPO)
- Phenotype and Trait Ontology (PATO)
- Organ specific: Kidney Tissue Atlas Ontology (KTAO) and LungMAP

Tissue/Data Collection
- Biological Spatial Ontology (BSPO)
- Ontology of Biomedical Investigations (OBI)
- EDAM (Bioinformatics concepts)

Open Biological and Biomedical Ontology (OBO)
Foundry is a collective of ontology developers that are committed to collaboration and adherence to shared principles.
BioPortal is a collection of ontologies for biomedical research.

(Sub-)Cellular
- Cell Ontology (CL)
- Gene Ontology (GO)
- Chemical Entities of Biological Interest (ChEBI)
- RNA Ontology (RNAO)
- Protein Ontology (PR)
- Cell Behavior Ontology (CBO)

Metadata
- Basic Formal Ontology (BFO)
- Information Artifact Ontology (IAO)
- Ontology of units of Measure (OM)
- Provenance, Authoring and Versioning ontology (PAV)
- VIVO (Identifying researchers)

MeSH and NCI Thesaurus
Current CCF Ontology

- Uses Uberon and user-supplied tables of terms to create a SLIM ontology
- Users (initially TMCs) can request missing terms as needed
- "partOf" and other partonomy terms used to help relate concepts
  - Requires domain expertise!
  - Individual TMCs will need to pitch in for their specific organs to refine
- Click here to visualize the current CCF ontology
Current CCF Ontology:

- Use Uberon and user-supplied tables of terms to create a SLIM ontology.
- Users (initially TMCs) can request missing terms as needed.
- "PartOf" and other partonomy terms used to help relate concepts.
- Requires domain expertise!
- Individual TMCs will need to pitch in for their specific organs to refine.

Click here to visualize the current CCF ontology.
Current CCF Ontology:

- Use Uberon and user-supplied tables of terms to create a SLIM ontology
- Users (initially TMCs) can request missing terms as needed
- “PartOf” and other partonomy terms used to help relate concepts
- Requires domain expertise!
- Individual TMCs will need to pitch in for their specific organs to refine

Click here to visualize the current CCF ontology
Y1 Q3 Progress – CCF Meta-Data Review
IU CCF Initial (v0.5.0) Image Formats

**Basic image**: OME-Tiff as the base “image” format based on:
- 2D to 4D data (includes movies)
- more than three “color channels”
- More flexible “color” data format (int, float, etc.)

**Regions of images**: SVG with annotations
(aligned with a particular OME-Tiff)

**Volumetric** (e.g., computed tomography, MR, ultrasound, ...)
- Data normally represented as volumes or surfaces
Data. Is very heterogeneous. Must provide guidance.

Kidney: Jeff Spraggins et al., VU
See data on Globus, BIOMIC_patient-64354

Clinical and Spatial Metadata (21 rows)

Cell Types, on right

Cell States (9 rows)

Cell cycle arrest
- G0
- G1/S
- G2/M

Heart: Shin Lin, UW
Year 1: Tissue data for 1-2cm cubed volumes from 9 sites for 1 heart from 1 individual.

Data Dictionary (115 rows)

Cell Types (14)
- endothelial cells
  - arterial
  - capillary
  - venous
  - lymphatic
- cardiomyocytes
  - atrial
  - ventricular
  - nodal
- fibroblasts
  - fibroblasts
  - myofibroblasts
- immune cells
  - macrophages
# Data: Clinical

**Kidney: Jeff Spraggins et al., VU**

### Clinical and Spatial Metadata (21 rows)

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Number</td>
<td>64354</td>
</tr>
<tr>
<td>Procedure ID</td>
<td>66598</td>
</tr>
<tr>
<td>Date</td>
<td>1/30/2019</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td><strong>38</strong></td>
</tr>
<tr>
<td>Gender</td>
<td>Female</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
</tr>
<tr>
<td>Height</td>
<td>165.1 cm</td>
</tr>
<tr>
<td>Weight</td>
<td>115.2 kg</td>
</tr>
<tr>
<td>BMI</td>
<td>42.3</td>
</tr>
<tr>
<td>Comorbidities</td>
<td>Obesity</td>
</tr>
<tr>
<td>Type of Procedure</td>
<td>Total Nephrectomy</td>
</tr>
<tr>
<td>Indications for Procedure</td>
<td>Renal tumor</td>
</tr>
<tr>
<td>Laterality</td>
<td>Left</td>
</tr>
<tr>
<td>Tissue Type</td>
<td>kidney</td>
</tr>
<tr>
<td>Dimensions (mm)</td>
<td>L: 19 x W: 13 x H: 7</td>
</tr>
<tr>
<td>Anatomical Landmark</td>
<td>Lower Pole</td>
</tr>
<tr>
<td>Distance from Tumor</td>
<td>7 cm</td>
</tr>
<tr>
<td>Sample Processing</td>
<td>Frozen</td>
</tr>
<tr>
<td>Method of Freezing</td>
<td>Dry Ice/Isopentane Slurry</td>
</tr>
<tr>
<td>Embedding Media</td>
<td>CMC</td>
</tr>
</tbody>
</table>

**Heart: Shin Lin, UW**

### Data Dictionary (115 rows)

<table>
<thead>
<tr>
<th>Field #</th>
<th>Sort</th>
<th>Field Label Sort</th>
<th>Field Name Sort</th>
<th>Field Units</th>
<th>Field Data <code>Lookup</code> <code>Low Value</code></th>
<th><code>High Value</code> <code>Valid value</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Donor // ABO</td>
<td>abo</td>
<td>char(3)</td>
<td>lookup_abo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Donor // Date of birth</td>
<td>dob</td>
<td>datetime</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Donor // Gender</td>
<td>gender</td>
<td>char(1)</td>
<td>lookup_gender</td>
<td>M,F</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Details // Age</td>
<td>age_in_months</td>
<td>smallint</td>
<td>0 1188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Details // Age Unit</td>
<td>age_unit</td>
<td>char(1)</td>
<td>lookup_age_unit</td>
<td>M,Y</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Details // Height</td>
<td>hgt_cm</td>
<td>cm</td>
<td>decimal(5, 2)</td>
<td>1 241.3</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Donor hgt_ft //</td>
<td>hgt_ft</td>
<td>ft</td>
<td>int</td>
<td>0 7</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Donor hgt_in //</td>
<td>hgt_in</td>
<td>in</td>
<td>int</td>
<td>0 11</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Details // Weight</td>
<td>wgt_kg</td>
<td>kg</td>
<td>decimal(7, 4)</td>
<td>0.454 294.835</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Donor wgt_lb //</td>
<td>wgt_lb</td>
<td>lbs</td>
<td>decimal(3, 0)</td>
<td>2 650</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Donor // Ethnicity/race</td>
<td>race</td>
<td>bigint</td>
<td>lookup_race_subcat_multi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Details // History of diabetes</td>
<td>diabetes</td>
<td>smallint</td>
<td>lookup_diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Donor // History of cancer</td>
<td>cancer</td>
<td>smallint</td>
<td>lookup_cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Donor History of cancer, cancer_other</td>
<td>cancer_other</td>
<td>varchar(50)</td>
<td>1 50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Details // History of hypertension</td>
<td>hypertension</td>
<td>smallint</td>
<td>lookup_hypertension</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data: 3D Tissue

Kidney: Jeff Spraggins et al., VU

See data on Globus, BIOMIC_patient-64354

Heart: Shin Lin, UW

Year 1: Tissue data for 1-2cm cubed volumes from 9 sites for 1 heart from 1 individual.

Terminology; Coordinates and photos to spatialize

Outer cortex and capsule

Inner cortex

Outer medulla

Inner medulla

Papilla
### Data: TMCs x Organs x Data Types x Technologies

**BUKMAP, Zhang Group**

<table>
<thead>
<tr>
<th>Organs (10)</th>
<th>Data Types (13)</th>
<th>Technologies (~25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bladder</td>
<td>1. Imaging - Proteins</td>
<td>CODEX;DART-FISHrp;IF;IHC;LRET-IF;MALDI Imaging MS;PER-DEI</td>
</tr>
<tr>
<td>2. Blood Vessel (Heart)</td>
<td>2. Imaging - RNA</td>
<td>DART-FISH;LRET-ISH;MERFISH;PER-DEI;seqFISH;smFISH</td>
</tr>
<tr>
<td>3. Breast</td>
<td>3. Imaging - DNA</td>
<td>PER-DEI</td>
</tr>
<tr>
<td>4. Colon</td>
<td>4. Imaging - Other</td>
<td>Lipid and Metabolite MALDI Imaging MS</td>
</tr>
<tr>
<td>5. Kidney</td>
<td>5. scRNAseq</td>
<td>snDropseq;scRNAseq</td>
</tr>
<tr>
<td>6. Liver</td>
<td>6. scDNAseq</td>
<td>scATACseq;scTHSseq;SNAREseq</td>
</tr>
<tr>
<td>7. Lung</td>
<td>7. scProteomics</td>
<td>IMC</td>
</tr>
<tr>
<td>8. Spleen</td>
<td>8. bulk-Proteomics</td>
<td>LC-MS/MS</td>
</tr>
<tr>
<td>9. Thymus</td>
<td>9. bulk-RNA</td>
<td>?</td>
</tr>
<tr>
<td>10. Tonsil</td>
<td>10. bulk-DNA</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>11. Metabolomics</td>
<td>LC-MS/MS;nano-POTS</td>
</tr>
<tr>
<td></td>
<td>12. Lipids</td>
<td>LC-MS/MS;nano-DESI</td>
</tr>
<tr>
<td></td>
<td>13. Other</td>
<td>Autofluorescence;PAS stained microscopy</td>
</tr>
</tbody>
</table>

No Bone Marrow and Pancreas.
Review of Meta-data formats of similar projects

Other NIH Centers
• GenitoUrinary Development Molecular Anatomy Project (GUDMAP)
• (Re)Building a Kidney (RBK)
• Kidney Precision Medicine Project (KPMP)

Other Efforts
• Human Cell Atlas (HCA)
• ApiNATOMY
Welcome to the CCF Data Wiki!

<table>
<thead>
<tr>
<th>Organ</th>
<th>CalTech</th>
<th>UCSD</th>
<th>Stanford</th>
<th>Vanderbilt</th>
<th>Florida</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowel</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thymus</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spleen</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymph Nodes</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bladder</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Colon</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Legend:
- ✓ - Organ proposed and survey submitted
- × - Organ was proposed, but no survey has been submitted

Y1 Q3 Progress - CCF User Interface (UI)
Y1 Q3 Progress - CCF User Interface (UI)

GitHub demo site: https://hubmapconsortium.github.io/ccf-ui/
Y1 Q3 Progress - CCF User Interface (UI)
MC-IU has released CCF user interface v0.5.0, a proof-of-concept version of the CCF UI. The user interface supports:

- Visual browsing of tissue samples and metadata at the whole body, organ, tissue, and cell level.
- Filtering by metadata (age, gender, TMC, and technology), results are presented at all views.
- Submission of questions and comments on the CCF UI.
- Semantic search by ontology, results are presented at all levels.
- Data download at the whole body, organ, tissue, and cell level, i.e., link to https://sampledata.hubmapconsortium.org.

See also:

Recorded demo: https://www.youtube.com/watch?v=rWMqKQc_00w&feature=youtu.be
GitHub link to code: https://github.com/hubmapconsortium/ccf-ui
GitHub demo site: https://hubmapconsortium.github.io/ccf-ui/
Original specs: https://drive.google.com/open?id=1tqUzmVLxwqcGprtR1evfY86YvHHPEsDR

Live Demo!
CFW Workshop

HuBMAP | COMMON COORDINATE FRAMEWORK WORKSHOP CCFWS-01

Time & Date

9:00am-5:00pm EDT on May 9, 2019

Goals

HuBMAP will develop a common reference map or coordinate system called the Common Coordinate Framework (CCF). As stated in the Common Coordinate Framework Meeting (CCFM) document, a CCF makes it possible to uniquely and effectively define and name any location in the human body. A set of robust origin points (serving as landmarks) make it possible to reference organs, tissues, cells over different anatomical scales; tolerate human variability and function across lifespan and disease, and help integrate heterogeneous data layers and a wide range of reference maps such as whole body spatial maps, genetic variant maps, and coordinate systems that align with vascular pathways.

This CCF workshop will focus on a kidney-specific CCF and atlas but also discuss other relevant CCF/atlas efforts. It will feature presentations and discussions on:

- CCF metadata—what data are currently captured, how can they be unified across tissue mapping centers (TMCs), what additional data are needed to meet stakeholder (research) needs.
- CCF ontologies—what ontologies exist and are used in what part of the data pipeline; what ‘desirable properties’ and ‘success criteria’ exist?
- CCF mapping and numerical construction—including dealing with human variation and using CCF user interfaces as a means to properly register data and review data completeness.
- General principles and processes that can inform CCF design for other organs and continuous adaptation of CCF to emerging technologies and ever changing user requirements.

The ultimate goal is a set of draft guidelines for TMCs detailing what data to provide in which formats to maximize CCF mapping accuracy and data utility.

All slides, video recordings are at https://ccfws.cns.iu.edu
Y2 Plans

- CCF Mapping and Ontologies
- CCF UI and Tissue Registration UI
- Visual Human Massive Open Online Course (VHMOOC)
- User Studies
Y2 Plans: CCF Mapping and Ontologies

- Complete kidney ontology and switch to “part of” hierarchy
- V1.0.0 Data format specifications for image data including specification of identified regions
- Patient, sample and technology metadata ontologies in collaboration with the tissue centers
- Lung ontology
Y2 Plans: CCF UI and Tissue Registration UI
Multiple TMC’s

- Provenance
- Patient
- Sample
- Sample Processing
- Technology (MS, IH, …)
- Analysis
- Etc.

- Only the data needed for the GUI

Propagate needs back to TMC’s

End user Interface

Tools to assist in sample spatial registration

TMC: Tissue Mapping Center
PSC: Pittsburgh Supercomputing Center

Data Wrangler

Data Filter (HIPPA, quality control, …)

- Global Data Store
- Data Traceback
- CCF Data Store
- User Interface (IU)

Additional Processing

CCF User Interface (UI)

Tissue Registration UI
Tissue Registration UI: Heart (depending data availability)

Align 9 tissue samples in 3D heart using a combi of

- Rough placement using human expertise/3D pattern matching and
- Fine adjustments using machine learning

Virtual tissue samples will be sized 1-2cm cubed, numbered (1 ... 9), and oriented (left-right, top and bottom tissue slice of z-stack).

Measure error from

- precision of tissue sample procurement and
- placement in the 3D browser

We hypothesize that placement accuracy will improve when additional information (e.g., landmarks, major scaffolds, MR/CT scan of heart after 9 samples were extracted) is being visible in virtual organ.
Human heart with data overlay
Developer: Andreas Bueckle

Heart model from NLM3D Data
https://lhncbc.nlm.nih.gov/project/visible-human-project
Tissue Registration UI: Kidney
(depending data availability)

• Exploit human pattern recognition and fine motor skills (by surgeons) to register tissue in organs.

• Add info on anatomical landmarks, cell types, molecular data to support alignment.

• LATER: Use human alignment data as training data for machine learning algorithms, to better support manual alignment OR to possibly fully automatize alignment.

Kidney model from NLM3D Data
https://lhncbc.nlm.nih.gov/project/visible-human-project

How many of you have used a VIVE or space mouse?
Kidney model from NLM3D Data

https://lhncbc.nlm.nih.gov/project/visible-human-project
For the kidney, there exist no predefined tissue extraction sites. The current kidney Registration UI uses a grid system and a picture of a kidney slice to guide placement, see **Fig. 3 left and middle**. Funding of this GLUE grant will make it possible to use 3D image volumes collected from intact kidney tissue within KPMP that feature anatomically meaningful structures of different cell types, see **Fig. 3 right**.

---

**Figure 3**: 3D grid system (left), 2D image (middle), 3D tissue reconstructions from KPMP (right)

Using the 3D grid system, a user can use 3D coordinates to refer to a particular area in the kidney, e.g., A-4-II would correspond to the lower-middle part of the kidney, on the inside (occluded in this view). Using the 2D image, sample placement within a semi-transparent reference kidney object is guided by anatomical structures. Using KPMP data, it will be possible to show complex volumetric structures inside the 3D reference kidney to allow for more granular alignment at the molecular level.

---

**KPMP GLUE grant proposal in progress.**

---

**Fig. 4.** KPMP volumetric data.  
*See presentation by Seth Winfree for details*  
[https://ccfws.cns.iu.edu](https://ccfws.cns.iu.edu)
Y2 Plans: CCF UI

• Use the CCF UI to serve a Kidney Micro Atlas.
• Go from Proof of Concept to fully-realized application
  • Fully integrate heart and kidney data
  • Integrate Tissue Registration UI data
  • Integrate expanded CCF Ontology
  • Connect to data/queries via IEC APIs and infrastructure
• Collaborate on a common Tissue Viewer
Y2 Plans: Visual Human Massive Open Online Course (VHMOOC)

Research and develop a Visual Human Massive Open Online Course (VHMOOC) that helps communicate the

- quality and coverage of HuBMAP data,
- utility and proper usage of CCF Tissue Registration, CCF UI, and HuBMAP tools, and
- demonstrate new single-cell analysis and CCF mapping techniques.

First interviews with experts will be run at the HuBMAP meeting at Stanford U. Please let MC-IU know if you are interested/available to showcase your work.
Y2 Plans: User Studies

Conducting in-depth interviews with experts to understand user needs (Y1Q3, will continue in Y2Q1)

Information gathered will inform both the Tissue Registration UI and the Common Coordinate Framework UI.

Identified 6 draft personas (Y1Q3, will refine in Y2)

Generator, Analyst, Technologist, Computationalist, Data Provider, Educator/Student

Based on interviews, user stories, surveys, and comparisons with similar projects (HCA, KPMP)

We welcome your feedback on their applicability across the project. Review and provide feedback at https://bit.ly/2ZFHWYW.

Developed methodology and obtained IRB approval for user studies to be conducted in Y2 (Y1Q3)

Task-oriented user testing to improve the CCF user interface. Administered online, with HuBMAP members and other biological researchers who are representative of users based on their research interests.