Kidney Metadata and Ontology Design (HuBMAP)

Katy Börner, Leonard Cross, Samuel H. Friedman, Randy Heiland, Bruce Herr II, Paul Macklin, Lisel Record, Ellen Quardokus, James Sluka, Griffin Weber

underlined: most directly involved in ontology development

Intelligent Systems Engineering
Indiana University

May 9, 2019

Motivation

Mapping Components (MC):
Spatial maps of biomolecular data

Given anatomical and molecular data, develop and validate:

1) Terminologies/Ontologies (<u>Semantics</u>)

- Reference concepts, e.g., organs, organ parts, cell types, cell states
- Fiduciary concepts: Well-defined landmarks that can be provided by TMCs and used by MC to spatially orient data with respect to 3D structures

2) 3D <u>Spatial</u> Models interlinked with terminology/ontology

- Across levels (gross anatomy/organ, tissue, cell level) using hierarchical containment to localize the sample within the body
- Make landmarks visible in 3D models

3) Interface for <u>semantic</u> and <u>spatial</u> search, filter, review, download of data.

- Use ontology for query expansion (elastic search), semantic browsing, and as controlled vocabulary (e.g., turning on/off male/female or different cell states).
- Use 3D models for spatial browsing, confirmation of proper tissue registration, exploring cell context, etc.

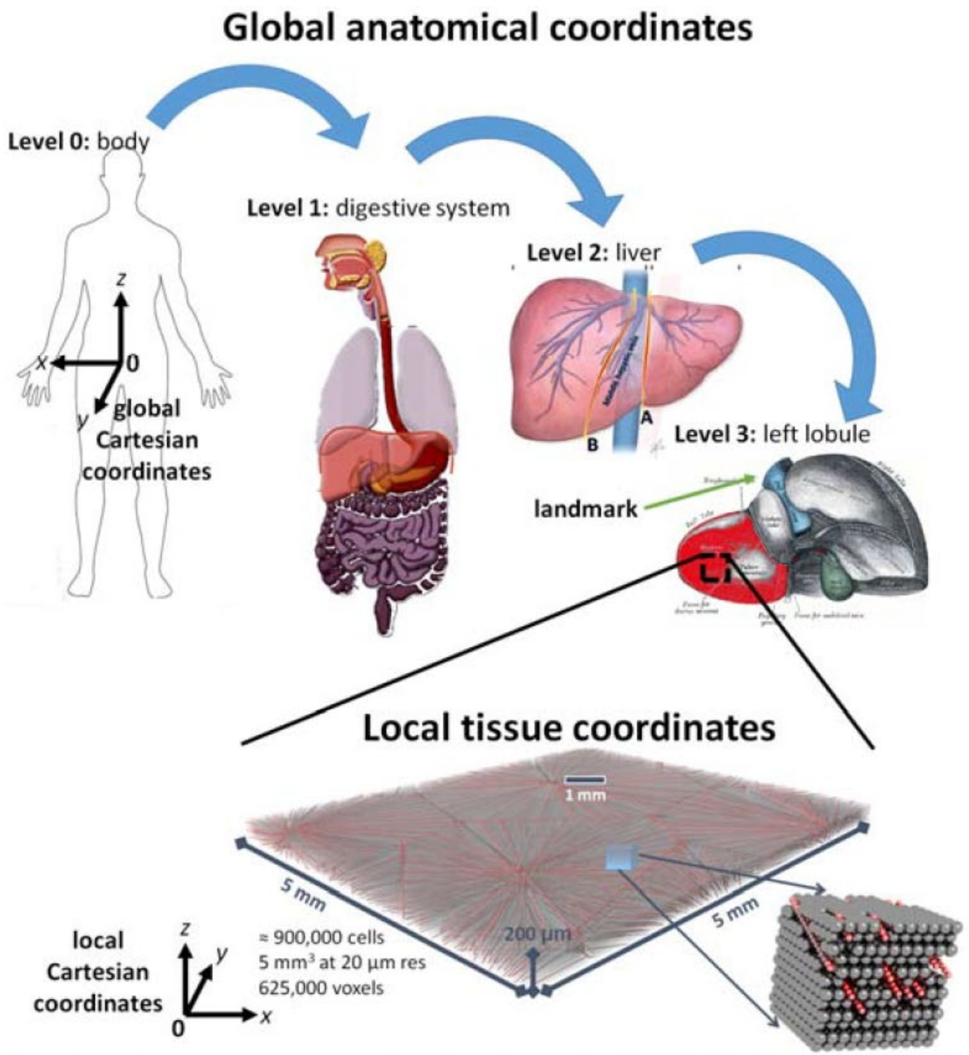
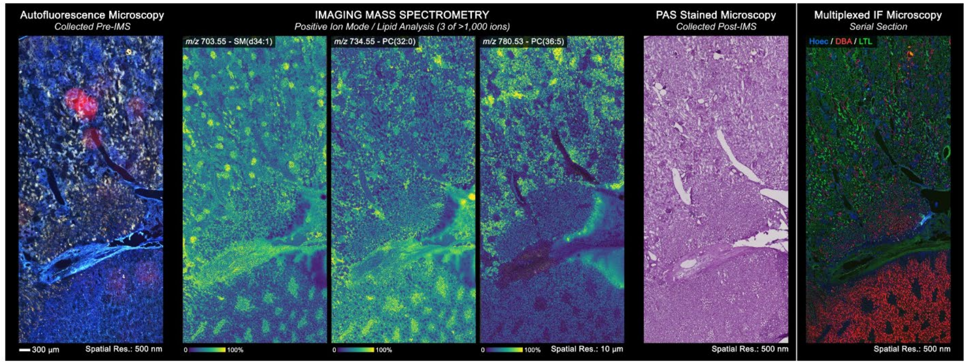


Fig. 5. CCF concept, navigating through the global anatomical coordinate system to insert a synthetic tissue sample (from PhysiCell[4]) into the left liver lobe with a local coordinate system.

We must plan for heterogeneous data

Kidney: Jeff Spraggins et al., VU

See data on Globus, BIOMIC_patient-64354



Clinical and Spatial Metadata (21 rows)

Cell Types, on right

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	

Cell States (9 rows)

Cell states	Subset A
Proliferating cells	S-phase
	G2/M
Cell cyle 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)

Field # Sort	Field Label Sort	Field Name Sort	Field Units	Field Data '	Lookup Tal	Low Value	High Value	Valid value	IsNullable S	Parent Fiel	Parent Fiel	Can Child b	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 hyper	hypertension		smallint	lkup_histhype_dur				TRUE			FALSE	FALSE

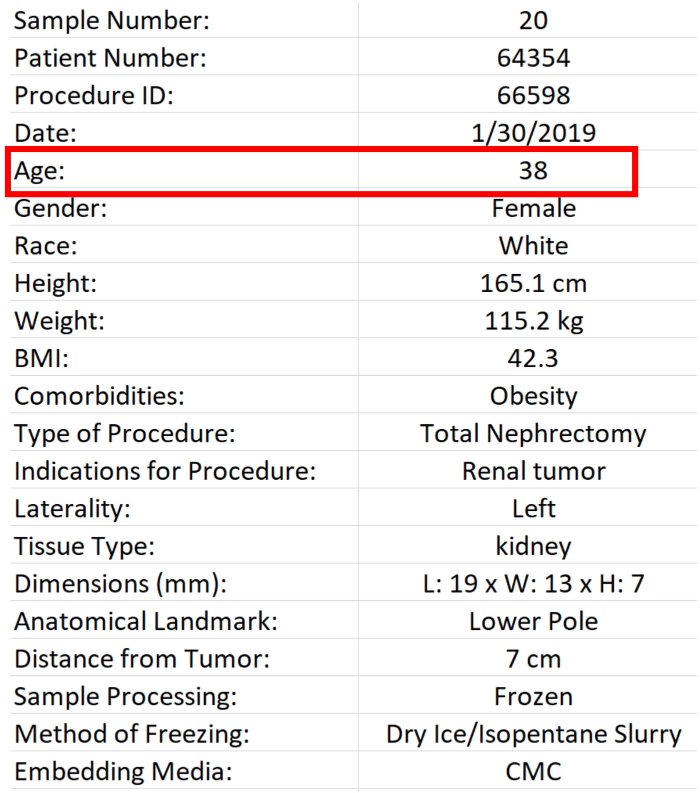
Cell Types (14)

endothelial cells	
	arterial
	capillary
	venous
	lymphatic
cardiomyocytes	
	atrial
	ventricular
	nodal
fibroblasts	
	fibroblasts
	myofibroblasts
immune cells	
	macrophages

Data

Kidney: Jeff Spraggins et al., VU

Clinical and Spatial Metadata (21 rows)

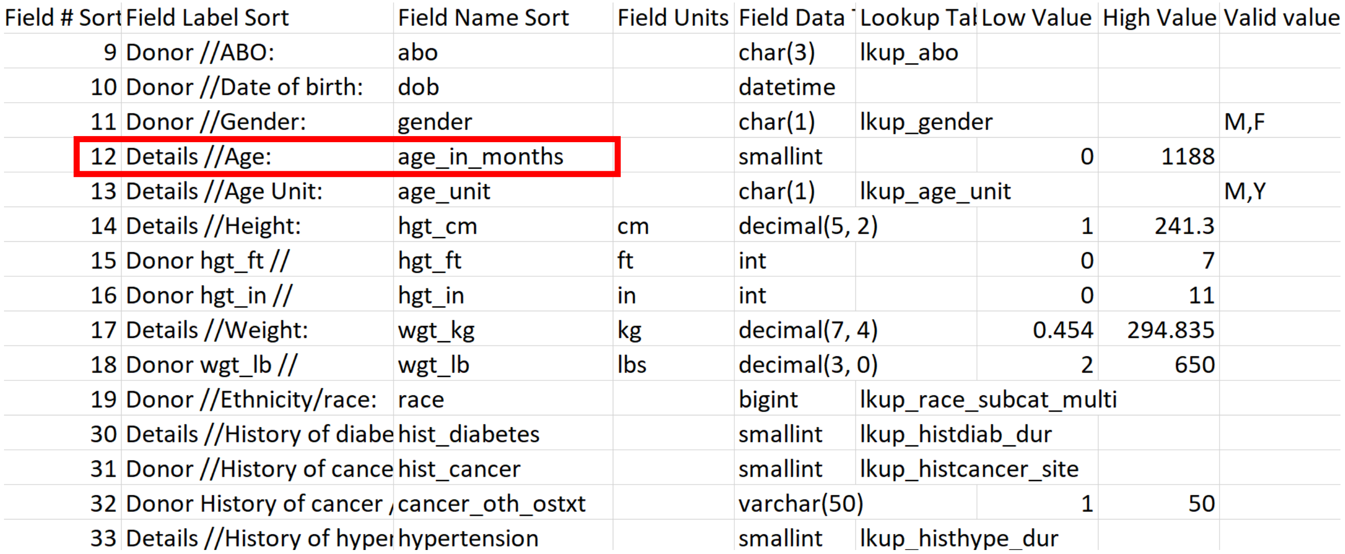


Sample Number:	20
Patient Number:	64354
Procedure ID:	66598
Date:	1/30/2019
Age:	38
Gender:	Female
Race:	White
Height:	165.1 cm
Weight:	115.2 kg
BMI:	42.3
Comorbidities:	Obesity
Type of Procedure:	Total Nephrectomy
Indications for Procedure:	Renal tumor
Laterality:	Left
Tissue Type:	kidney
Dimensions (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

Jeff invited feedback on current fields and format.

Heart: Shin Lin, UW

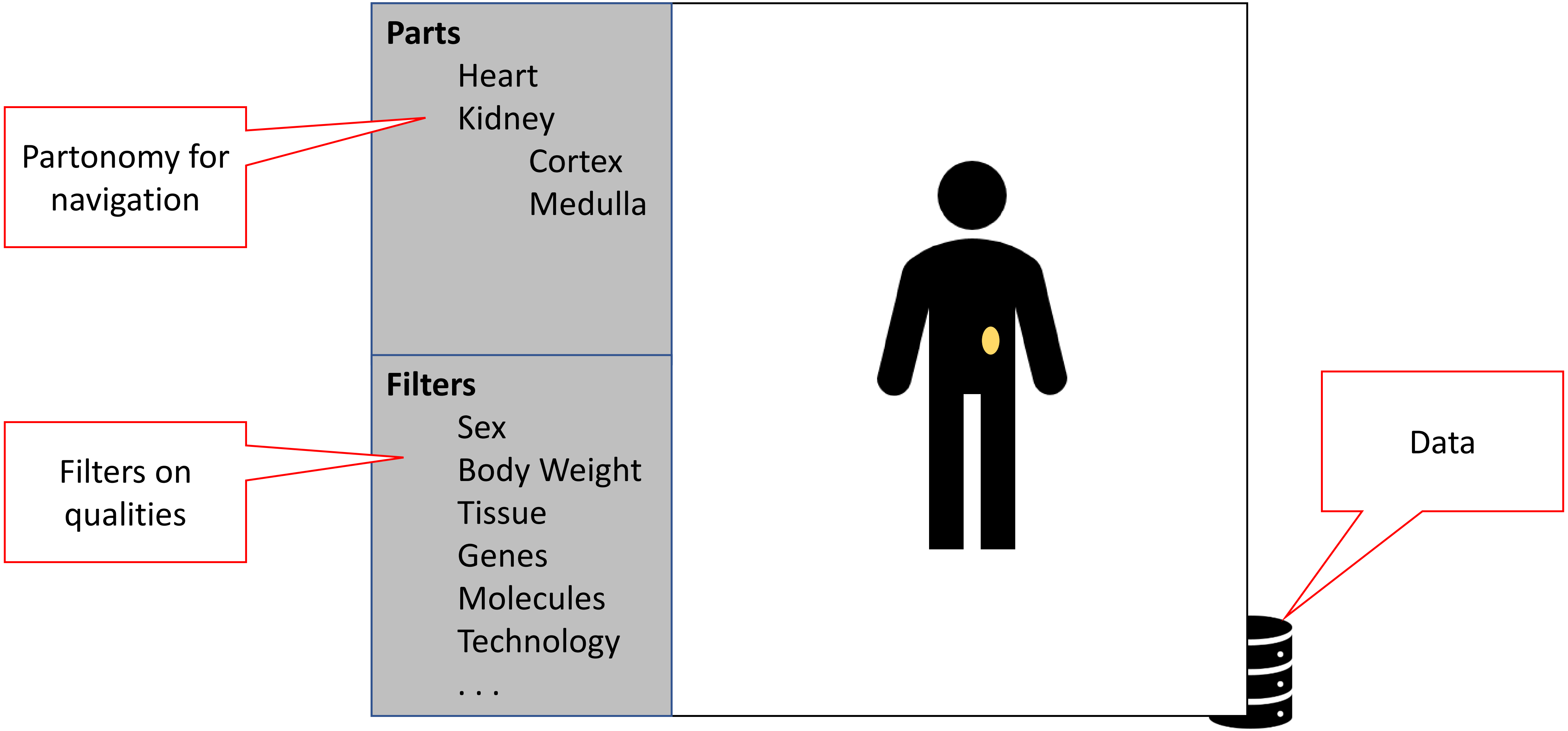
Data Dictionary (115 rows)



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 hyper	hypertension		smallint	lkup_histhype_dur			

Please complete **TMC Landmarks Survey** at
https://goo.gl/forms/x9F8cP1GIzprDxbI2
(complete one survey per organ)

Goal: Facilitate navigation of multiscale data



Partonomy for navigation

Parts
- Heart
- Kidney
  - Cortex
  - Medulla

Filters on qualities

Filters
- Sex
- Body Weight
- Tissue
- Genes
- Molecules
- Technology
- . . .

Data

Overall CCF Approach

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
- May need separate partOf (or class/subclass) 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)

(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

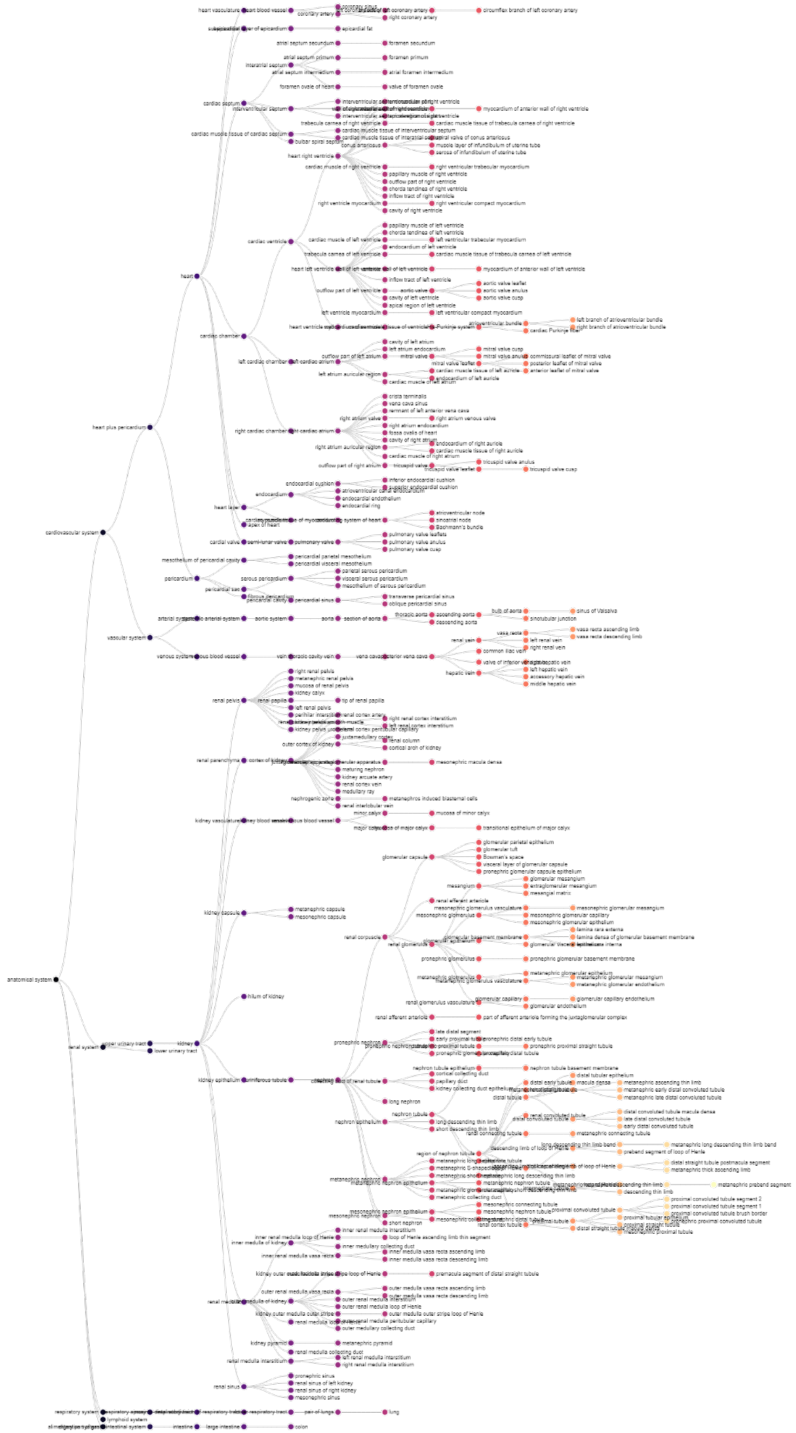
- Basic Formal Ontology (BFO)
- Information Artifact Ontology (IAO)
- Ontology of units of Measure (OM)
- Provenance, Authoring and Versioning ontology (PAV)
- VIVO (Identifying researchers)

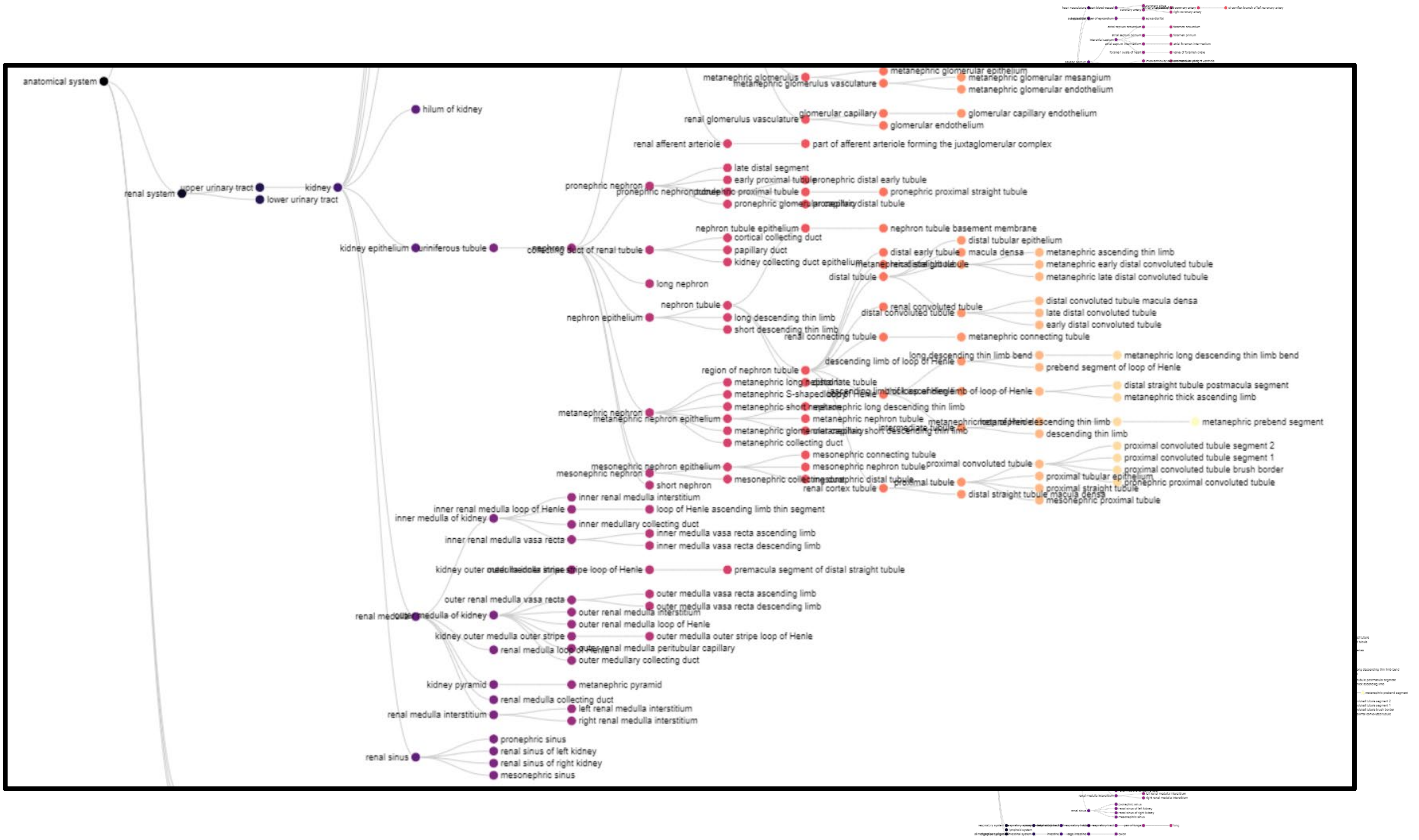
MeSH and NCI Thesaurus

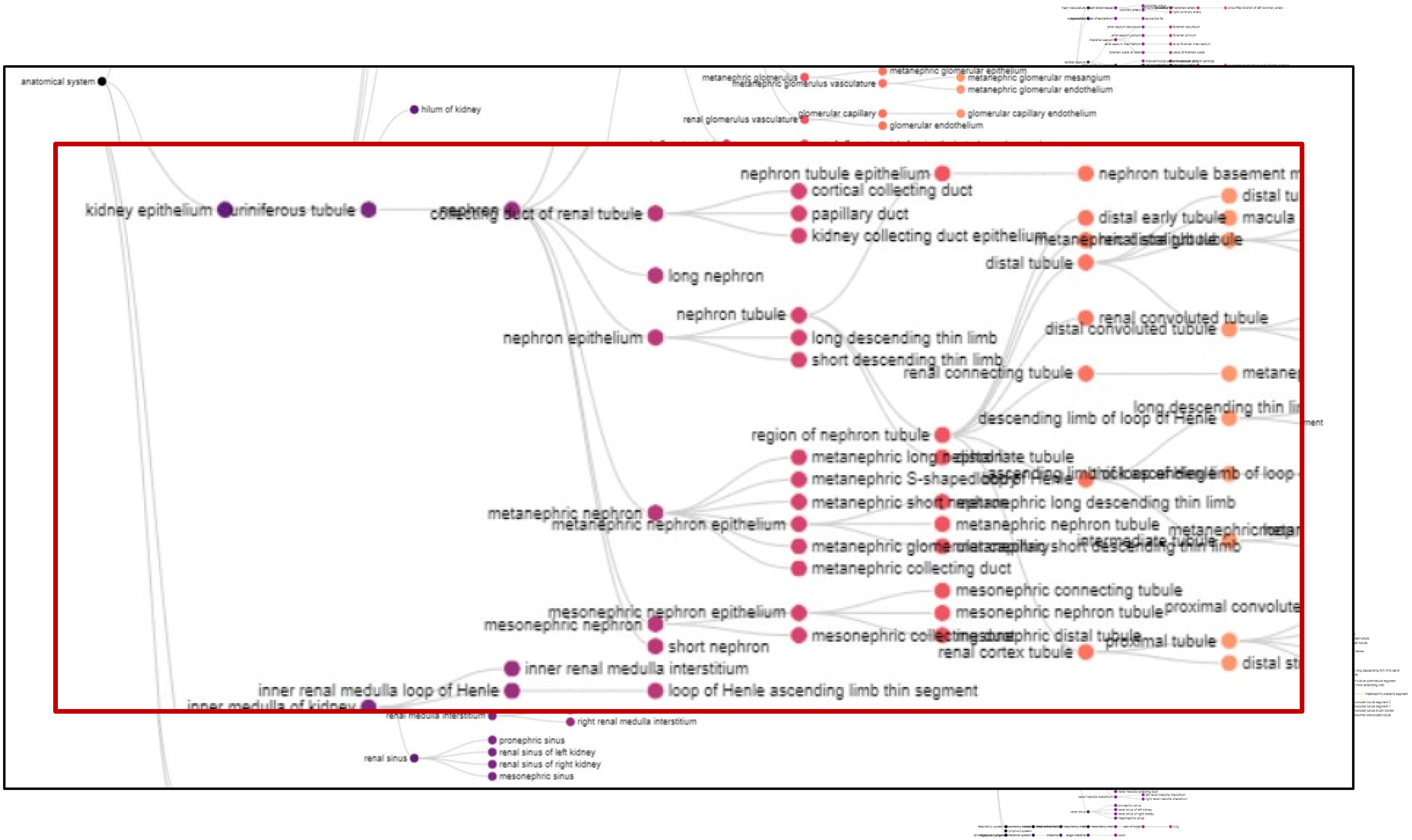
Ontology

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]()


heart vasculature
coronary sinus
coronary artery
right coronary artery
circumflex branch of left coronary artery
epicardial fat
atrial septum secundum
foramen secundum
atrial septum primum
foramen primum
interatrial septum
atrial septum intermedium
atrial foramen intermedium
foramen ovale of heart
valve of foramen ovale
cardiac septum
interventricular septum
myocardium of anterior wall of right ventricle
trabecula carnea of right ventricle
cardiac muscle tissue of trabecula carnea of right ventricle
cardiac muscle tissue of cardiac septum
cardiac muscle tissue of interventricular septum
bulbar spiral septum
conus arteriosus
muscle layer of infundibulum of uterine tube
serosa of infundibulum of uterine tube
heart right ventricle
cardiac muscle of right ventricle
right ventricular trabecular myocardium
papillary muscle of right ventricle
outflow part of right ventricle
chorda tendinea of right ventricle
inflow tract of right ventricle
right ventricle myocardium
right ventricular compact myocardium
cavity of right ventricle
papillary muscle of left ventricle
chorda tendinea of left ventricle
cardiac muscle of left ventricle
left ventricular trabecular myocardium
cardiac ventricle
endocardium of left ventricle
trabecula carnea of left ventricle
cardiac muscle tissue of trabecula carnea of left ventricle
heart
heart left ventricle
myocardium of anterior wall of left ventricle
inflow tract of left ventricle
aortic valve leaflet
outflow part of left ventricle
aortic valve
aortic valve anulus
cavity of left ventricle
aortic valve cusp
apical region of left ventricle
left ventricle myocardium
left ventricular compact myocardium
heart ventricle
Purkinje system
atrioventricular bundle
left branch of atrioventricular bundle
right branch of atrioventricular bundle
cardiac Purkinje fiber
cardiac chamber
cavity of left atrium
left atrium endocardium
mitral valve cusp
outflow part of left atrium
mitral valve
mitral valve anulus
commissural leaflet of mitral valve
left cardiac chamber
left cardiac atrium
mitral valve leaflet
posterior leaflet of mitral valve
left atrium auricular region
cardiac muscle tissue of left auricle
anterior leaflet of mitral valve
endocardium of left auricle
cardiac muscle of left atrium
crista terminalis
vena cava sinus
remnant of left anterior vena cava
right atrium valve
right atrium venous valve
heart plus pericardium
right atrium endocardium
right cardiac chamber
right cardiac atrium
fossa ovalis of heart
cavity of right atrium
right atrium auricular region
endocardium of right auricle
cardiac muscle tissue of right auricle
cardiac muscle of right atrium
outflow part of right atrium
tricuspid valve
tricuspid valve anulus
tricuspid valve leaflet
tricuspid valve cusp
endocardial cushion
inferior endocardial cushion
superior endocardial cushion
endocardium
atrioventricular canal endocardium
endocardial endothelium
endocardial ring
heart layer
conducting system of heart
atrioventricular node
apex of heart
sinoatrial node
Bachmann's bundle
cardiovascular system
pulmonary valve leaflets
cardial valve
semi-lunar valve
pulmonary valve
pulmonary valve anulus
pulmonary valve cusp
mesothelium of pericardial cavity
pericardial parietal mesothelium
pericardial visceral mesothelium
parietal serous pericardium
pericardium
serous pericardium
visceral serous pericardium
mesothelium of serous pericardium
pericardial sac
fibrous pericardium
pericardial cavity
pericardial sinus
transverse pericardial sinus
oblique pericardial sinus
arterial system
systemic arterial system
aortic system
aorta
section of aorta
thoracic aorta
ascending aorta
bulb of aorta
sinus of Valsalva
descending aorta
sinotubular junction
vasa recta
vasa recta ascending limb
vasa recta descending limb
vascular system
renal vein
left renal vein
right renal vein
common iliac vein
venous system
venous blood vessel
vein
thoracic cavity vein
vena cava
posterior vena cava
valve of inferior vena cava
left hepatic vein
hepatic vein
accessory hepatic vein
middle hepatic vein
right renal pelvis
metanephric renal pelvis
mucosa of renal pelvis
kidney calyx
renal pelvis
renal papilla
tip of renal papilla
left renal pelvis
renal cortex artery
right renal cortex interstitium
left renal cortex interstitium
kidney pelvis urothelium
renal cortex peritubular capillary
juxtamedullary cortex
renal column
outer cortex of kidney
cortical arch of kidney
renal parenchyma
cortex of kidney
juxtaglomerular apparatus
mesonephric macula densa
maturing nephron
kidney arcuate artery
renal cortex vein
medullary ray
nephrogenic zone
metanephros induced blastemal cells
renal interlobular vein
minor calyx
mucosa of minor calyx
kidney vasculature
kidney blood vessel
venous blood vessel
major calyx
mucosa of major calyx
transitional epithelium of major calyx
glomerular parietal epithelium
glomerular tuft
glomerular capsule
Bowman's space
visceral layer of glomerular capsule
pronephric glomerular capsule epithelium
glomerular mesangium
mesangium
extraglomerular mesangium
mesangial matrix
renal efferent arteriole
mesonephric glomerulus vasculature
mesonephric glomerular mesangium
mesonephric glomerulus
mesonephric glomerular capillary
kidney capsule
metanephric capsule
mesonephric capsule
mesonephric glomerular epithelium
lamina rara externa
renal corpuscle
glomerular basement membrane
lamina densa of glomerular basement membrane
renal glomerulus
glomerular epithelium
glomerular visceral epithelium
lamina rara interna
pronephric glomerulus
pronephric glomerular basement membrane
metanephric glomerulus
metanephric glomerular epithelium
metanephric glomerular mesangium
metanephric glomerulus vasculature
metanephric glomerular endothelium
anatomical system
glomerular capillary
glomerular capillary endothelium
hilum of kidney
renal glomerulus vasculature
glomerular endothelium
renal afferent arteriole
part of afferent arteriole forming the juxtaglomerular complex
late distal segment
early proximal tubule
pronephric distal early tubule
pronephric nephron
pronephric proximal tubule
pronephric proximal straight tubule
upper urinary tract
kidney
renal system
lower urinary tract
nephron tubule epithelium
nephron tubule basement membrane
cortical collecting duct
distal tubular epithelium
kidney epithelium
uriniferous tubule
nephron
collecting duct of renal tubule
papillary duct
distal early tubule
macula densa
metanephric ascending thin limb
kidney collecting duct epithelium
metanephric early distal convoluted tubule
distal tubule
metanephric late distal convoluted tubule
long nephron
renal convoluted tubule
distal convoluted tubule macula densa
nephron tubule
distal convoluted tubule
late distal convoluted tubule
nephron epithelium
long descending thin limb
early distal convoluted tubule
short descending thin limb
renal connecting tubule
metanephric connecting tubule
long descending thin limb bend
metanephric long descending thin limb bend
descending limb of loop of Henle
prebend segment of loop of Henle
region of nephron tubule
metanephric S-shaped body
distal straight tubule postmacula segment
metanephric thick ascending limb
metanephric nephron
metanephric long descending thin limb
metanephric nephron epithelium
metanephric nephron tubule
metanephric prebend segment
descending thin limb
metanephric collecting duct
proximal convoluted tubule segment 2
mesonephric connecting tubule
proximal convoluted tubule segment 1
mesonephric nephron epithelium
mesonephric nephron tubule
proximal convoluted tubule
proximal convoluted tubule brush border
mesonephric nephron
mesonephric distal tubule
proximal tubular epithelium
renal cortex tubule
proximal tubule
proximal straight tubule
metanephric proximal convoluted tubule
short nephron
distal straight tubule
mesonephric proximal tubule
inner renal medulla interstitium
inner renal medulla loop of Henle
loop of Henle ascending limb thin segment
inner medulla of kidney
inner medullary collecting duct
inner medulla vasa recta ascending limb
inner renal medulla vasa recta
inner medulla vasa recta descending limb
premacula segment of distal straight tubule
outer medulla vasa recta ascending limb
outer renal medulla vasa recta
outer medulla vasa recta descending limb
renal medulla
outer medulla of kidney
outer renal medulla interstitium
outer renal medulla loop of Henle
kidney outer medulla outer stripe
outer medulla outer stripe loop of Henle
renal medulla loop of Henle
outer renal medulla peritubular capillary
outer medullary collecting duct
kidney pyramid
metanephric pyramid
renal medulla collecting duct
left renal medulla interstitium
renal medulla interstitium
right renal medulla interstitium
pronephric sinus
renal sinus of left kidney
renal sinus
renal sinus of right kidney
mesonephric sinus
respiratory system
respiratory tract
pair of lungs
lung
lymphoid system
intestine
large intestine
colon


anatomical system
renal system
upper urinary tract
lower urinary tract
kidney
hilum of kidney
kidney epithelium
uriniferous tubule
nephron
collecting duct of renal tubule
metanephric glomerulus
metanephric glomerulus vasculature
metanephric glomerular epithelium
metanephric glomerular mesangium
metanephric glomerular endothelium
renal glomerulus vasculature
glomerular capillary
glomerular capillary endothelium
glomerular endothelium
renal afferent arteriole
part of afferent arteriole forming the juxtaglomerular complex
pronephric nephron
pronephric nephron tubule
late distal segment
early proximal tubule
pronephric distal early tubule
pronephric proximal tubule
pronephric proximal straight tubule
pronephric glomerular capillary
pronephric distal tubule
nephron tubule epithelium
nephron tubule basement membrane
cortical collecting duct
papillary duct
kidney collecting duct epithelium
metanephric distal tubule
distal tubular epithelium
distal early tubule
macula densa
metanephric ascending thin limb
metanephric early distal convoluted tubule
metanephric late distal convoluted tubule
distal straight tubule
distal tubule
long nephron
nephron tubule
nephron epithelium
long descending thin limb
short descending thin limb
renal convoluted tubule
distal convoluted tubule
distal convoluted tubule macula densa
late distal convoluted tubule
early distal convoluted tubule
renal connecting tubule
metanephric connecting tubule
long descending thin limb bend
metanephric long descending thin limb bend
descending limb of loop of Henle
prebend segment of loop of Henle
region of nephron tubule
metanephric long nephron
metanephric S-shaped body
metanephric short nephron
intermediate tubule
ascending limb of loop of Henle
thick ascending limb of loop of Henle
distal straight tubule postmacula segment
metanephric thick ascending limb
metanephric long descending thin limb
metanephric nephron tubule
metanephric nephron
metanephric nephron epithelium
metanephric glomerular capillary
metanephric short descending thin limb
metanephric loop of Henle
metanephric descending thin limb
metanephric prebend segment
descending thin limb
metanephric collecting duct
mesonephric connecting tubule
mesonephric nephron tubule
proximal convoluted tubule
proximal convoluted tubule segment 2
proximal convoluted tubule segment 1
proximal convoluted tubule brush border
mesonephric nephron epithelium
mesonephric nephron
mesonephric collecting duct
mesonephric distal tubule
proximal tubule
proximal tubular epithelium
pronephric proximal convoluted tubule
proximal straight tubule
renal cortex tubule
distal straight tubule macula densa
mesonephric proximal tubule
short nephron
inner renal medulla interstitium
inner renal medulla loop of Henle
loop of Henle ascending limb thin segment
inner medulla of kidney
inner medullary collecting duct
inner renal medulla vasa recta
inner medulla vasa recta ascending limb
inner medulla vasa recta descending limb
kidney outer medulla inner stripe
outer medulla inner stripe loop of Henle
premacula segment of distal straight tubule
outer renal medulla vasa recta
outer medulla vasa recta ascending limb
outer medulla vasa recta descending limb
renal medulla
outer medulla of kidney
outer renal medulla interstitium
outer renal medulla loop of Henle
kidney outer medulla outer stripe
outer medulla outer stripe loop of Henle
renal medulla loop of Henle
outer renal medulla peritubular capillary
outer medullary collecting duct
kidney pyramid
metanephric pyramid
renal medulla collecting duct
renal medulla interstitium
left renal medulla interstitium
right renal medulla interstitium
renal sinus
pronephric sinus
renal sinus of left kidney
renal sinus of right kidney
mesonephric sinus


anatomical system
hilum of kidney
metanephric glomerulus
metanephric glomerulus vasculature
metanephric glomerular epithelium
metanephric glomerular mesangium
metanephric glomerular endothelium
renal glomerulus vasculature
glomerular capillary
glomerular capillary endothelium
glomerular endothelium
kidney epithelium
uriniferous tubule
nephron
collecting duct of renal tubule
nephron tubule epithelium
nephron tubule basement m
cortical collecting duct
papillary duct
kidney collecting duct epithelium
distal tu
distal early tubule
macula
metanephric distal tubule
distal straight tubule
distal tubule
long nephron
nephron tubule
nephron epithelium
long descending thin limb
short descending thin limb
renal convoluted tubule
distal convoluted tubule
renal connecting tubule
metane
long descending thin li
descending limb of loop of Henle
ment
region of nephron tubule
metanephric long nephron
intermediate tubule
metanephric S-shaped body
loop of Henle
ascending limb of loop of Henle
thick ascending limb of loop
metanephric short nephron
metanephric long descending thin limb
metanephric nephron
metanephric nephron epithelium
metanephric nephron tubule
metanephric glomerular capillary
metanephric short descending thin limb
intermediate tubule
metanephric collecting duct
mesonephric connecting tubule
mesonephric nephron epithelium
mesonephric nephron tubule
proximal convolute
mesonephric nephron
mesonephric collecting duct
mesonephric distal tubule
proximal tubule
renal cortex tubule
short nephron
distal st
inner renal medulla interstitium
inner renal medulla loop of Henle
loop of Henle ascending limb thin segment
inner medulla of kidney
renal medulla interstitium
right renal medulla interstitium
pronephric sinus
renal sinus
renal sinus of left kidney
renal sinus of right kidney
mesonephric sinus

Data Formats

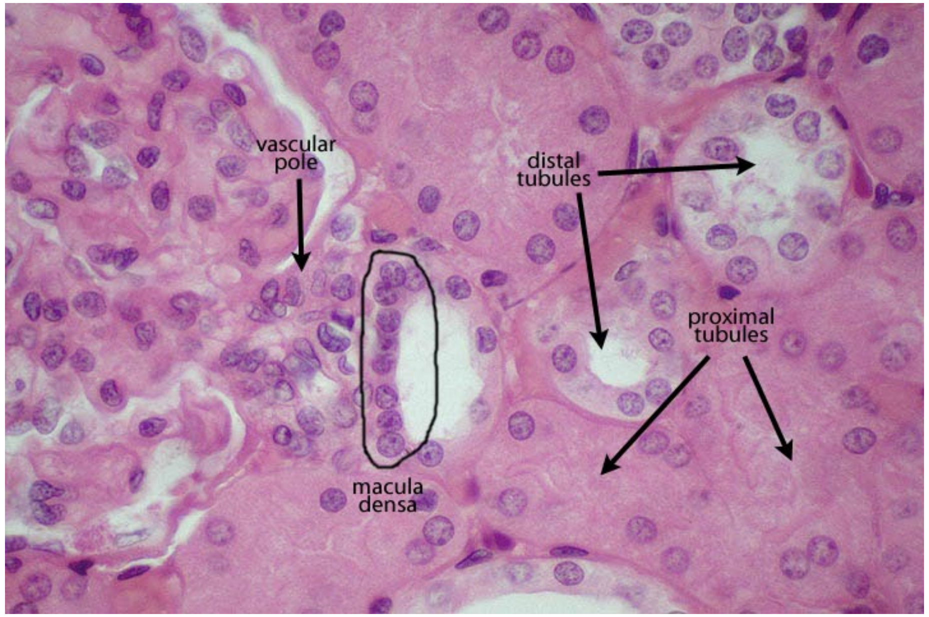
IU CCF Initial (v0.5.0) Image Formats

- **Basic image**: OME-Tiff
  - 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

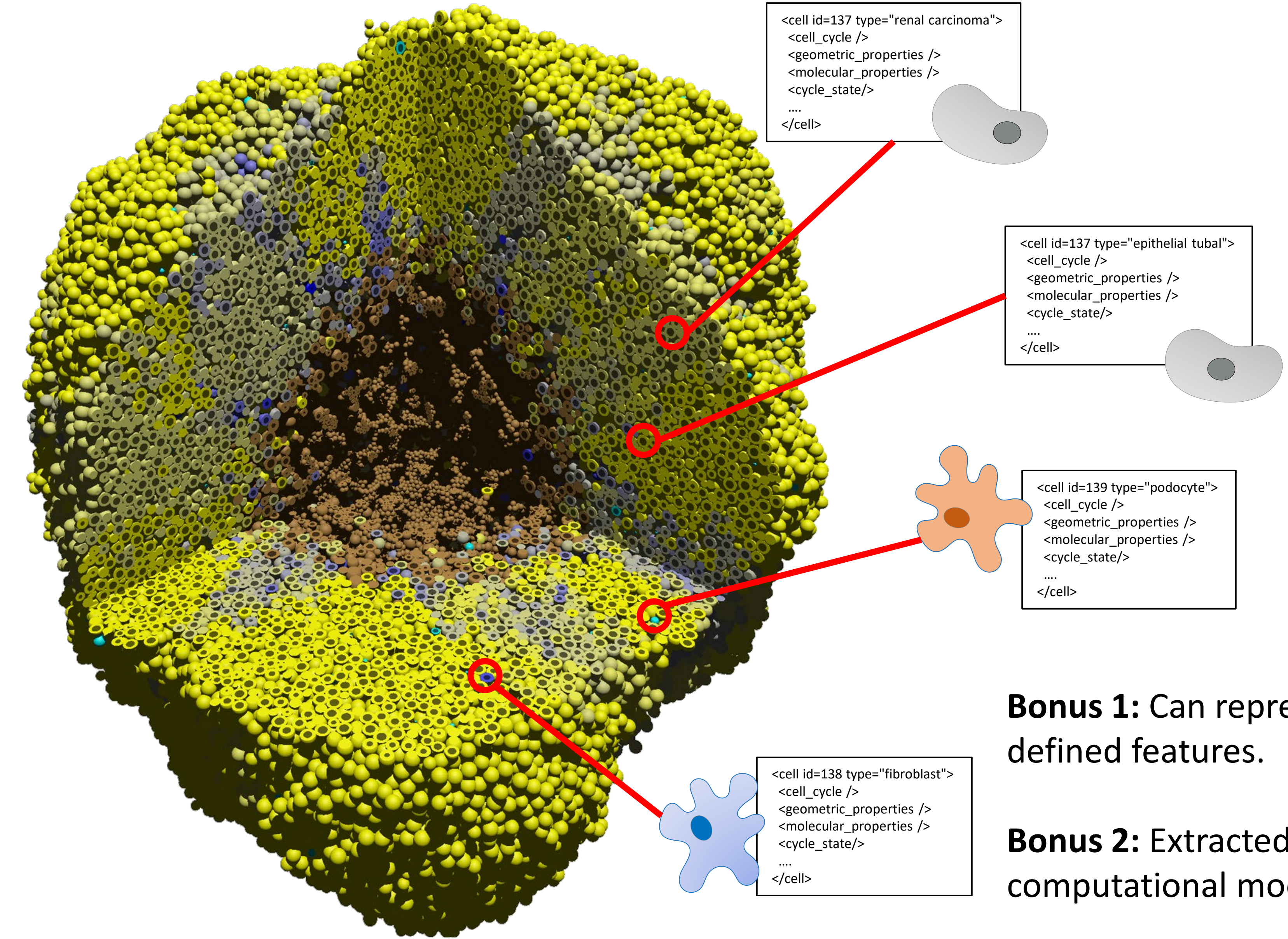
More data



- RNAseq and other OMICS data
  - Challenges in:
    - Data formatting and visual representation of spatial data
    - Extracting *knowledge* from complex, often noisy data
    - Harmonizing data sets from different platforms
    - Validating and benchmarking
- Extracted information
  - Segmented cells and structures
    - SVG overlays
      - e.g., cell apical surfaces, glomerulus podocyte, renal corpuscle, Bowman's space, …
    - Vectorized annotations (e.g., as in MultiCellDS – see below)
  - Multiplexed and massively multichannel imaging (e.g., MALDI)
    - Associate vectors of measurements with segmented structures
    - Additional ontology-driven annotations for the structures
      - e.g., cell type and state by Cell Ontology …
      - Cell morphometric annotations …

Vectorized annotations of extracted cell features:

a step from data towards knowledge

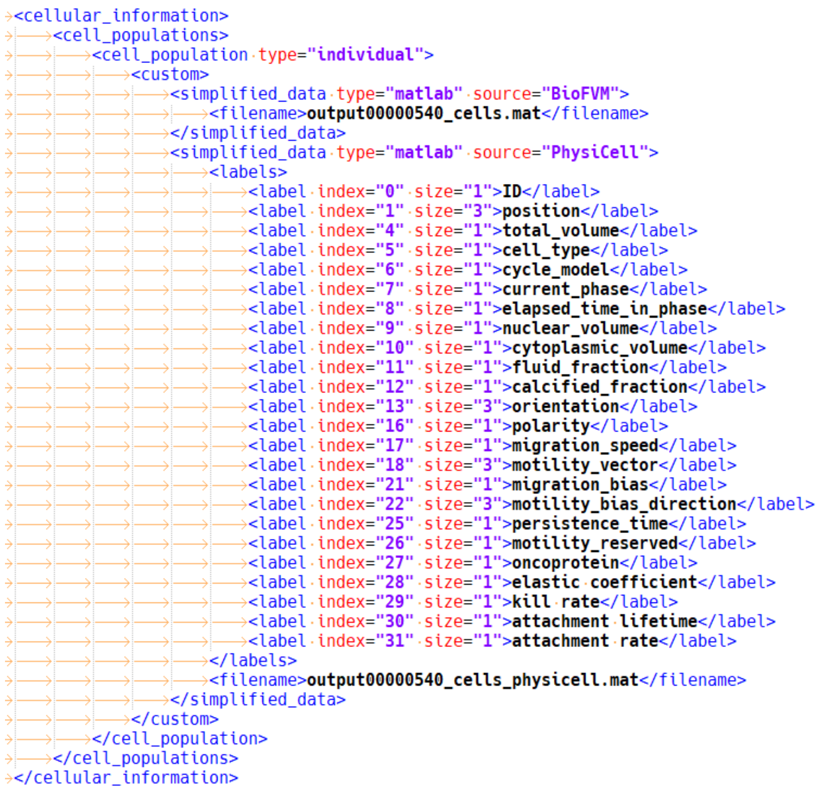


```
<cell id=137 type="renal carcinoma">
  <cell_cycle />
  <geometric_properties />
  <molecular_properties />
  <cycle_state/>
  ....
</cell>
```

```
<cell id=137 type="epithelial tubal">
  <cell_cycle />
  <geometric_properties />
  <molecular_properties />
  <cycle_state/>
  ....
</cell>
```

```
<cell id=139 type="podocyte">
  <cell_cycle />
  <geometric_properties />
  <molecular_properties />
  <cycle_state/>
  ....
</cell>
```

```
<cell id=138 type="fibroblast">
  <cell_cycle />
  <geometric_properties />
  <molecular_properties />
  <cycle_state/>
  ....
</cell>
```



```
<cellular_information>
    <cell_populations>
        <cell_population type="individual">
            <custom>
                <simplified_data type="matlab" source="BioFVM">
                    <filename>output00000540_cells.mat</filename>
                </simplified_data>
                <simplified_data type="matlab" source="PhysiCell">
                    <labels>
                        <label index="0" size="1">ID</label>
                        <label index="1" size="3">position</label>
                        <label index="4" size="1">total_volume</label>
                        <label index="5" size="1">cell_type</label>
                        <label index="6" size="1">cycle_model</label>
                        <label index="7" size="1">current_phase</label>
                        <label index="8" size="1">elapsed_time_in_phase</label>
                        <label index="9" size="1">nuclear_volume</label>
                        <label index="10" size="1">cytoplasmic_volume</label>
                        <label index="11" size="1">fluid_fraction</label>
                        <label index="12" size="1">calcified_fraction</label>
                        <label index="13" size="3">orientation</label>
                        <label index="16" size="1">polarity</label>
                        <label index="17" size="1">migration_speed</label>
                        <label index="18" size="3">motility_vector</label>
                        <label index="21" size="1">migration_bias</label>
                        <label index="22" size="3">motility_bias_direction</label>
                        <label index="25" size="1">persistence_time</label>
                        <label index="26" size="1">motility_reserved</label>
                        <label index="27" size="1">oncoprotein</label>
                        <label index="28" size="1">elastic coefficient</label>
                        <label index="29" size="1">kill rate</label>
                        <label index="30" size="1">attachment lifetime</label>
                        <label index="31" size="1">attachment rate</label>
                    </labels>
                    <filename>output00000540_cells_physicell.mat</filename>
                </simplified_data>
            </custom>
        </cell_population>
    </cell_populations>
</cellular_information>
```

Bonus 1: Can represent domain expert knowledge via expert-defined features.

Bonus 2: Extracted features could be ***directly imported*** into computational models.

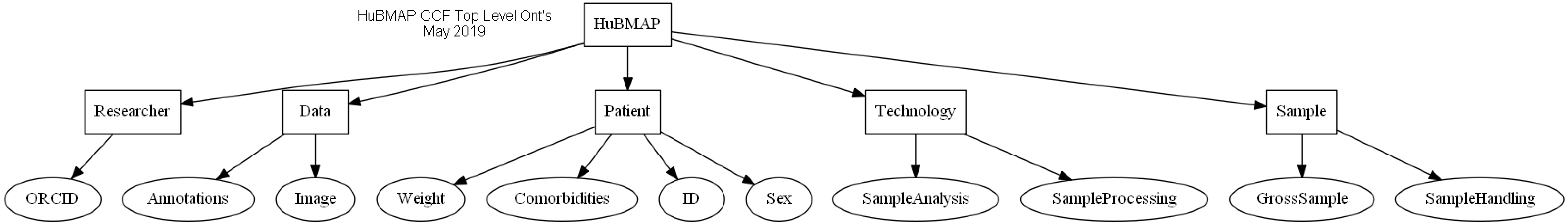
Metadata

Some Key Types of Metadata:

- Biological (patient specific data)
  - May include patient differentia such as disease state
- Technological (analysis tools; MS, immunohistochemical, RNAseq, …)
  - Where possible, respect and reuse any technology domain-specific standards, ontologies, etc.
  - Will often include the biological results (e.g., gene expression levels)
- Interpretive
  - Summary of findings, e.g., conversion of gene expression lists into list of highly impacted pathways
  - Algorithms and software used in interpretation and analysis
- Provenance
  - Who processed and analyzed samples
  - Points of contact (to answer questions on the samples and analyses)
  - Can be important for QC and account
- Use metadata
  - Licenses, citation information, ….

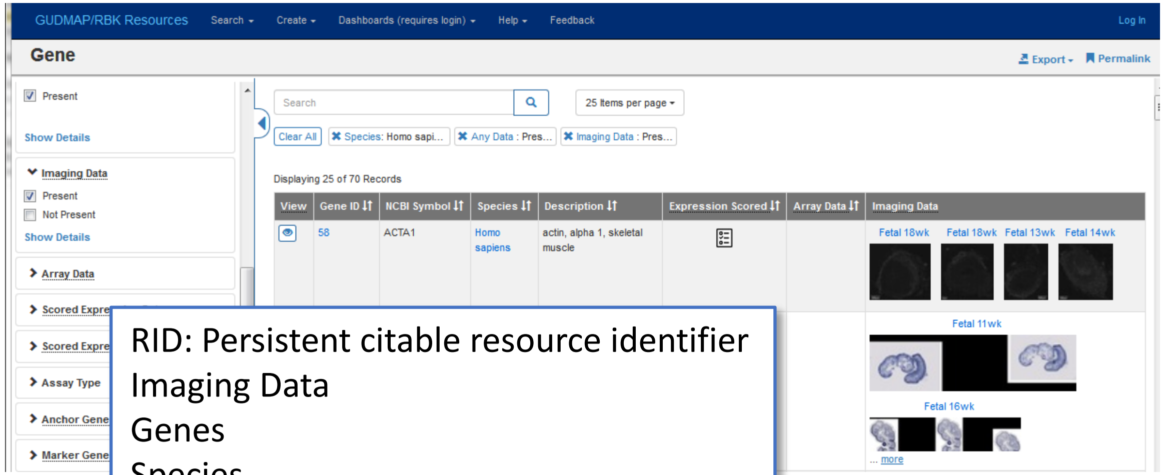
Many data and metadata types and files

- Patient
- Sample
- Analysis Technology
- Results
- Researcher
- . . .



Example: Metadata for Similar Large-Scale Data & Imaging Projects

GUDMAP:



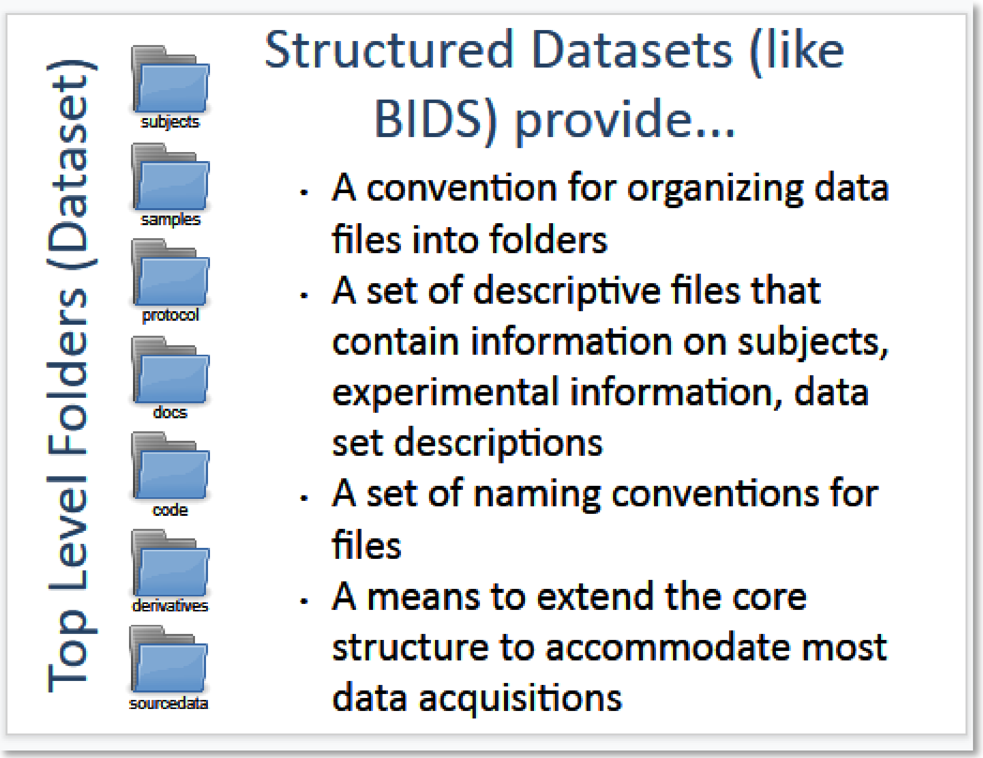
Displaying 25 of 70 Records

View	Gene ID	NCBI Symbol	Species	Description	Expression Scored	Array Data	Imaging Data
	58	ACTA1	Homo sapiens	actin, alpha 1, skeletal muscle			Fetal 18wk Fetal 18wk Fetal 13wk Fetal 14wk
							Fetal 11wk Fetal 16wk ... more

RID: Persistent citable resource identifier
Imaging Data
Genes
Species
Stage
Anatomical Sources
Assay Type
Preparation
Principle Investigator . . .

Kidney Tissue Atlas Ontology from KPMP

Human cell Atlas, SPARC (informed by BIDS)



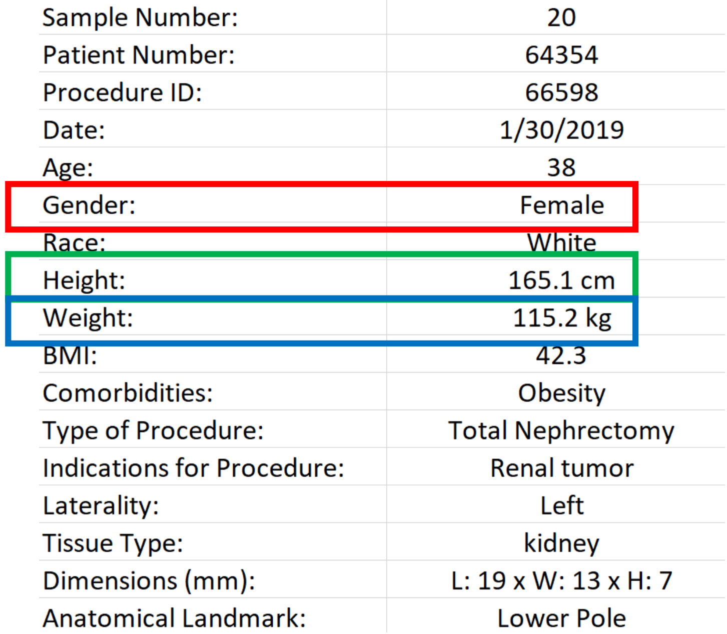
Structured Datasets (like BIDS) provide...

- A convention for organizing data files into folders
- A set of descriptive files that contain information on subjects, experimental information, data set descriptions
- A set of naming conventions for files
- A means to extend the core structure to accommodate most data acquisitions

Challenge: Inconsistent metadata specifications

Kidney: Jeff Spraggins et al., VU

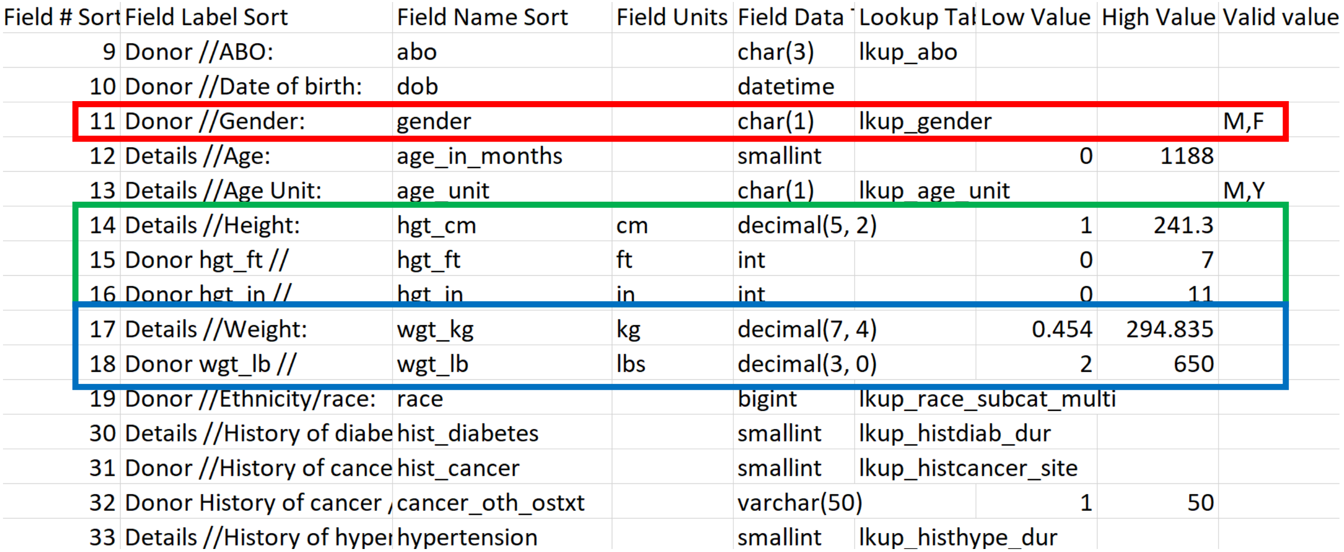
Clinical and Spatial Metadata (21 rows)



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

Heart: Shin Lin, UW

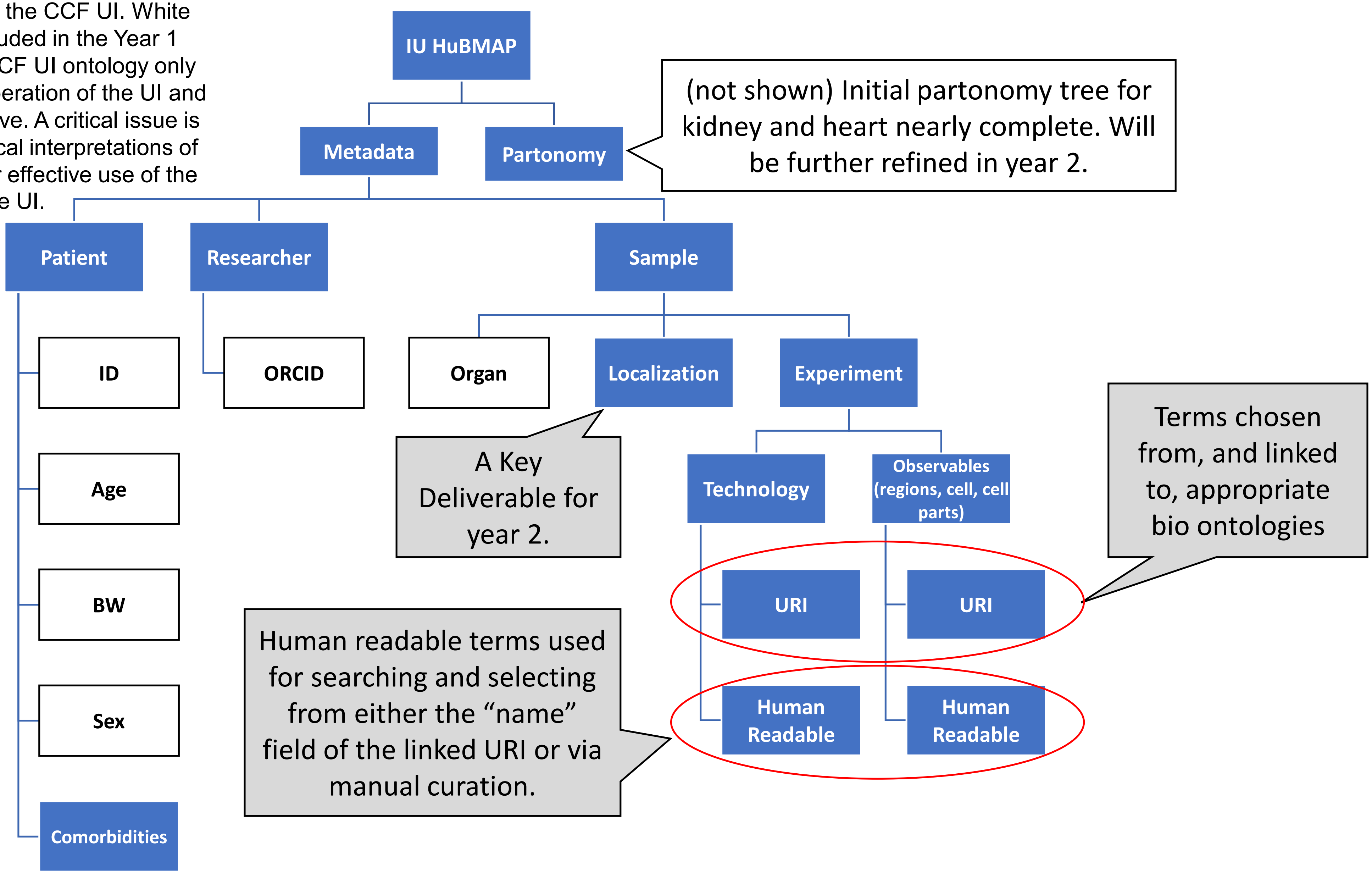
Data Dictionary (115 rows)



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 hyper	hypertension		smallint	lkup_histhype_dur			

AMIS (absolute minimal information solution)

Figure 2: Partial AMIS ontology for the CCF UI. White boxes indicate metadata terms included in the Year 1 metadata ontology. Note that the CCF UI ontology only covers the CCF data needed for operation of the UI and that is reflected in the ontology above. A critical issue is that human understandable biological interpretations of the various data sets is required for effective use of the image datasets by end users via the UI.



(not shown) Initial partonomy tree for kidney and heart nearly complete. Will be further refined in year 2.

A Key Deliverable for year 2.

Terms chosen from, and linked to, appropriate bio ontologies

Human readable terms used for searching and selecting from either the “name” field of the linked URI or via manual curation.

IU CCF Initial (v0.5.0) Patient Metadata

Column Header	Data Type	*Comments*
HuBMAP Sample ID	string	*assigned by PSE/IEC*
HuBMAP Patient Number	string	
Source Sample ID	string	*assigned by clinical unit, deidentified.*
Source Patient Number	string	
Procedure ID	string	
~~Procedure Date~~	~~formatted date~~	
Species	Human (STY:T016)	*required, though always human*
Age	decimal years	
Sex	M/F/u (SNOMED extended with "unkown")	
Race		*Race, ethnicity, strain*
Height	meter	
Weight	killogram	
BMI	float (calculated localy from height and weight)	
Comorbidities and other clinical classifications	MEDRA, SNOMED CT or MeSH terms	
Type of Procedure	MEDRA, SNOMED CT or MeSH terms	
Indications for Procedure	MEDRA, SNOMED CT or MeSH terms	
Laterality	MEDRA, SNOMED CT or MeSH terms	
Tissue Type	MEDRA, SNOMED CT or MeSH terms	
Anatomical Landmark	MEDRA, SNOMED CT or MeSH terms	
Displacement from Landmark	Affine transformation matrix	

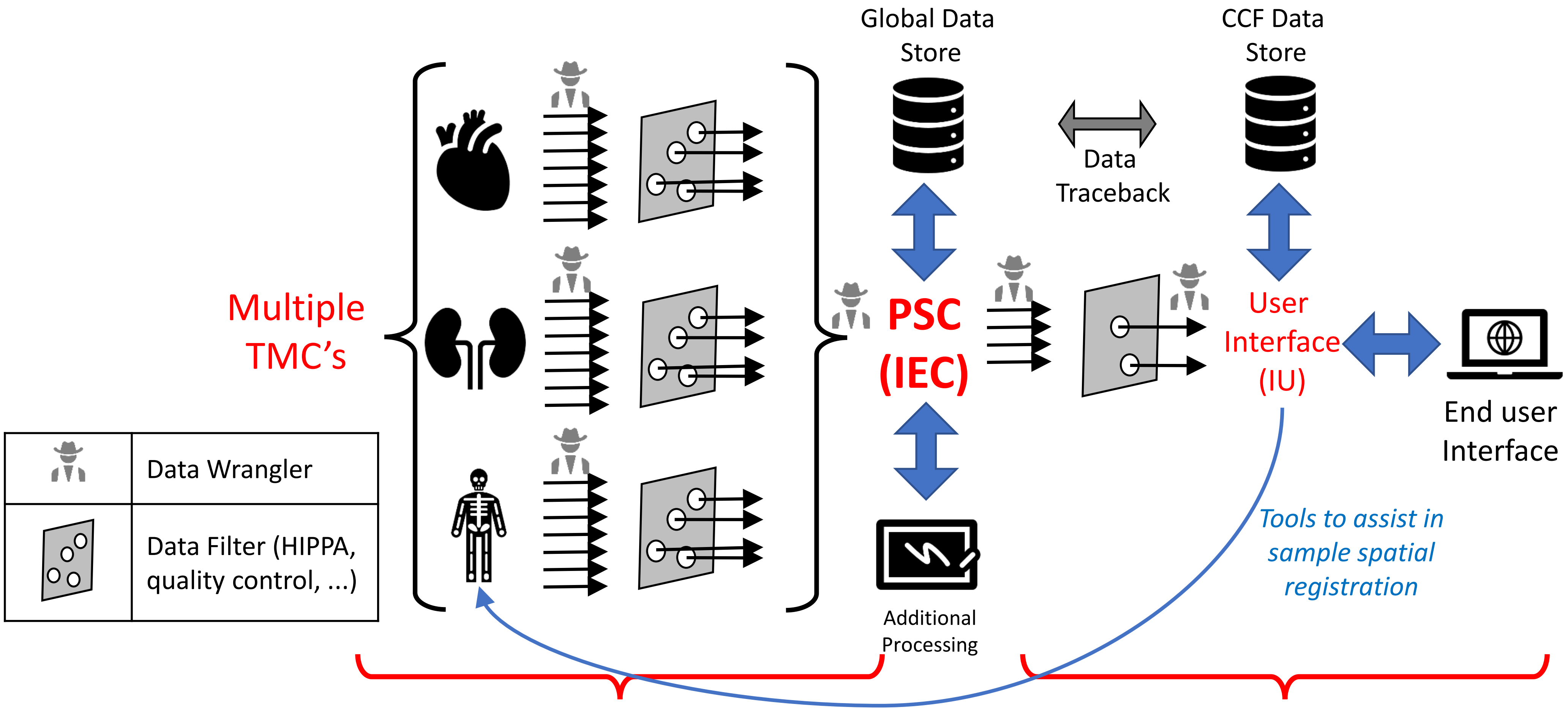
Next?

Discussion Points

- **Year 2 Plans**
  - Finalize metadata formats with input from the TMCs
  - Unify data formats across TMCs’ technologies
  - Work closely with the CCF IU team to ensure data content and formatting is compatible with the needs and goals of the user interface.
  - Work closely with TMCs to insure all anatomical scales are represented with ontology terms

- **Other TMC needs?**
  - Are there unmet needs?
  - Any "must have" features or terms to describe your data?

Extra reference materials



Multiple TMC's

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

Global Data Store

CCF Data Store

Data Traceback

PSC (IEC)

User Interface (IU)

End user Interface

Additional Processing

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

What is an ontology?

An ontology is a particular view of reality that encompasses a defined set of objects, processes and relationships within that reality.

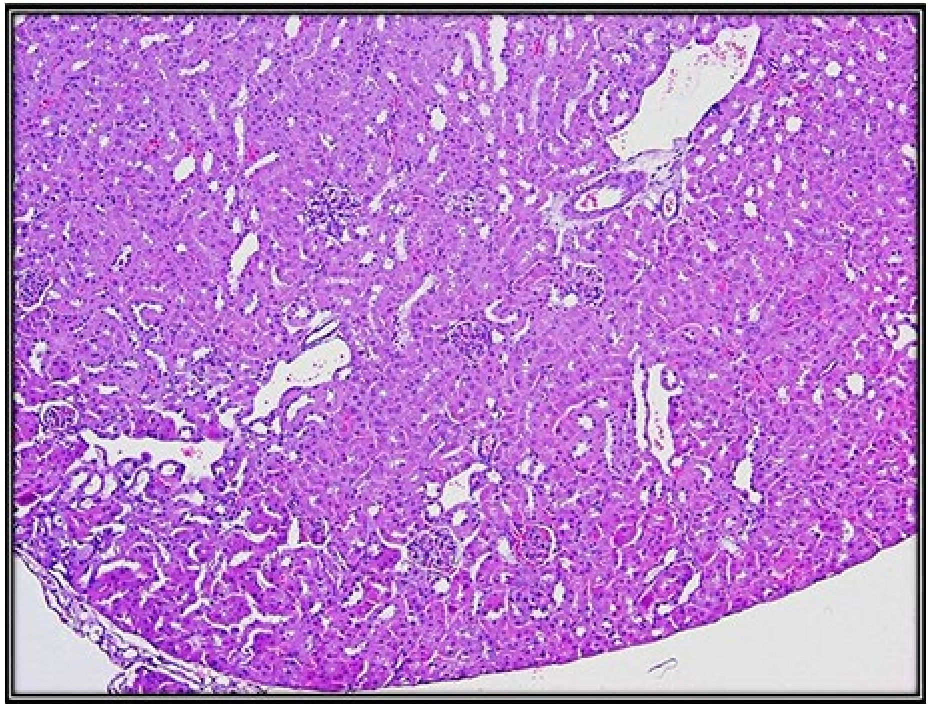
"Ontological Commitment"

Controlled Vocabulary	Hierarchy of Terms (isA)	Full Ontology
Cell Hepatocyte Leukocyte Organ Heart Liver	1. Cell a. Hepatocyte b. Leukocyte 2. Organ a. Heart b. Liver	1. Cell a. Hepatocyte b. Leukocyte 2. Organ a. Heart b. Liver *partOf*

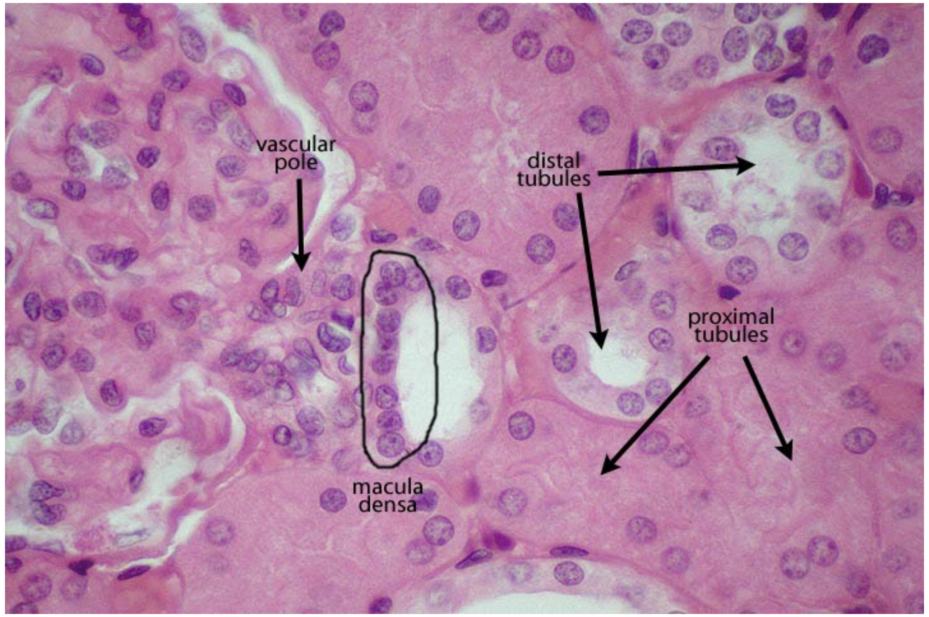
adjacentTo, containedIn, derivesFrom, definedBy, participatesIn, contributesTo, downStreamOf, …

Annotating Images and Images containing identified (not just identifiable) things

In the H&E stained kidney image below what regions (e.g., nephron, tubule, …), cell types, cell states etc. are present? Regions must be annotated, presumably by the TMCs. It is not enough to simply say this image contains cell types X, Y and Z in cell states 1, 2, and 3:



This image has been annoted (though the regions aren't very clear):

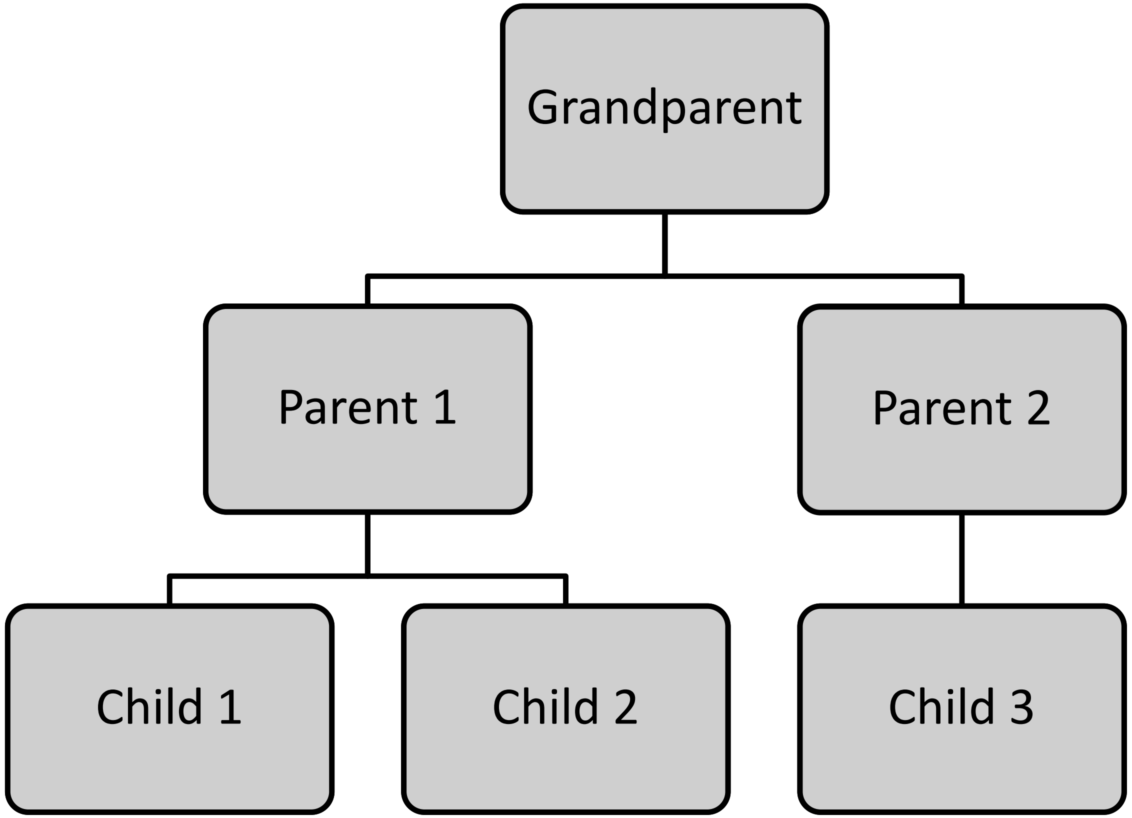


Solution: Multiple Slim Ontologies

- Gene Ontology (GO) is a massive ontology. Researchers didn’t need the entire ontology, but only a subset of it.
- GO users create “slim” versions of GO:
  - Subsets that contained the terms needed along with the necessary ancestor and children terms
  - Can better illuminate the science of what is going on rather than being overwhelmed by too much information
- Terms in the main ontology can receive “slim” annotations, indicating their use in the corresponding “slim” ontology

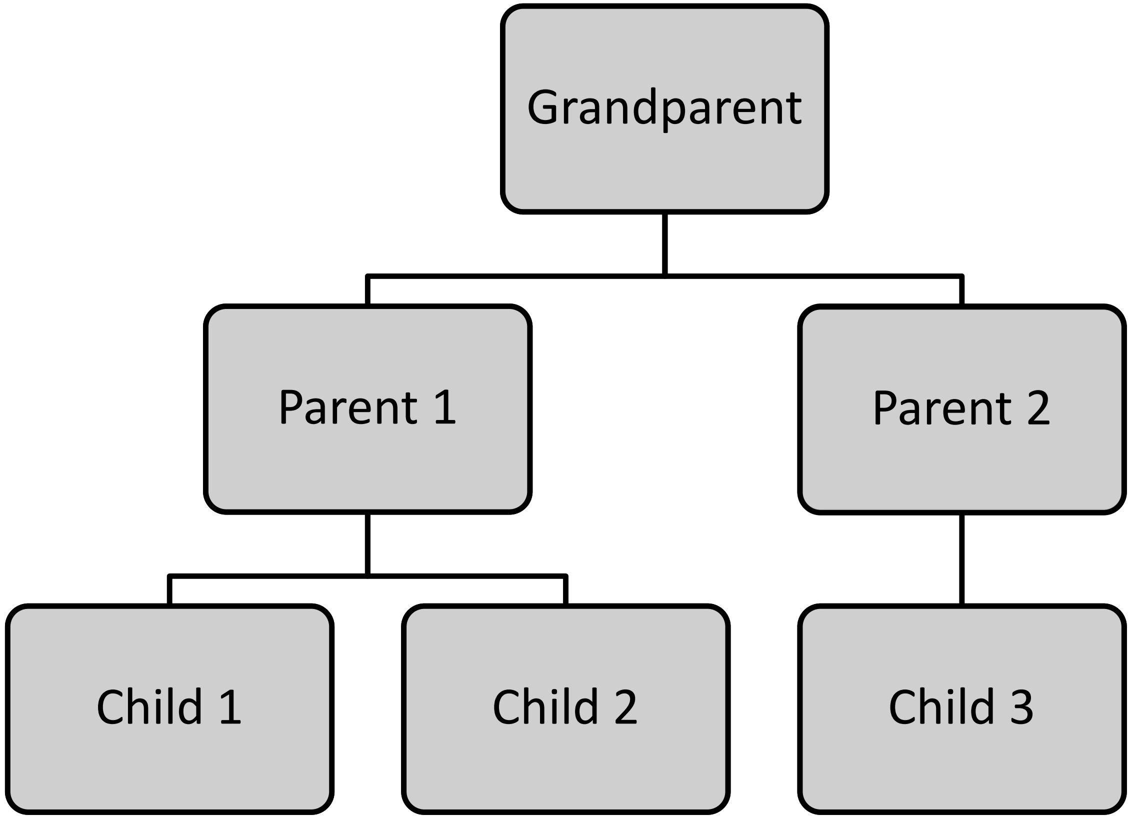
Multiple Slim Ontologies to Main HuBMAP/CCF Ontology

- Instead of creating a single ontology file, we have software create the ontology based on the needed terms.
- We can ensure that we obtain all appropriate parent/child nodes are included



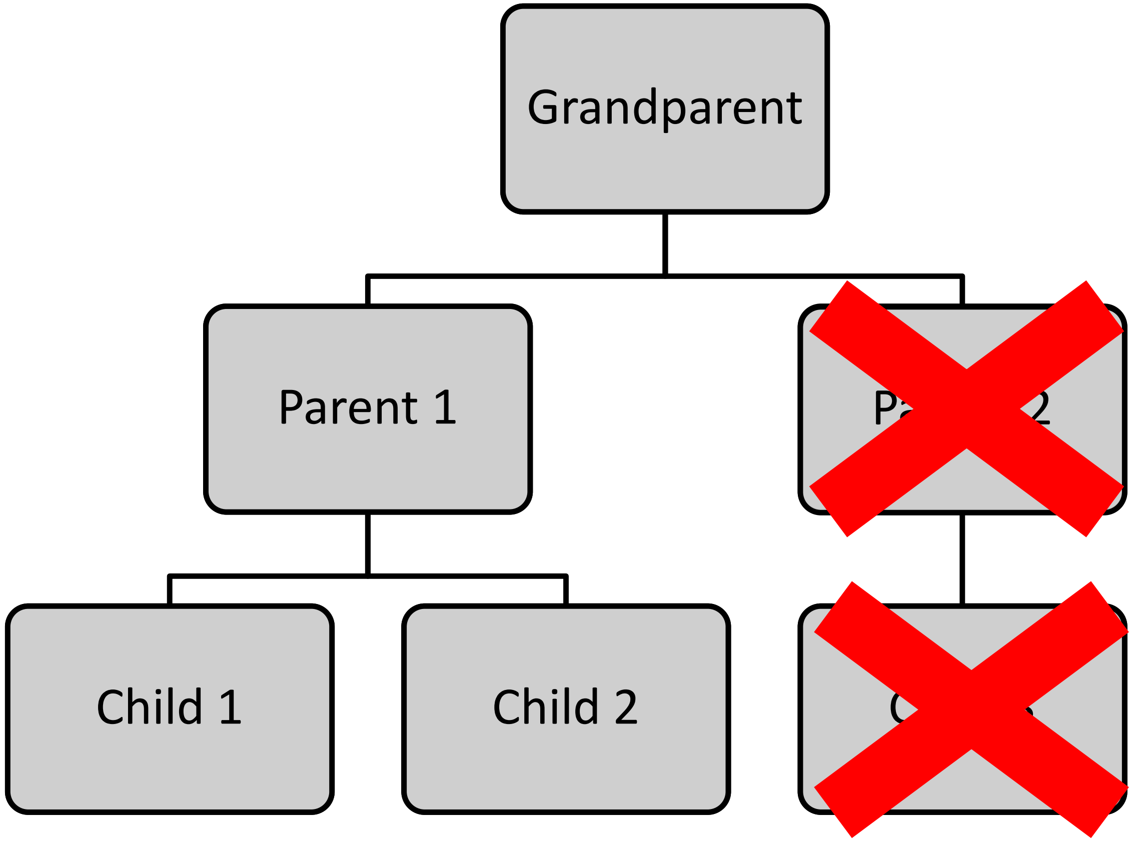
Multiple Slim Ontologies to Main HuBMAP/CCF Ontology

- We could specify for the main file we want nodes Parent 1 (from Ontology A) and Parent 2 (from Ontology B) and all their descendent nodes.
- We could specify we want nodes Children 1, 2, and 3 and all of their ancestor nodes.
- We could do a mix and still get the ontology.



Multiple Slim Ontologies to Main HuBMAP/CCF Ontology

- When viewing the ontology, we can easily eliminate extraneous terms for UI/navigation (e.g. eliminate Parent 2)
- When in a specific part of the body, certain terms could just disappear (e.g. eliminate Parent 2 and descendants)
- With a graph library, very easy to remove nodes or branches.



Pros/Cons of Multiple Slim Ontologies

Pros

- Ensures that we have the correct biology in the base ontology.
  - Lots of work already done by using pre-existing ontologies
- Can have slim ontologies for each zoom level, organ, or system
- As the base ontologies continue to update, new information propagates in
- Easy to add additional ontologies through crosswalks
- Elimination of some of the hand editing of ontologies

Cons

- Need to specify which versions of the input ontologies we are using
- Tracking of slim ontologies could become burdensome
- Lack of hand editing could make slim ontologies harder to use
- Hand editing could be replaced with code, but then additional coding effort is necessary

- Most of these cons are easily surmountable by using GitHub and using a triplestore repository like LungMAP did for its ontology

Dimensions of CCF Ontology

- Overall anatomical location
  - Heart aorta
- Cell line/type location
  - Endothelial cells
- Chemical location
  - Areas for positive staining for H&E

- Real data will have to be mapped into/annotated with the ontologies
- Some of this data will need to come from the experimental protocols (possible semantic integration with protocols.io ?) and some will likely come from Machine Learning