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.

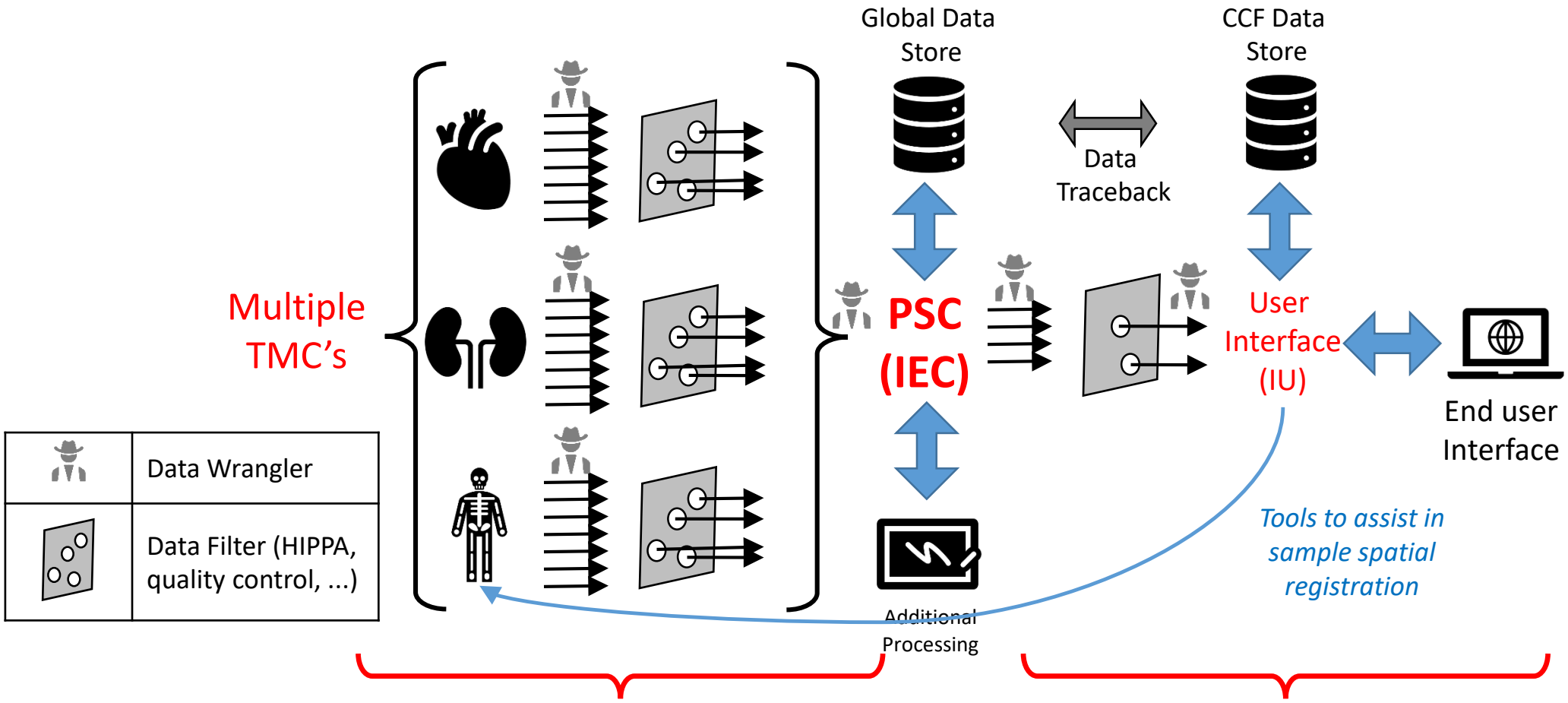
Common Coordinate Framework (CCF) Workshop

Indiana University, Bloomington

May 9, 2019

Overview

- Data
- 3D Spatial Models Linked to Terminology/Ontology
- CCF Tissue Registration Interface (for Data Providers)
- CCF User Interface (for Data Users)
- Year 2 Plans
- Open Questions



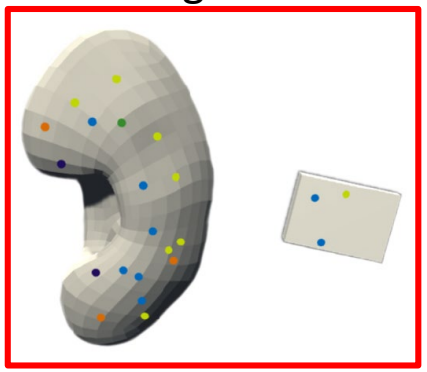
- Provenance
- Patient
- Sample
- Sample Processing
- Technology (MS, IH, …)
- Analysis
- Etc.

Propagate needs back to TMC's

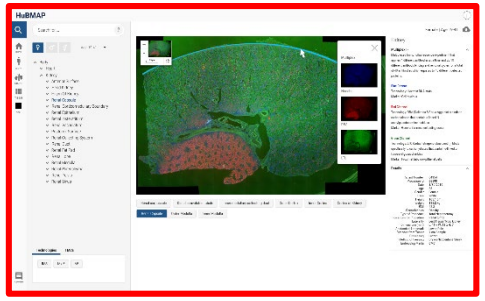
- Only the data needed for the GUI

TMC: Tissue Mapping Center
PSC: Pittsburgh Supercomputing center

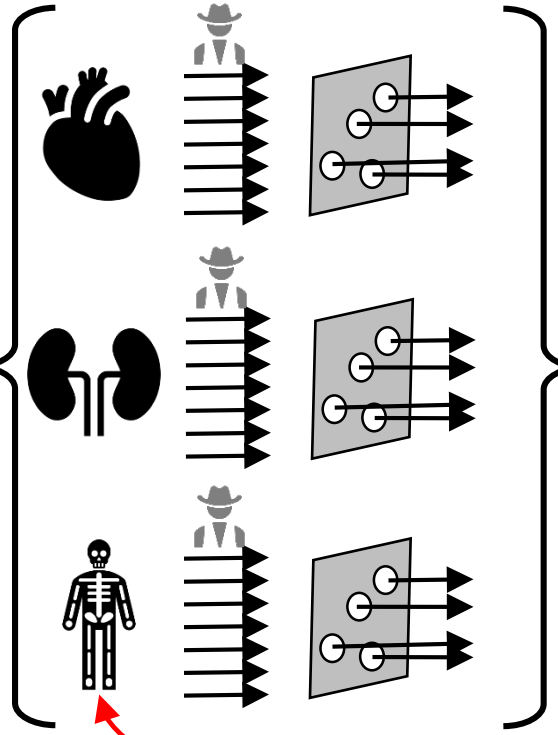
Tissue Registration UI



CCF User Interface (UI)



Multiple TMC's



Global Data Store

CCF Data Store

Data Traceback

PSC (IEC)

User Interface (IU)

End user Interface

Additional Processing

	Data Wrangler
	Data Filter (HIPPA, quality control, ...)

Tools to assist in sample spatial registration

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

Propagate needs back to TMC's

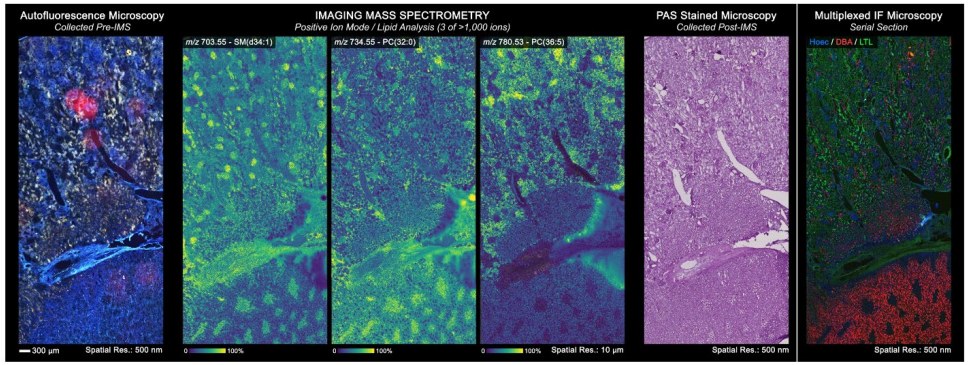
- Only the data needed for the GUI

TMC: Tissue Mapping Center
PSC: Pittsburgh Supercomputing center

Data: Biological

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 states	Subset A
Proliferating cells	S-phase
	G2/M
Cell cyle arrest	G0
	G1/S
	G2/M

Cell type	Subset A	Subset B	Subset C
Tubular Epithelium	Proximal tubular cells	S1	
		S2	
		S3	
	Loop on Henle	Thin descending limg	
		Thin ascending limb	
		Thick limb	medullary
			cortical
		Macula Densa	
	Distal convoluted tubule		
	Connecting segment		
	Collecting duct	Principal cells	
		Intercalated cells	Type A
			Type B
Glomerulus	Epithelium	Visceral	
		Parietal	
	Mesangial cells		
Vasculature	Endothelium	Glomerular	
		Peritubular	
		Lymphatic	
	Pericytes		
	Juxta Glomerular Cells		
Interstitium	Fibroblasts	Myofibroblasts	
		EPO producing cells	
		Medullary fibroblasts	
	Mononuclear cells	Resident macrophages	
		Dendritic cells	
	Lymphocytes	T cells	
		B cells	
		NK cells	

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)

Field # Sort	Field Label Sort	Field Name Sort	Field Units	Field Data	Lookup Tal	Low Value	High Value	Valid value	IsNullable	Parent Fiel	Parent Fiel	Can Child E	ReadOnly Sort
9	Donor //ABO:	abo		char(3)	lkup_abo				TRUE				FALSE
10	Donor //Date of birth:	dob		datetime					TRUE				FALSE
11	Donor //Gender:	gender		char(1)	lkup_gender			M,F	TRUE				FALSE
12	Details //Age:	age_in_months		smallint		0	1188		TRUE			FALSE	FALSE
13	Details //Age Unit:	age_unit		char(1)	lkup_age_unit			M,Y	TRUE	age_in_months			TRUE
14	Details //Height:	hgt_cm	cm	decimal(5, 2)		1	241.3		TRUE				FALSE
15	Donor hgt_ft //	hgt_ft	ft	int		0	7		TRUE				TRUE
16	Donor hgt_in //	hgt_in	in	int		0	11		TRUE				TRUE
17	Details //Weight:	wgt_kg	kg	decimal(7, 4)		0.454	294.835		TRUE				FALSE
18	Donor wgt_lb //	wgt_lb	lbs	decimal(3, 0)		2	650		TRUE				TRUE
19	Donor //Ethnicity/race:	race		bigint	lkup_race_subcat_multi				FALSE				FALSE
30	Details //History of diabe	hist_diabetes		smallint	lkup_histdiab_dur				TRUE				FALSE
31	Donor //History of cance	hist_cancer		smallint	lkup_histcancer_site				TRUE			FALSE	FALSE
32	Donor History of cancer ,	cancer_oth_ostxt		varchar(50)		1	50		TRUE	hist_cance	999		FALSE
33	Details //History of hype	hypertension		smallint	lkup_histhype_dur				TRUE			FALSE	FALSE

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

Heart: Shin Lin, UW

Clinical and Spatial Metadata (21 rows)

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

Data Dictionary (115 rows)

Field # Sort	Field Label Sort	Field Name Sort	Field Units	Field Data	Lookup Tal	Low Value	High Value	Valid value
9	Donor //ABO:	abo		char(3)	lkup_abo			
10	Donor //Date of birth:	dob		datetime				
11	Donor //Gender:	gender		char(1)	lkup_gender			M,F
12	Details //Age:	age_in_months		smallint		0	1188	
13	Details //Age Unit:	age_unit		char(1)	lkup_age_unit			M,Y
14	Details //Height:	hgt_cm	cm	decimal(5, 2)		1	241.3	
15	Donor hgt_ft //	hgt_ft	ft	int		0	7	
16	Donor hgt_in //	hgt_in	in	int		0	11	
17	Details //Weight:	wgt_kg	kg	decimal(7, 4)		0.454	294.835	
18	Donor wgt_lb //	wgt_lb	lbs	decimal(3, 0)		2	650	
19	Donor //Ethnicity/race:	race		bigint	lkup_race_subcat_multi			
30	Details //History of diabe	hist_diabetes		smallint	lkup_histdiab_dur			
31	Donor //History of cance	hist_cancer		smallint	lkup_histcancer_site			
32	Donor History of cancer ,	cancer_oth_ostxt		varchar(50)		1	50	
33	Details //History of hype	hypertension		smallint	lkup_histhype_dur			

Data: CCF

Minimum Information Standard

hubmapconsortium / ccf-data-wiki Private

Unwatch 3 | Star 0 | Fork 0

Code | Issues 0 | Pull requests 0 | Wiki | Insights

Home

Edit | New Page

Bruce Herr II edited this page 2 days ago · 5 revisions

Welcome to the CCF Data Wiki!

Organ	CalTech	UCSD	Stanford	Vanderbilt	Florida
Heart	✓				
Kidney		✓		✓	
Bowel			✓		
Thymus					✓
Spleen					✓
Lymph Nodes					✓
Lung		✓			
Bladder		×			
Colon			×		

Legend:

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

+ Add a custom footer

Pages 12

Links

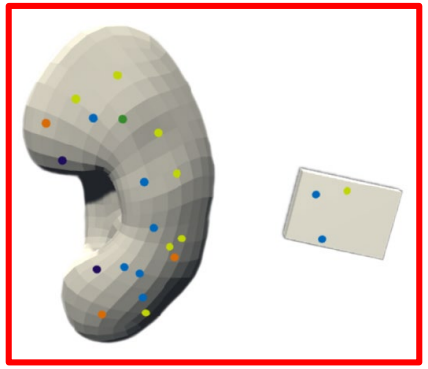
- Home

Templates

- TMC-Organ-Template

Clone this wiki locally

https://github.com/hubmapcc



https://github.com/hubmapconsortium/ccf-data-wiki/wiki

Data: TMCs x Organs x Data Types x Technologies

BUKMAP, Zhang Group

Organs (10)

1. Bladder
2. Blood Vessel (Heart)
3. Breast
4. Colon
5. Kidney
6. Liver
7. Lung
8. Spleen
9. Thymus
10. Tonsil

No Bone Marrow and Pancreas.

Data Types (13)	**Technologies (~25)**
1. Imaging - Proteins	CODEX;DART-FISHrp;IF;IHC;LRET-IF;MALDI Imaging MS;PER-DEI
2. Imaging - RNA	DART-FISH;LRET-ISH;MERFISH;PER-DEI;seqFISH;smFISH
3. Imaging - DNA	PER-DEI
4. Imaging - Other	Lipid and Metabolite MALDI Imaging MS
5. scRNAseq	snDropseq;scRNAseq
6. scDNAseq	scATACseq;scTHSseq;SNAREseq
7. scProteomics	IMC
8. bulk-Proteomics	LC-MS/MS
9. bulk-RNA	?
10. bulk-DNA	?
11. Metabolomics	LC-MS/MS;nano-POTS
12. Lipids	LC-MS/MS;nano-DESI
13. Other	Autofluorescence;PAS stained microscopy

3D Spatial Models of Organs - Individual Differences

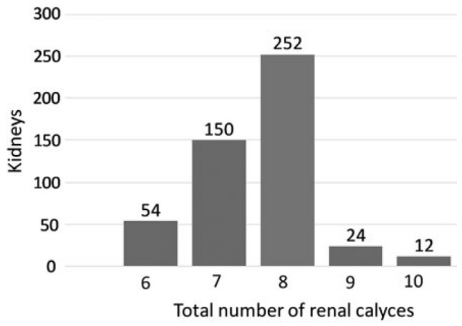
Example: Kidney

Takazawa et al. analyzed a total of 492 kidneys to identify differences across individuals.

Within HuBMAP, we need to decide what level of detail is required to serve HuBMAP data use cases.

In Year 2, we plan to

- Work closely with TMCs to identify and review existing literature on organ specific individual differences—across sex, ethnicity, age groups, etc.
- Compile recommendations on how to represent individual differences.



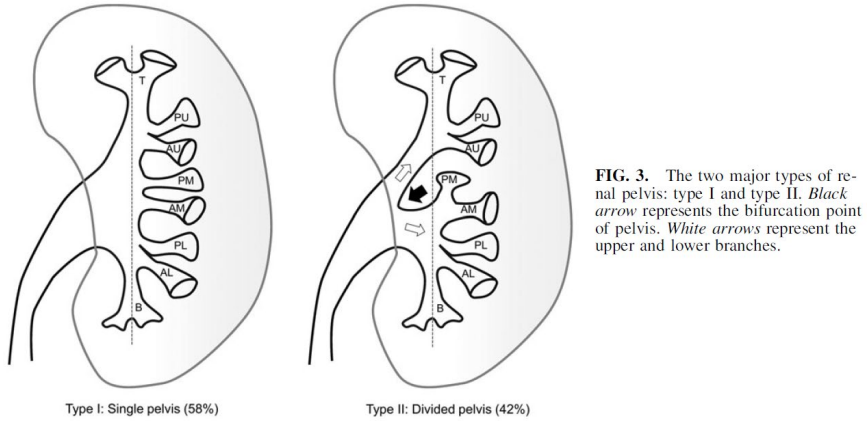


FIG. 3. The two major types of renal pelvis: type I and type II. *Black arrow* represents the bifurcation point of pelvis. *White arrows* represent the upper and lower branches.

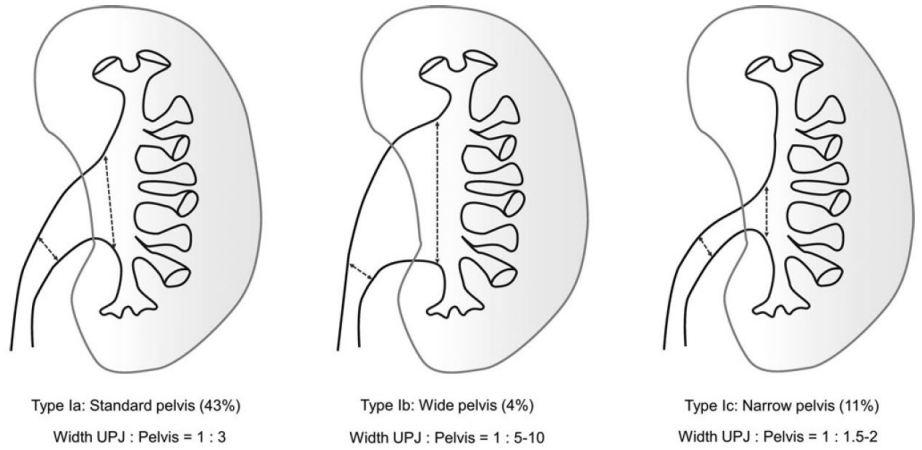


FIG. 4. The three subtypes of type I: type Ia, type Ib, and type Ic.

Takazawa et al. (2018) Proposal for a Simple Anatomical Classification of the Pelvicaliceal System for Endoscopic Surgery. *JOURNAL OF ENDOUROLOGY*, 32:8, 753–758.

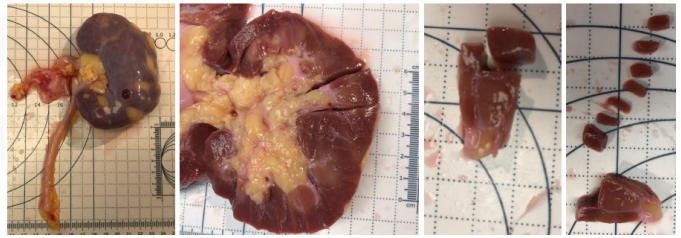
3D Spatial Models interlinked with terminology/ontology

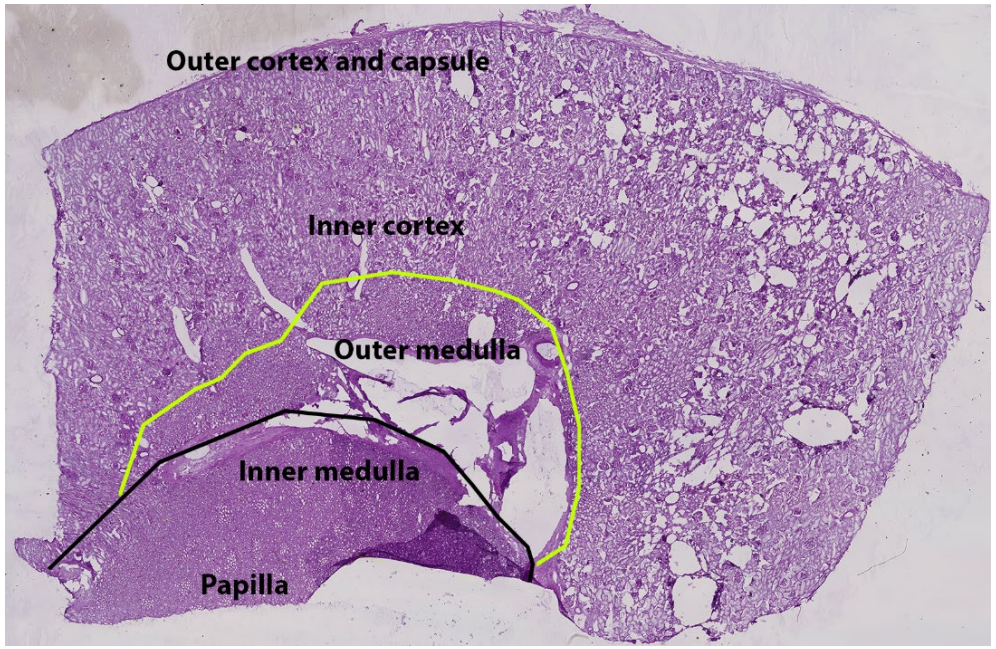
Kidney: Jeff Spraggins et al., VU

See data on Globus, BIOMIC_patient-64354

- BIOMIC_patient-64354_clinical_and_spatial_metadata.xlsx
- BIOMIC_patient-64354_data_guide.pptx
- BIOMIC_patient-64354_overview.png
- BIOMIC_patient-64354_Sample-20-Histology.tif
- neg_ion_mode_section
- pos_ion_mode_section

BUKMAP, Zhang Group

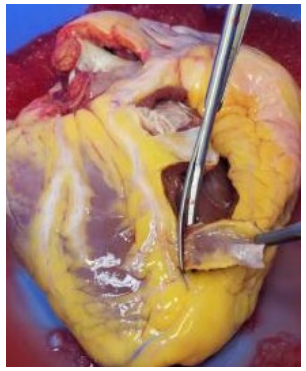




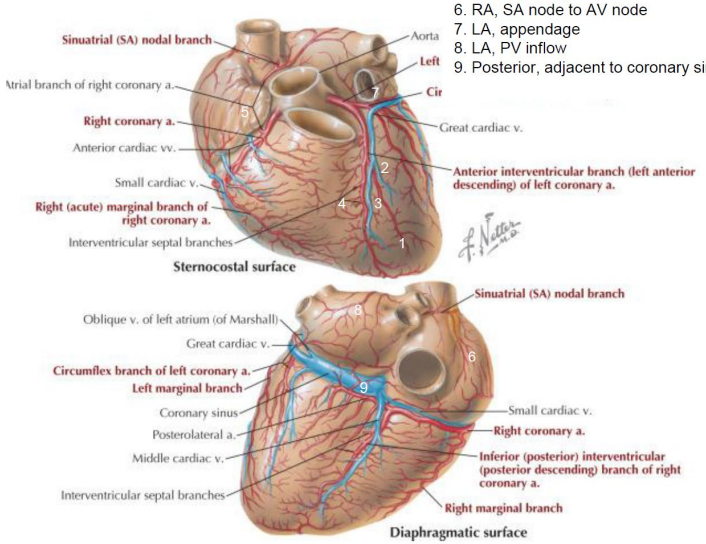
Heart: Shin Lin, UW

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

Terminology; Coordinates and photos to spatialize



Sites	Distinctive features
1. LV, apex	
2. LV, free wall 3 cm from apex	
3. septum, 3cm from apex including LAD	major arterial vessel, Purkinje fiber CM
4. RV, free wall 3 cm from apex	
5. RA appendage	
6. RA, SA node to AV node	pacemaker CM
7. LA, appendage	
8. LA, PV inflow	
9. Posterior, adjacent to coronary sinus	major venous vessel



3D Spatial Models interlinked with terminology/ontology

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

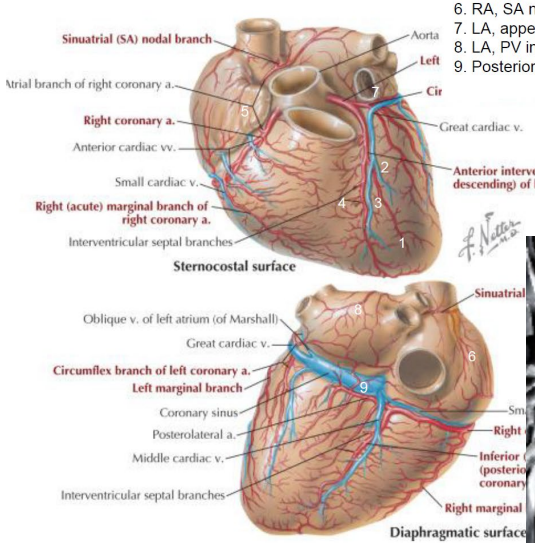
Virtual tissue samples will be sized 1-2cm cubed, numbered (1 … 9), and oriented (left-right, top and bottom tissue slice of z-stack).

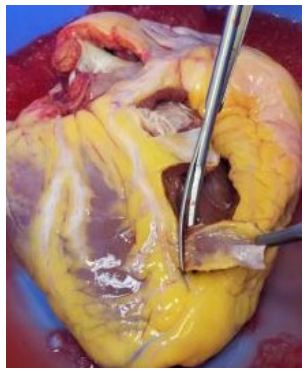
Measure error from

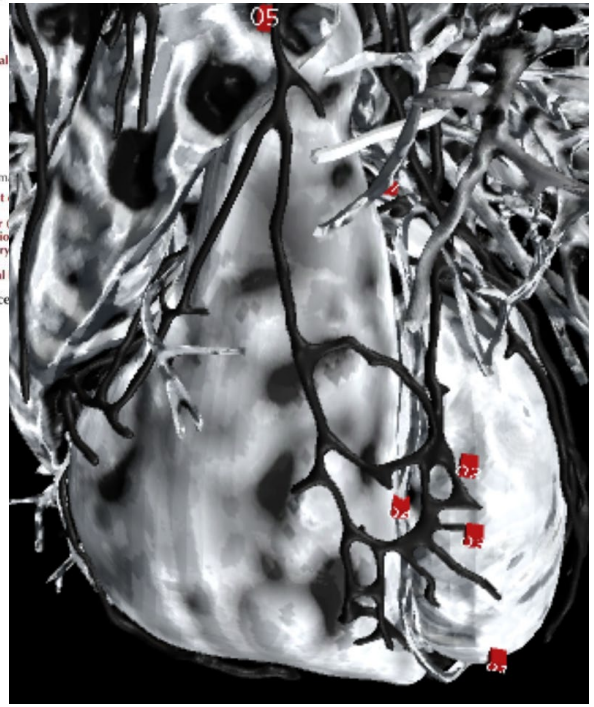
- precision of tissue sample procurement and
- placement in the 3D browser

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.

Sites	Distinctive features
1. LV, apex	
2. LV, free wall 3 cm from apex	
3. septum, 3cm from apex including LAD	major arterial vessel, Purkinje fiber CM
4. RV, free wall 3 cm from apex	
5. RA appendage	
6. RA, SA node to AV node	pacemaker CM
7. LA, appendage	
8. LA, PV inflow	
9. Posterior, adjacent to coronary sinus	major venous vessel





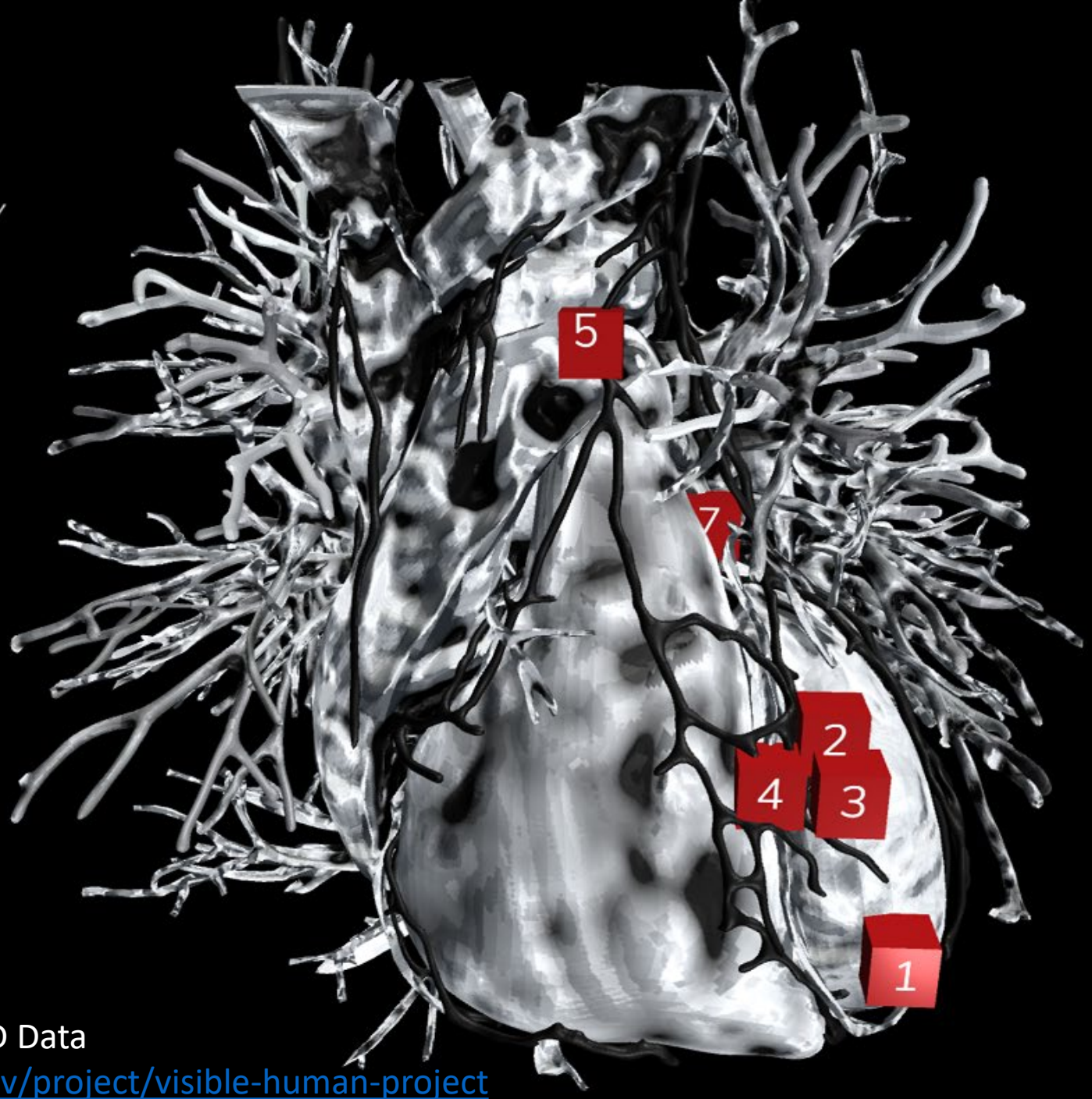


Human heart with data overlay
Developer: Andreas Bueckle

Show/hide

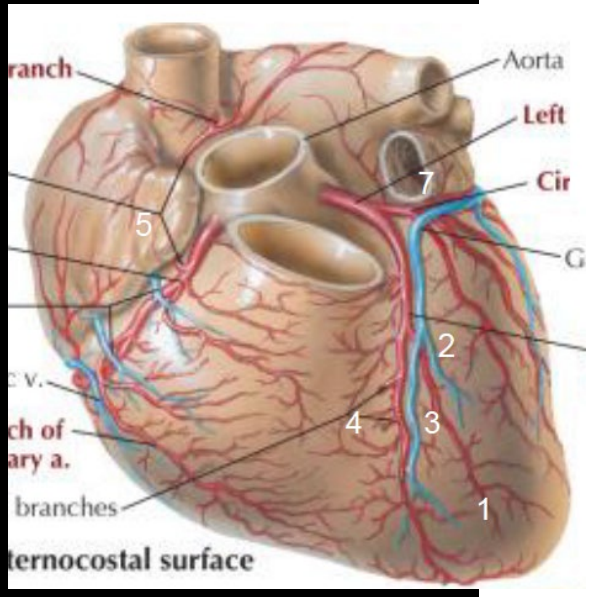
- [x] Coronary arteries
- [x] Coronary veins
- [x] Left atrium
- [x] Left ventricle
- [x] Right atrium
- [x] Right ventricle
- [x] Markers

Adjust camera speed



Currently Selected

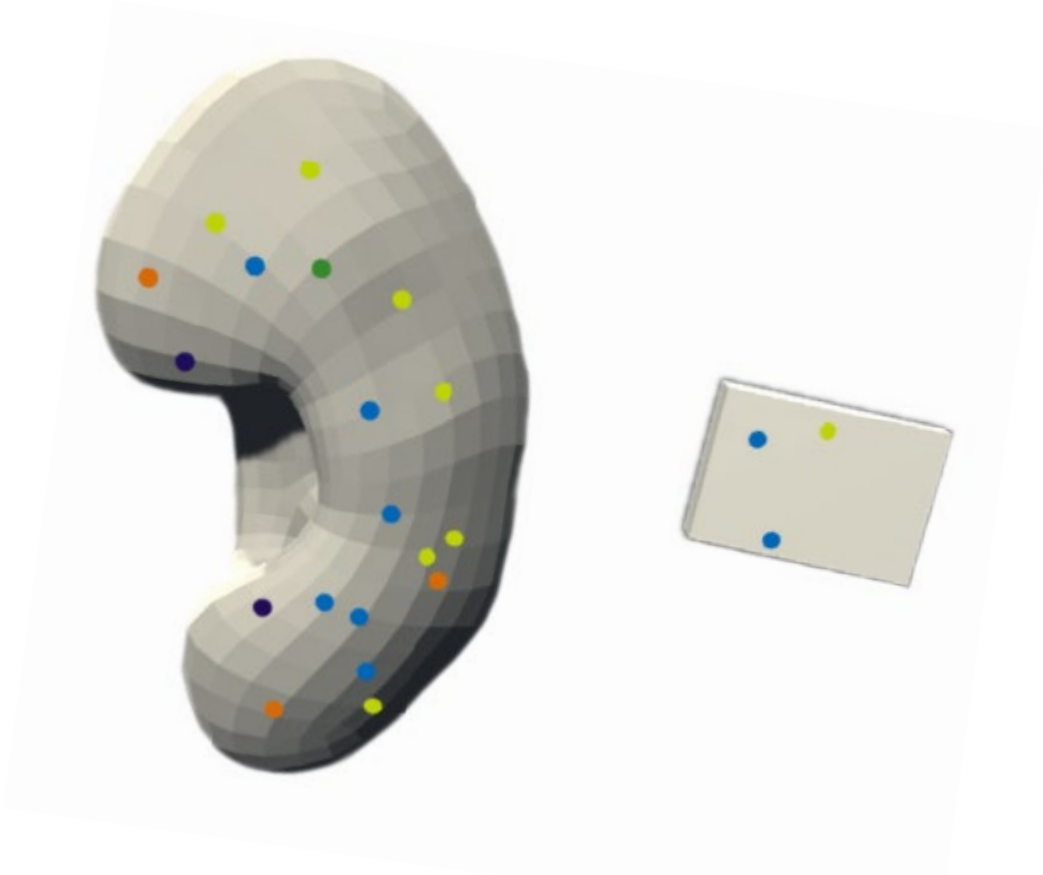
Please click any of the red markers!



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

CCF Tissue Registration Interface

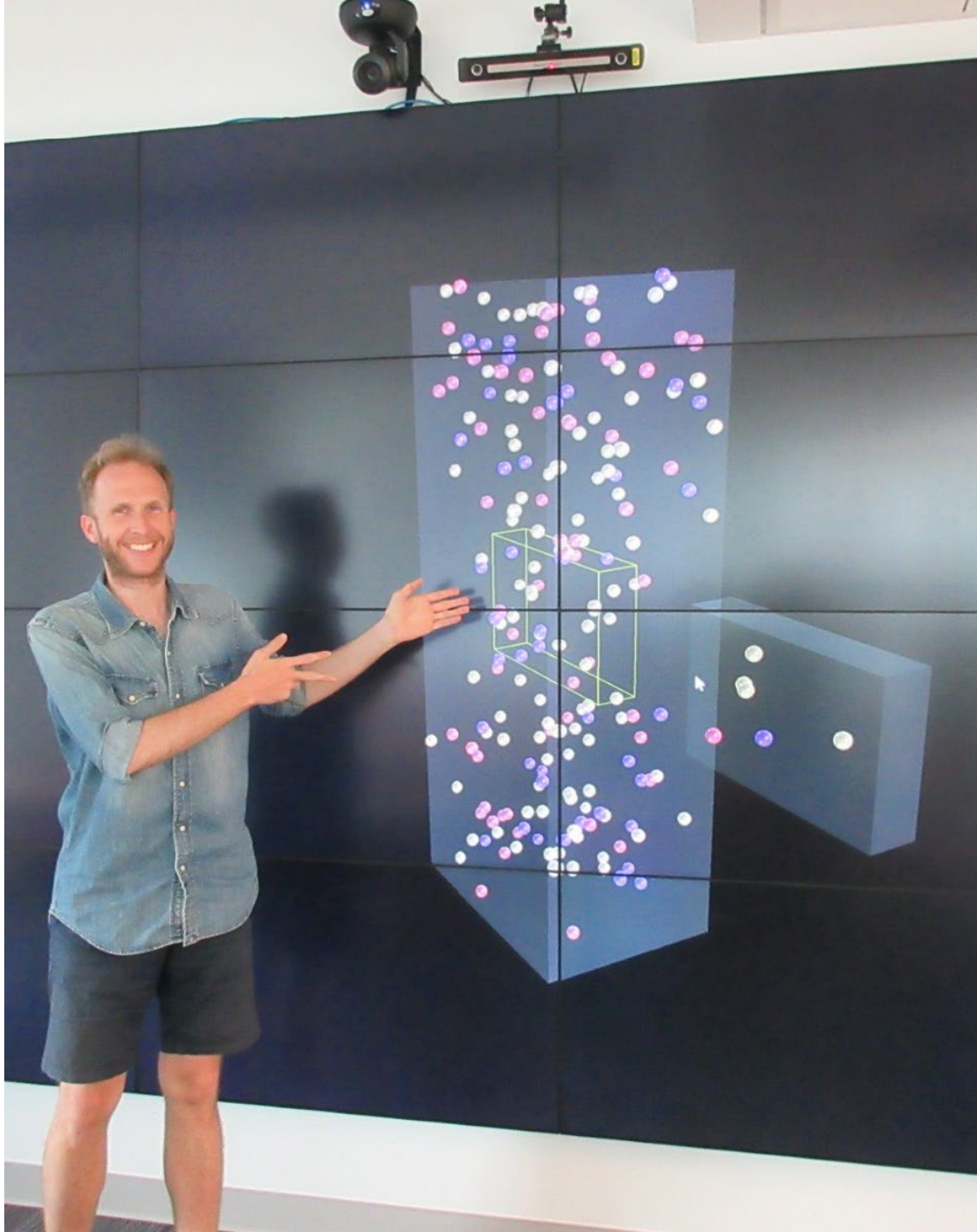
- Exploit human pattern recognition and fine motor skills (by surgeons) to register tissue in organs.
- Add info on anatomical landmarks, cell types, molecular data to support alignment.
- LATER: Use human alignment data as training data for machine learning algorithms, to better support manual alignment OR to possibly fully automatize alignment.



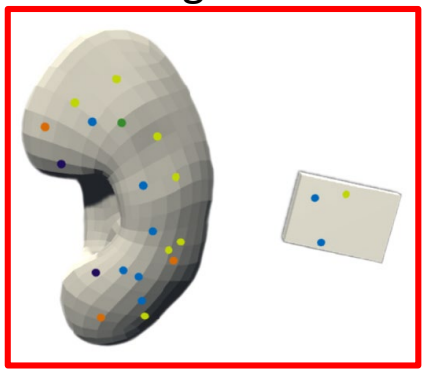
Kidney model from NLM3D Data
https://lhncbc.nlm.nih.gov/project/visible-human-project

CCF Tissue Registration Interface

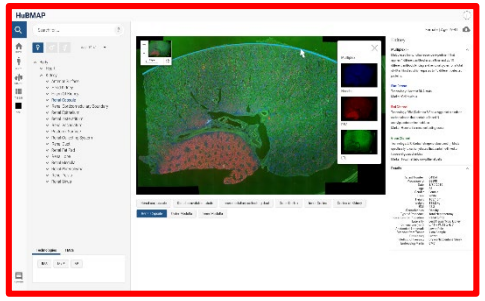
- Exploit human pattern recognition and fine motor skills (by surgeons) to register tissue in organs.
- Add info on anatomical landmarks, cell types, molecular data to support alignment.
- LATER: Use human alignment data as training data for machine learning algorithms, to better support manual alignment OR to possibly fully automatize alignment.



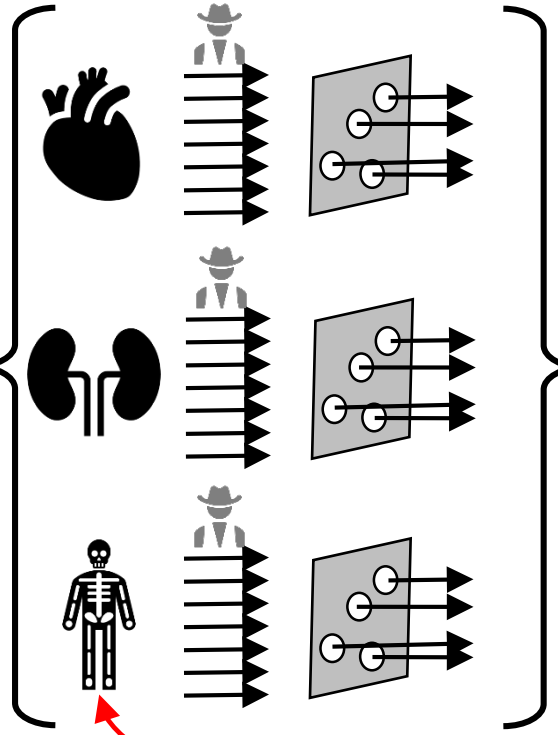
Tissue Registration UI



CCF User Interface (UI)



Multiple TMC's



Global Data Store

CCF Data Store

Data Traceback

PSC (IEC)

User Interface (IU)

End user Interface

Additional Processing

	Data Wrangler
	Data Filter (HIPPA, quality control, ...)

Tools to assist in sample spatial registration

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

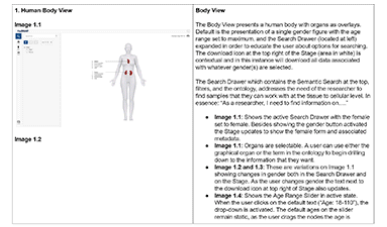
Propagate needs back to TMC's

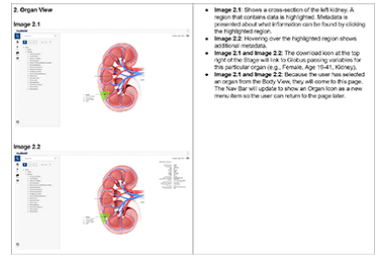
- Only the data needed for the GUI

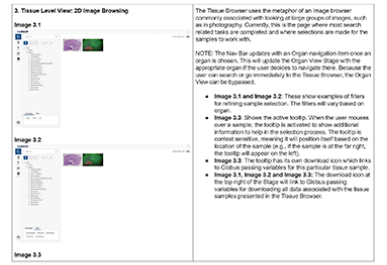
TMC: Tissue Mapping Center
PSC: Pittsburgh Supercomputing center

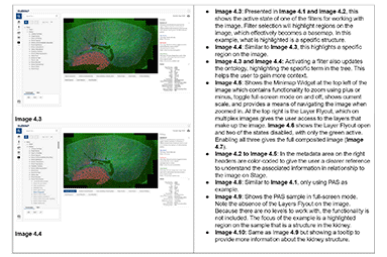
CCF User Interface for semantic + spatial search, filter, review, download

CCF UI Spec v0.5.0 v2 (Initial Writeup) https://tinyurl.com/y4ywy95t









The initial 'user story' features a researcher interested to search for, filter out, review, and download biomolecular data in the context of the whole human body or in spatial relationship to specific organ(s), tissues, or cell types. The researcher is also able to learn more about how the data was acquired, to connect with data authors, and to submit questions and comments on the CCF UI.

Consequently, the UI will support

- Search (ontology supported search) and filter by ontology, anatomy, and metadata
- Visual browsing of tissue samples and metadata at the whole body, organ, tissue, and cell level
- Connect with data authors to inquire about technology details.
- Data download at the whole body, organ, tissue, and cell level.
- Submit questions and comments on the CCF UI.

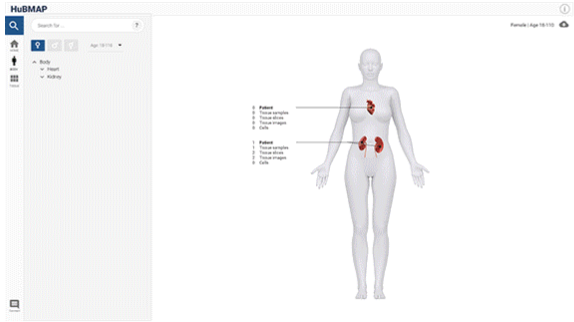
In the initial 9m, proof of concept versions of the whole body, organ, tissue and cell level views will be implemented.

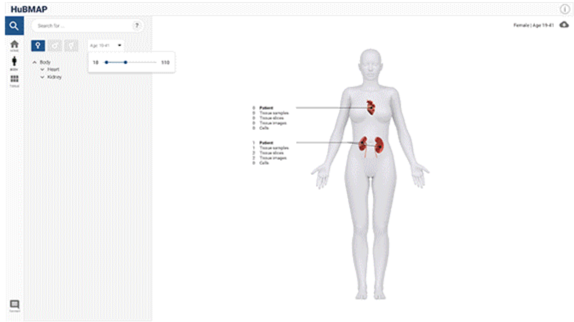
The github repo is at: https://github.com/hubmapconsortium/ccf-ui

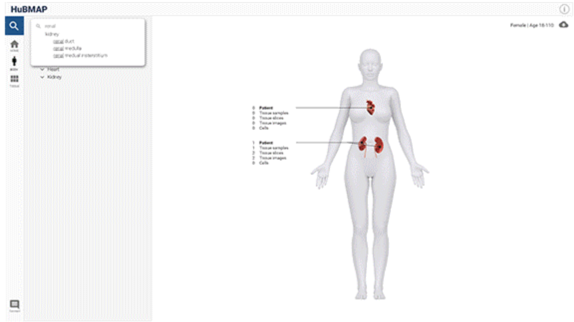
The deployed release build after 6w of development is at: https://hubmapconsortium.github.io/ccf-ui/

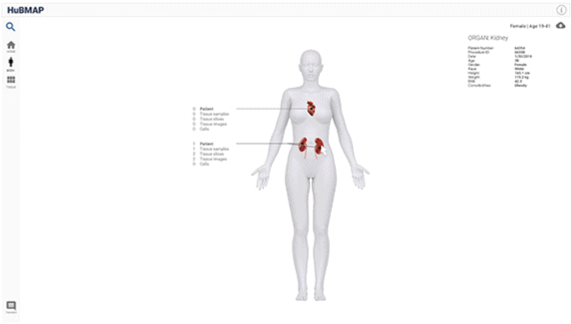
CCF User Interface for semantic + spatial search, filter, review, download

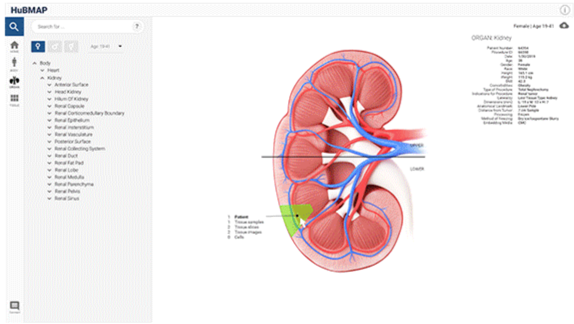
CCF UI Spec v0.5.0 interface mockups (PDF) https://tinyurl.com/y2d43zds

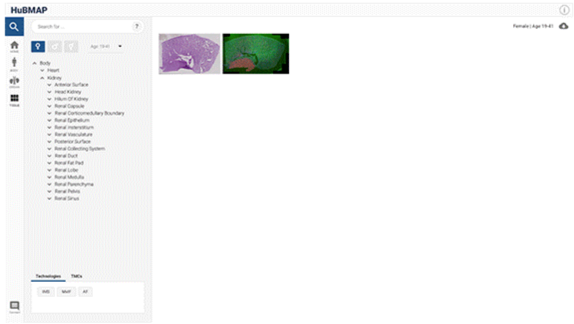


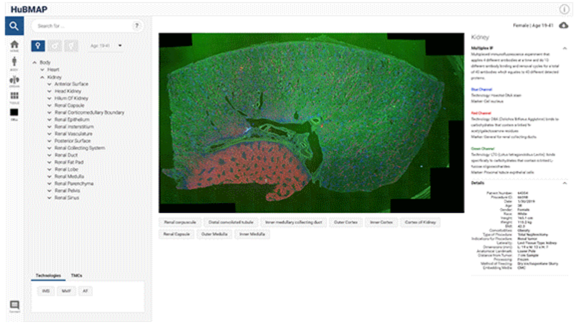


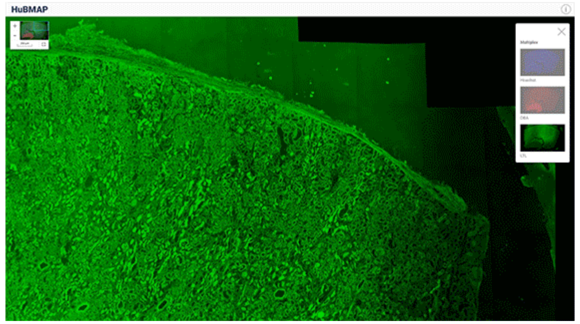


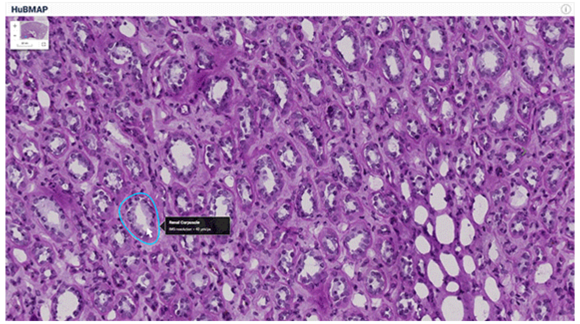












HuBMAP

Semantic Search

Search for ...

Age: 19-41

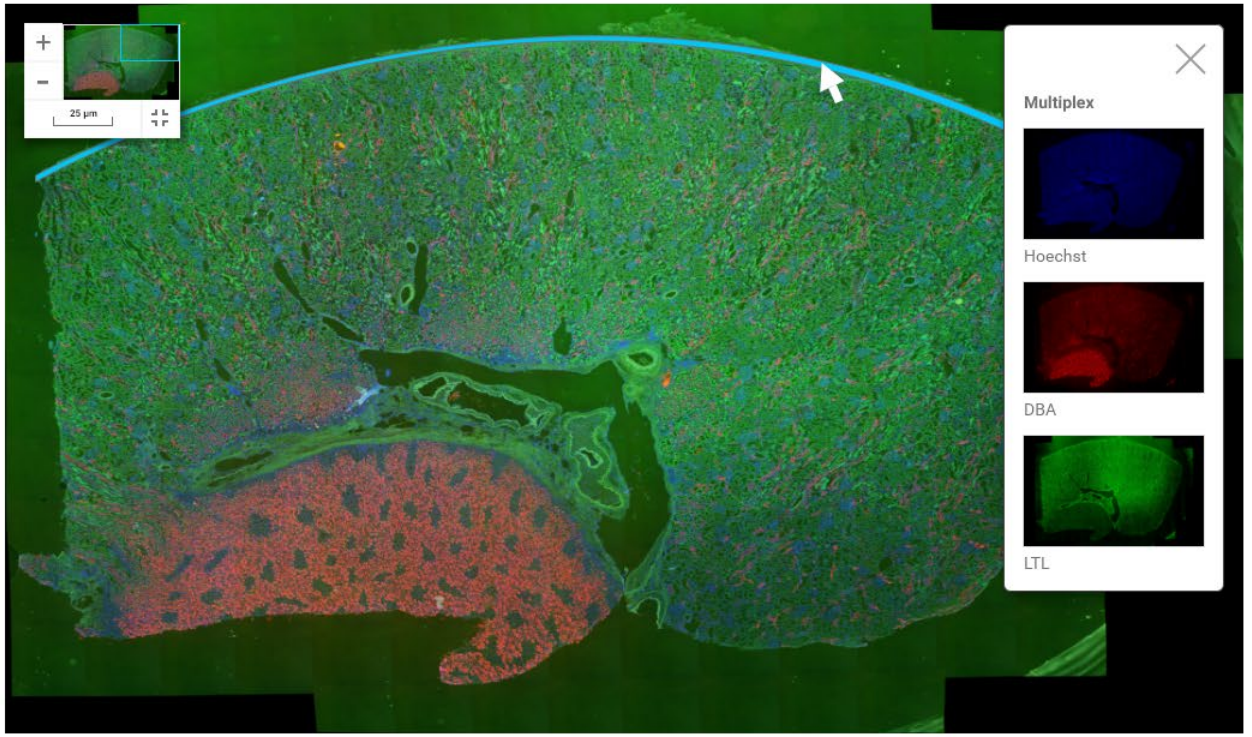
- Body
  - Heart
  - Kidney
    - Anterior Surface
    - Head Kidney
    - Hilum Of Kidney
    - Renal Capsule
    - Renal Corticomedullary Boundary
    - Renal Epithelium
    - Renal Insterstitium
    - Renal Vasculature
    - Posterior Surface
    - Renal Collecting System
    - Renal Duct
    - Renal Fat Pad
    - Renal Lobe
    - Renal Medulla
    - Renal Parenchyma
    - Renal Pelvis
    - Renal Sinus

Semantic Browsing

Semantic Filters

Technologies | TMCs

IMS | MxIF | AF



Renal corpuscule | Distal convoluted tubule | Inner medullary collecting duct | Outer Cortex | Inner Cortex | Cortex of Kidney

Renal Capsule | Outer Medulla | Inner Medulla

Female | Age 19-41

Kidney

Multiplex IF

Multiplexed immunofluorescence experiment that applies 4 different antibodies at a time and do 10 different antibody binding and removal cycles for a total of 40 antibodies which equates to 40 different detected proteins.

Blue Channel
Technology: Hoechst DNA stain
Marker: Cell nucleus

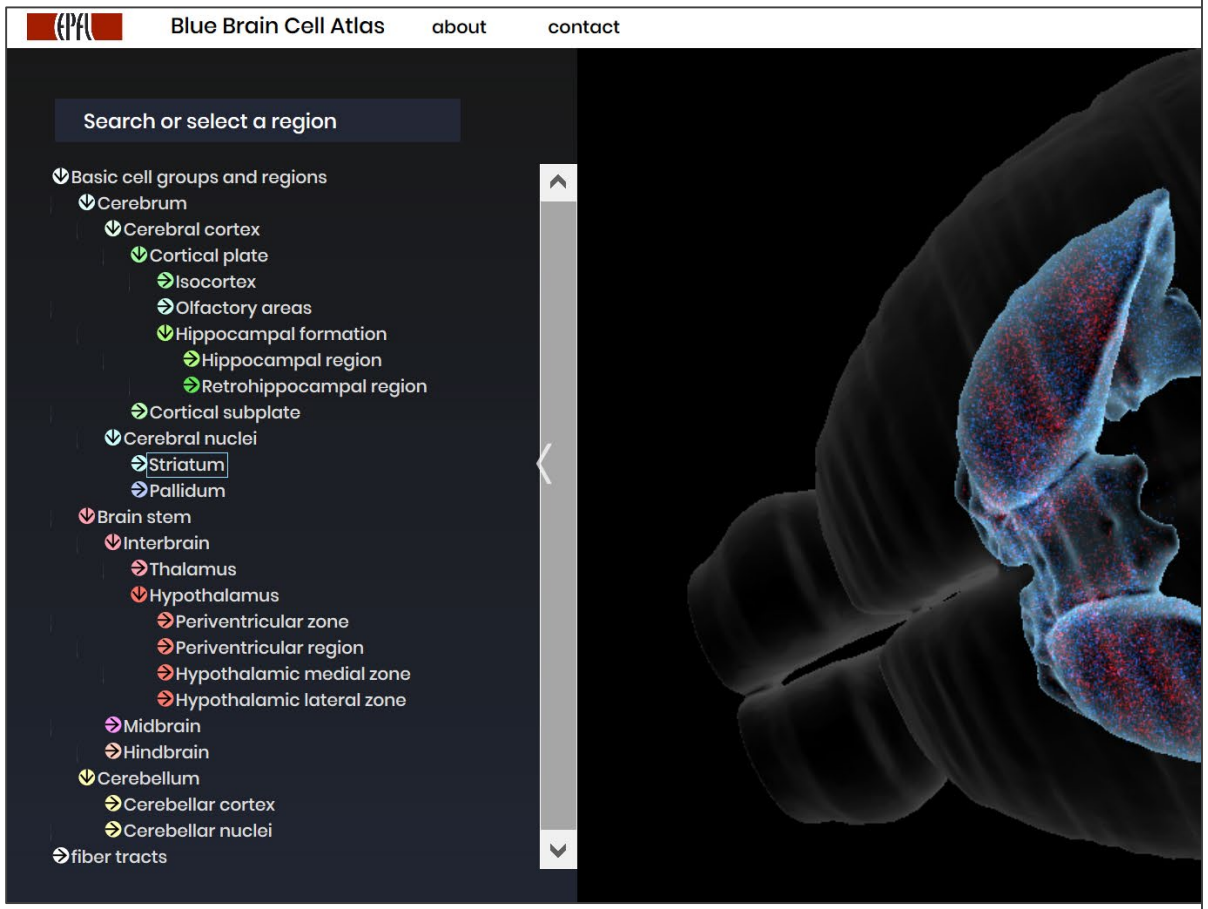
Red Channel
Technology: DBA (Dolichos Biflorus Agglutinin) binds to carbohydrates that contain a-linked N-acetylgalactosamine residues
Marker: General for renal collecting ducts.

Green Channel
Technology: LTO (Lotus tetragonolobus Lectin) binds specifically to carbohydrates that contain α-linked L-fucose oligosaccharides
Marker: Proximal tubule expithelial cells

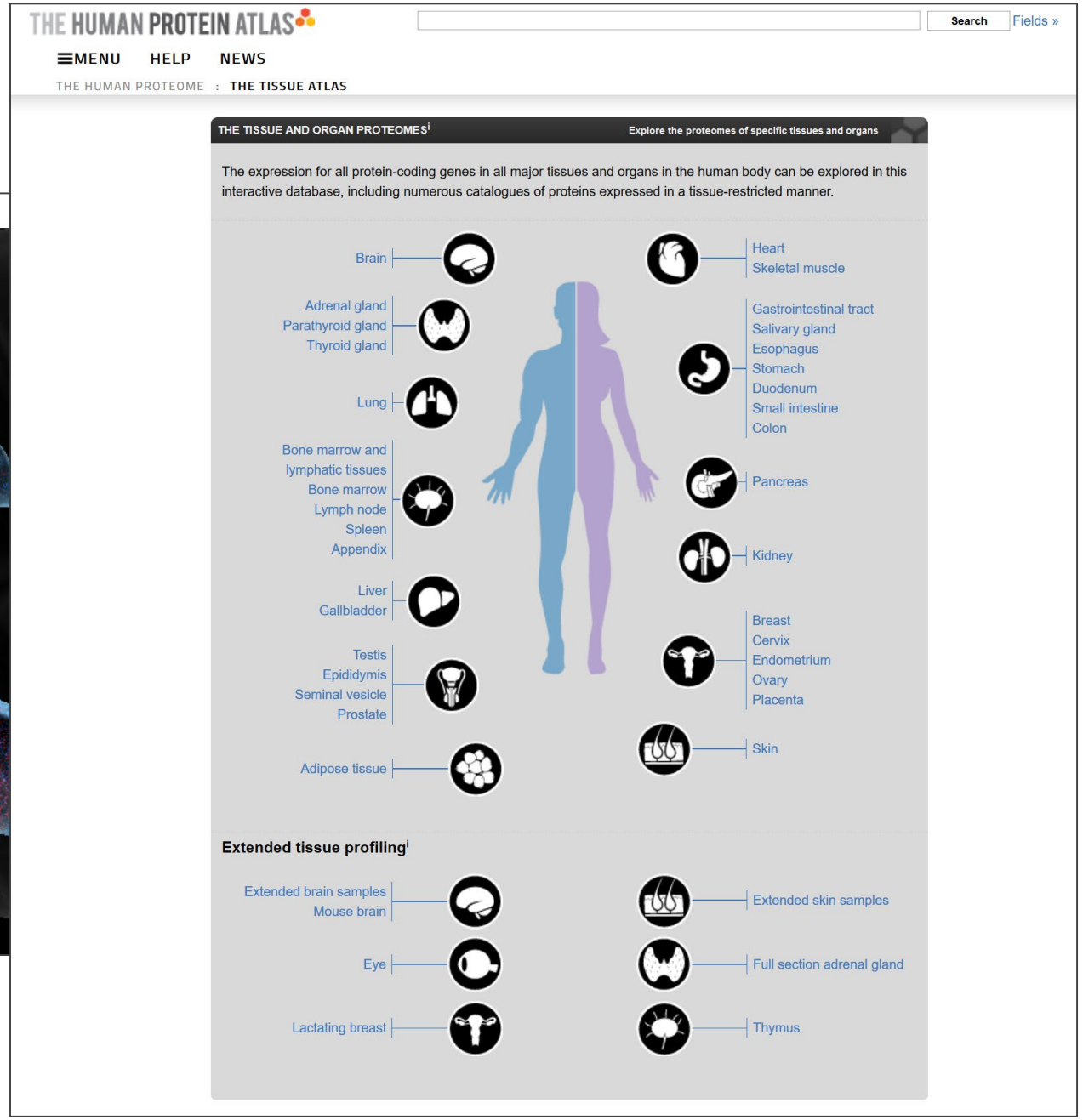
Details

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

Prior Work



https://bbp.epfl.ch/nexus/cell-atlas



https://www.proteinatlas.org/humanproteome/tissue

Year 2 Plans (June 21, 2019 - June 20, 2020)

- Develop, test, optimize different **Tissue Registration UI**, optimized for HuBMAP organs.
- Evaluate and enhance functionality of **CCF User Interface (UI)**.
- Use the CCF UI to serve a **Kidney Micro Atlas**.
- Run **user studies** for CCF and Registration UI.
- Research and develop a **Visual Human Massive Open Online Course (VHMOOC)** hat helps communicate the
  - quality and coverage of HuBMAP data,
  - utility and proper usage of CCF UI and HuBMAP tools, and
  - demonstrate new single-cell analysis and mapping techniques.
- Host another **CCF Workshop** in collaboration with NYGC in DC.

Open Questions

- What datasets (in what unified formats) will become available when? Which will be included in first HuBMAP data release in summer 2020?
- What tasks do existing ontologies perform well? Where do they fall short (e.g., uncertainty, variability).
- What tasks do existing user interfaces perform well? Where do they fall short?