

# The GTEx-HuBMAP tissue mapping Collaboration (for the CFDE)

Kristin Ardlie & Katy Börner

# The Challenge

Disparate Common Fund projects such as GTEx and HuBMAP (and others: MoTrPAC, SPARC, Kids First...etc.) are generating large molecular and spatial datasets on human tissues (contributing to even larger, similar efforts, e.g HCA, HPA).

For these data to be maximally “*Interoperable*”, **we need to accurately map and compare** data from the same organs, tissues, and cell types across these projects.

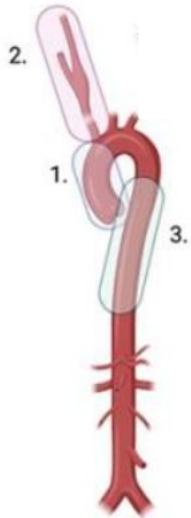
Our goal was to map & connect:

- **GTEx tissue samples** - collected across MANY individuals at the same tissue site per individual (using SOPs) – population level data
- **HuBMAP tissue samples** - collected for few individuals at MANY specific sites within a tissue (mapped to tissue locations using CCF, & ASCT+B details) – individual level data

# Goals of our CFDE-funded project

- Spatially map samples collected under GTEx biospecimen SOPs to the CCF to which HuBMAP samples are mapped
- Generate and make available spatially, semantically, and ontologically explicit FAIR metadata for GTEx and HuBMAP data housed on several cloud environment
- Extend and serve as web components data visualization and exploration user interfaces that were initially developed for HuBMAP but are valuable for other consortia
- Make a combination of anatomically mapped and ontologically linked human tissue data available across both HuBMAP and GTEx portals

# Location, Location, Location matters

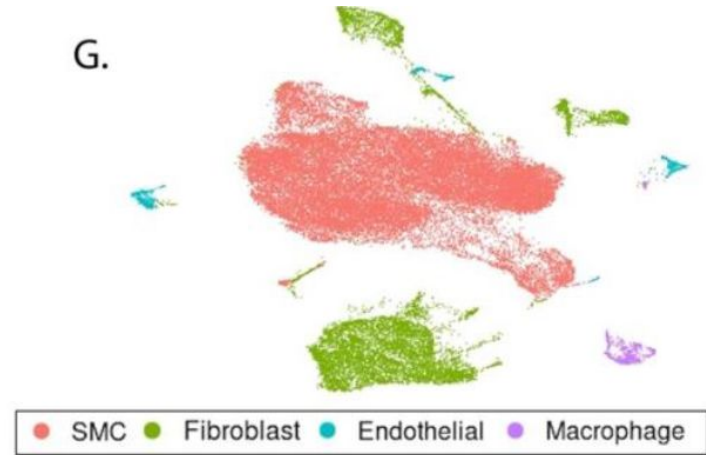


1. *Ascending aorta and aortic root*  
*Secondary heart field and neural crest*

2. *Brachiocephalic and carotid artery*  
*Neural crest*

3. *Descending thoracic aorta*  
*Somitic mesoderm*

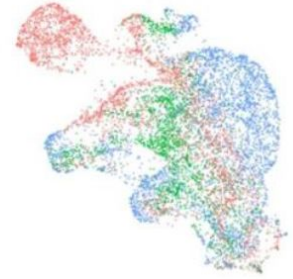
G.



I. *SMC subset*



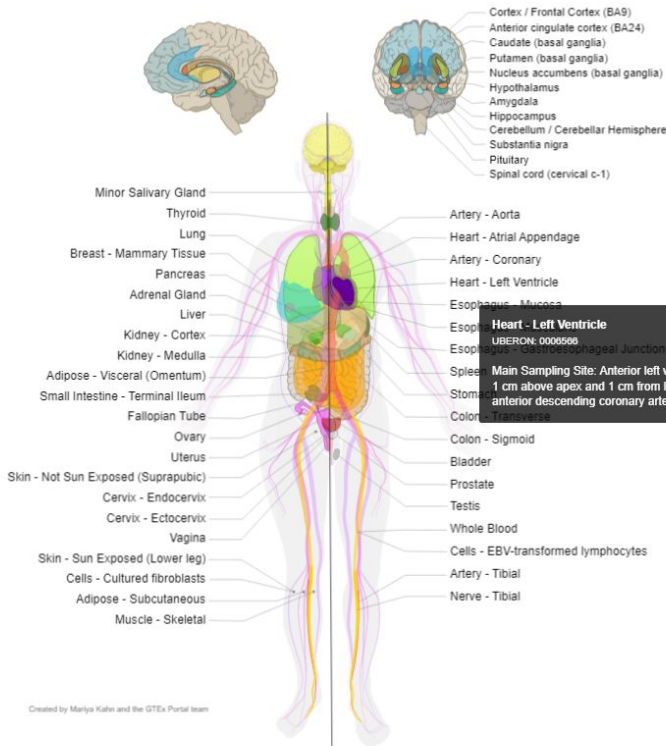
J. *Fibroblast subset*



# GTEX Tissue sites

## Tissue Sampling Sites

This page provides a visual representation of the biospecimen source sites (BSSs) for the collection of tissue from postmortem/organ procurement cases for the Genotype-Tissue Expression (GTEx) project. The full documentation on tissue collection procedures can be found on the [GTEx Tissue Harvesting Work Instruction](#).



Created by Marilya Kahn and the GTEx Portal team

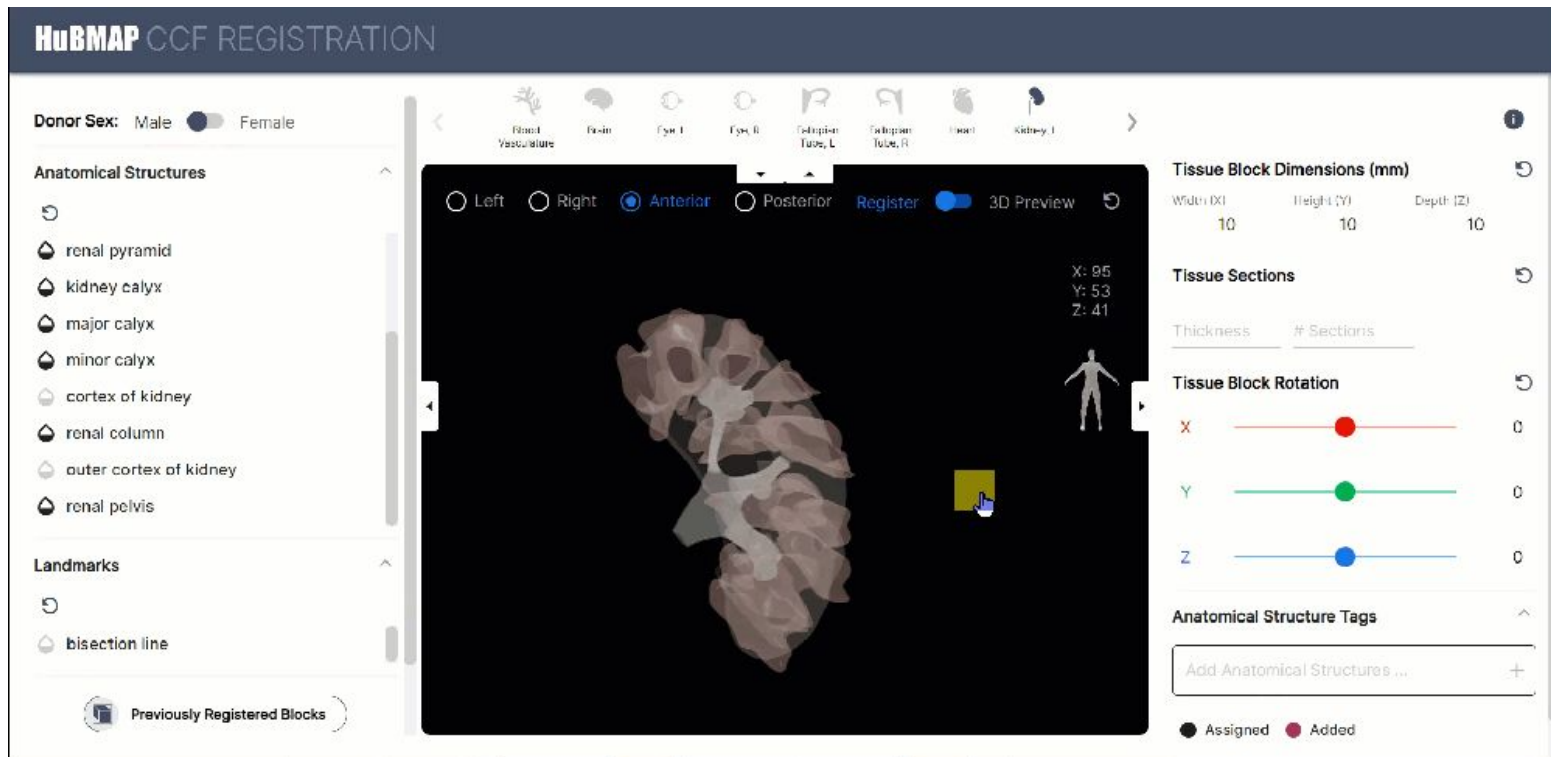
Tissue sampling site pre-defined “SOP”

Goal was to collect same “normal/healthy” site across ~1000 individuals

Tissue	UBERON	Main Sampling Site
Adipose - Subcutaneous	0002190	Subcutaneous tissue beneath the leg's skin sample.
Adipose - Visceral (Omentum)	0010414	Adipose tissue on the large fold of peritoneum that hangs down from the greater curvature of the stomach, passing in front of the small intestines.
Adrenal Gland	0002369	Left, followed by the right if necessary for sufficient aliquots.
Artery - Aorta	0001496	Ascending aorta or other thoracic regions (nonatherosclerotic).
Artery - Coronary	0001621	Left and right, noncalcific regions only.
Artery - Tibial	0007610	Left Tibial. Peripheral tibial artery from gastrocnemius region
Bladder	0001255	Central posterior urinary bladder, trimming from the outer wall if necessary.
Brain - Amygdala	0001876	Amygdala (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Anterior cingulate cortex (BA24)	0009835	Anterior cingulate cortex (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Caudate (basal ganglia)	0001873	Caudate (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Cerebellar Hemisphere	0002037	Cerebellar hemisphere (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Cerebellum	0002037	Right cerebellum (sampled at donor collection site and preserved in PAXgene fixative).
Brain - Cortex	0001870	Right cerebral frontal pole cortex (sampled at donor collection site and preserved in PAXgene fixative).
Brain - Frontal Cortex (BA9)	0009834	Right cerebral frontal pole cortex (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Hippocampus	0001954	Hippocampus (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Hypothalamus	0001898	Hypothalamus (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Nucleus accumbens (basal ganglia)	0001882	Nucleus accumbens basal ganglia (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Putamen (basal ganglia)	0001874	Putamen basal ganglia (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Spinal cord (cervical c-1)	0006469	Spinal cord (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Brain - Substantia nigra	0002038	Substantia nigra (sampled at Miami Brain Bank and preserved as fresh frozen tissue).
Breast - Mammary Tissue	0008367	Central breast subareolar region of the right breast, 1-2 cm under the skin surface of nipple region.
Cells - Cultured fibroblasts	EFO_0002009	Cultured primary fibroblasts.
Cells - EBV-transformed lymphocytes	EFO_0000572	Transformed lymphocytes.
Cervix - Ectocervix	0012249	Ectocervix (squamous) from uterus. Bivalve uterus along the endocervical canal to open uterus and cervix from external os to fundus.
Cervix - Endocervix	0000458	Endocervix (glandular) from uterus. Bivalve uterus along the endocervical canal to open uterus and cervix from external os to fundus.
Colon - Sigmoid	0001159	Sigmoid colon, Obtain muscularis only; discard mucosa.
Colon - Transverse	0001157	Transverse, Full thickness: mucosa and muscularis

# HuBMAP CCF Registration User Interface (RUI)

Mapping efforts are developing a *common coordinate framework* (CCF) for the healthy human body, to support cataloging different types of individual cells, understanding the functions of and relationships between those cell types, and modeling their individual and collective function.



<https://hubmapconsortium.github.io/ccf-ui/rui/>

# Example: Mapping GTEx Colon to HuBMAP RUI

		<b>GTEx Tissue Harvesting Work Instruction</b>	
PR-0004-W1	VER. 03.05	Effective Date: mm/dd/yyyy	Page 13 of 21

4.3.6.20 Colon

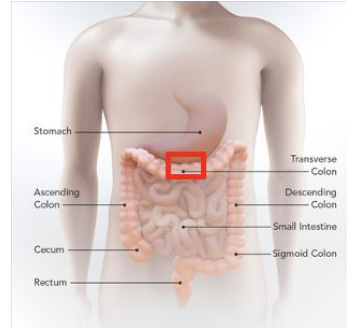
4.3.6.20.1 **Preferred Location:** Transverse colon. Gently rinse mucosa with normal saline before aliquot preparation. Aliquots should contain the full thickness of the colonic wall, i.e., **mucosa and muscularis propria**. **Trim adjacent adipose tissue.**

4.3.6.20.2 **Preferred Aliquot:** 20 mm x 10 mm x thickness ( $\leq 4$  mm), divided into two adjacent 10 mm x 10 mm x thickness aliquots. Each cassette should contain two 10 mm x 10 mm x thickness aliquots.

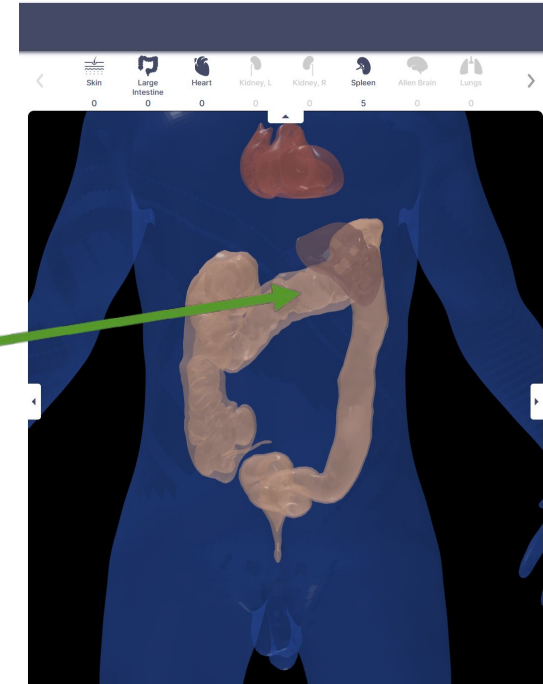
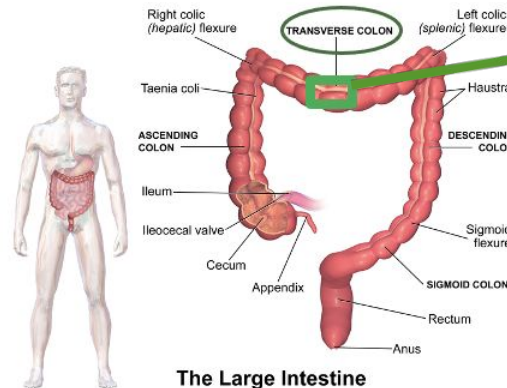
4.3.6.20.3 **Preferred Location:** Sigmoid colon. **Preferred Location:** Sigmoid colon. Gently rinse mucosa with normal saline before aliquot preparation. **Obtain only muscularis propria**; discard mucosa and any serosal adipose tissue.

4.3.6.20.4 **Preferred Aliquot:** 20 mm x 10 mm x thickness ( $\leq 4$  mm), divided into two adjacent 10 mm x 10 mm x thickness aliquots. Each cassette should contain two 10 mm x 10 mm x thickness aliquots.

Sigmoid Colon ('pelvic colon') Dissection Guide (Diagram 4)



Recover the transverse colon starting 10 cm back from the right colic (hepatic) flexure.



# CCF Registration User Interface (RUI) - GTEx samples

The RUI was used to capture 29 extraction sites. RUI usage is well-defined via an [SOP](#) and [video demo](#). These extraction sites can be associated with up to **5,171** tissue blocks (and GTEx population-level summary information, such as eQTLs).

**Heart Atrial Appendage** | 293 Male | 136 Female

**Heart Left Ventricle** | 294 Male | 138 Female

**Spleen** | 154 Male | 87 Female

**Kidney Cortex** | 66 Male | 19 Female

**Colon Sigmoid** | 240 Male | 133 Female

**Colon Transverse** | 259 Male | 147 Female

**Lung** | 395 Male | 183 Female

**Artery Coronary (Blood Vasculature)** | 146 Male | 94 Female

**Small Intestine Terminal Ileum** | 120 Male | 67 Female

**Pancreas** | 207 Male | 121 Female

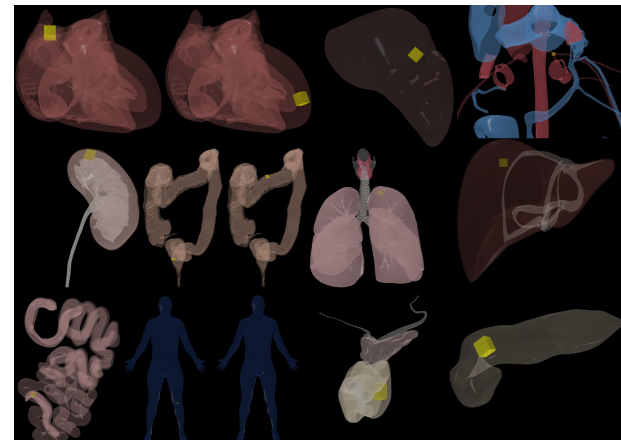
**Uterus** | 142 Female

**Ovary** | 180 Female

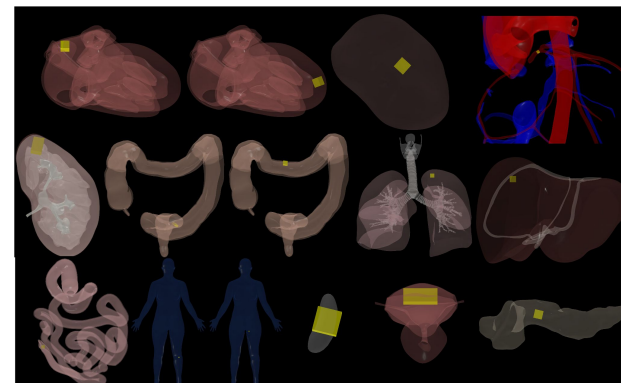
**Prostate** | 245 Male

**Skin Sun-Exposed Lower Leg** | 467 Male | 234 Female

**Skin Not Sun-Exposed Suprapubic** | 411 Male | 193 Female



Extraction sites (male)

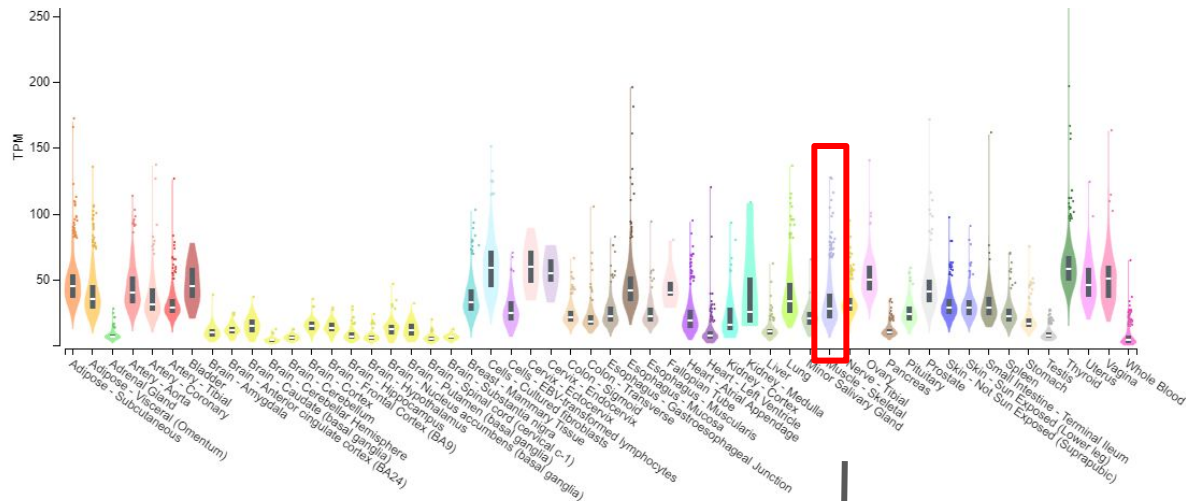


Extraction sites (female)



# Checking Accuracy

# Tissue level expression

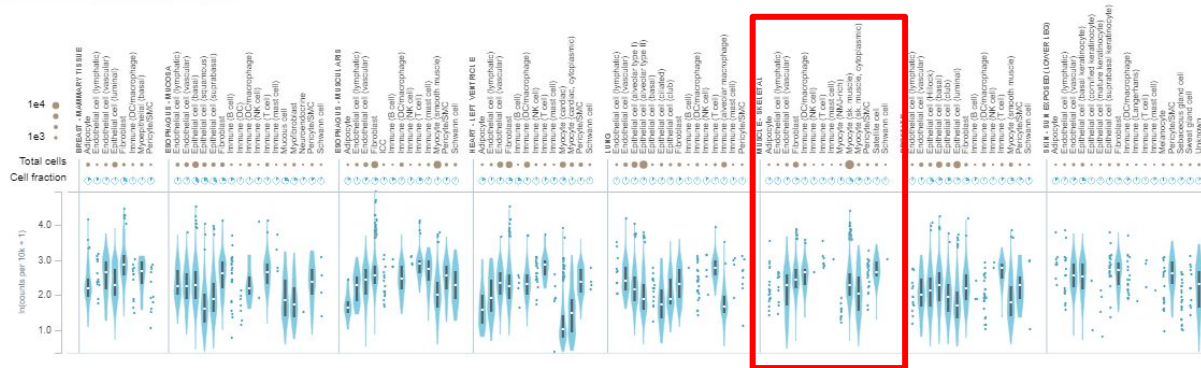


Single tissue expression for SMAD3 (ENSG00000166949.15)

Data Source: Single cell snRNA-seq pilot

All Nonzero Split C T

# cell level expression



# GTEX Cell Populations

## Tissue Blocks Comparison

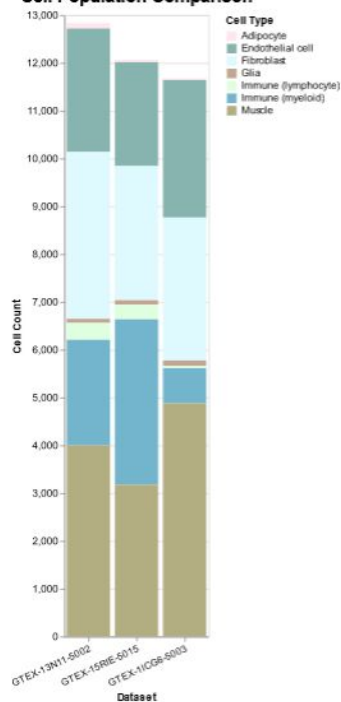
Source:

Sort By:

Group By:

Y-Axis:  Raw Count  Percentage

### Cell Population Comparison



GTEX provides high quality data for Human Reference Atlas (HRA) design.

Shown are cell populations (i.e., cell types and number per type) across tissue blocks that derive from the same spatial location within three human reference organs (heart, prostate, skeletal muscle).

Given sufficient cell population data for the same location and major demographic groups, it will become possible to predict cell populations for new tissue block data.

## Tissue Blocks Comparison

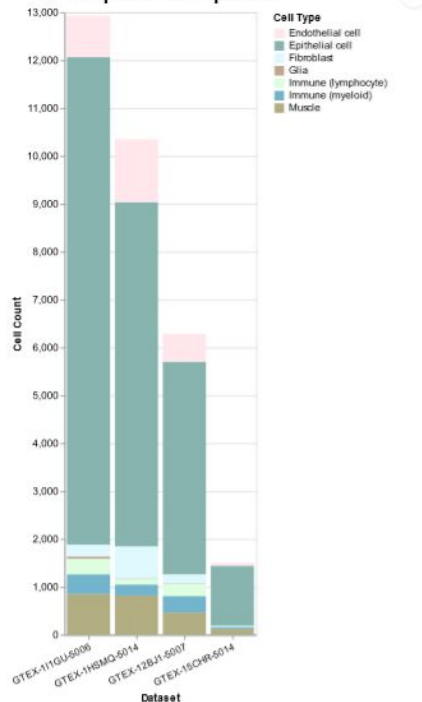
Source:

Sort By:

Group By:

Y-Axis:  Raw Count  Percentage

### Cell Population Comparison



## Tissue Blocks Comparison

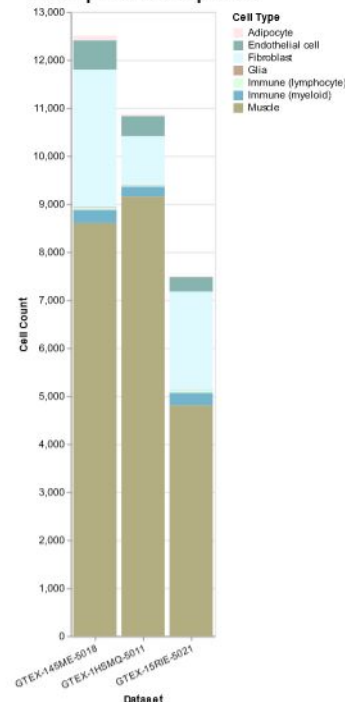
Source:

Sort By:

Group By:

Y-Axis:  Raw Count  Percentage

### Cell Population Comparison



# GTEX / HuBMAP Deliverables

GTEX-branded CCF EUI for GTEX Portal, including cross-consortia Cell Type search

The screenshot shows the GTEX Portal interface. At the top, there are navigation links for Home, Downloads, Expression, Single Cell, QTL, IGV Browser, Tissues & Histology, and Documentation. A search bar for 'Search Gene or SNP ID...' is visible. Below the navigation, there are filters for Sex (Both), Age (1-110), and BMI (13-83). The main area is divided into two search sections: 'Search anatomical structures...' and 'Search cell types...'. The anatomical structures section lists various organs and tissues with counts, such as 'body' (309), 'brain' (0), 'lymph node' (27), 'eye' (0), 'fallopian tube' (0), 'heart' (27), 'kidney' (58), 'knee' (0), 'liver' (11), 'lung' (4), 'ovary' (1), and 'pancreas' (4). The cell types section lists various cell types with counts, such as 'absorptive' (46), 'adipocyte' (27), 'adipocyte' (70), 'adipocyte' (16), 'adventitial stromal cell' (55), 'afferent arteriole endothelial cell' (58), and 'airway smooth muscle' (11). A central 3D anatomical model of a human torso is shown, with various organs highlighted in different colors. To the right of the model, there are filters for 'body | cell' and a list of tissue data providers, donors, blocks, sections, and datasets. Below the model, there are several filters for demographic and clinical data, such as 'Females (n=84) Mean Age 53.5 (range 21 - ...)', 'Males (n=148) Mean Age 54.3 (range 20 - 7...)', 'Females (n=133) Mean Age 51.7 (range 21 - ...)', 'Males (n=240) Mean Age 52.3 (range 20 - 7...)', 'Females (n=147) Mean Age 49.3 (range 21 - ...)', 'Males (n=259) Mean Age 50.3 (range 21 - 7...)', 'Females (n=138) Mean Age 55 (range 21 - 7...)', 'Males (n=293) Mean Age 56.4 (range 20 - ...)', 'Females (n=138) Mean Age 53.2 (range 21 - ...)', 'Males (n=294) Mean Age 54.9 (range 20 - ...)', and 'Females (n=19) Mean Age 56.4 (range 30 - ...)'.

More to explore here:

<https://hubmapconsortium.github.io/ccf-gtex-pilot/>

## CCF-API Client Libraries published

- CCF API
  - Production: <https://ccf-api.hubmapconsortium.org>
  - Staging: <https://ccf-api--staging.herokuapp.com>
  - OpenAPI Specification: [ccf-api-spec.yaml](https://github.com/ccf-api-spec)
- CCF API Client Libraries:
  - JavaScript: [npm v1.0.0 @ccf-openapi/js-client](https://www.npmjs.com/package/@ccf-openapi/js-client)
  - TypeScript: [npm v1.0.0 @ccf-openapi/ts-client](https://www.npmjs.com/package/@ccf-openapi/ts-client)
  - Angular 12+: [npm v1.0.0 @ccf-openapi/ng-client](https://www.npmjs.com/package/@ccf-openapi/ng-client)
  - Python 3.6+: [pypi v1.0.0 ccf-openapi](https://pypi.org/project/ccf-openapi/)

## CCF-API Implemented and deployed with Interactive Documentation Site

### Get aggregate results / statistics

[GET https://ccf-api.hubmapconsortium.org/v1/aggregate-results](https://ccf-api.hubmapconsortium.org/v1/aggregate-results)

### Request

#### Query Parameters

**age** Number Range

Filter result by age.

Style: Comma separated values

**min** number

**max** number

**age-range** string

Filter result by age range.

Example: "10-20"

Style: Comma separated values

**bnk** Number Range

Filter result by bnk.

Style: Comma separated values

**min** number

**max** number

**bnk-range** string

Filter result by bnk range.

Example: "0.45"

Style: Comma separated values

**cache** boolean

Hint to cache (or not) the response of this request.

Default: true

Style: Comma separated values

**ca11-tissue-forms** array[boolean]

The screenshot shows the interactive documentation site for the CCF-API. It displays the 'Parameters' section for the 'GET /v1/aggregate-results' endpoint. The parameters listed are 'age', 'age-range', 'bnk', and 'bnk-range', each with a 'Filter result by' description and a 'Style' of 'Comma separated values'. There are also 'min' and 'max' fields for 'age' and 'bnk'. A 'Send API Request' button is visible, along with a 'Response' section showing a 'Request Sample: Shell / cURL' and a 'Response Example' in JSON format.

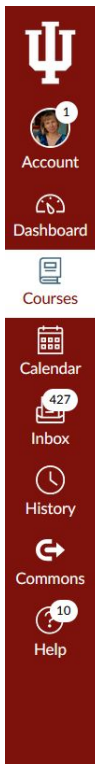
```
Request Sample: Shell / cURL
curl --request GET \
  --url https://ccf-api.hubmapconsortium.org/v1/aggregate-results \
  --header 'Content-Type: application/json'
```

```
Response Example
{
  1: [
  2: {
  3:   "label": "string",
  4:   "count": 0
  5: }
  6: ]
}
```

# GTEx / HuBMAP Deliverables

GTEx Training as part of the HuBMAP Visible Human MOOC (VHMOOC)

<https://expand.iu.edu/browse/sice/cns/courses/hubmap-visible-human-mooc>



- Home
- Announcements
- Syllabus
- Modules
- Discussions
- People
- Pages
- Files
- Grades
- Rubrics
- Outcomes
- Assignments
- Quizzes
- Collaborations
- Settings

No Expiration

## 3rd HuBMAP Portal Release (June 2022)



### HuBMAP Halfway Point

- HuBMAP consortia members reflect on the past four years and discuss their plans and hopes for the future.



### Introduction to the HRA-CCF

- An introduction to the three ontologies at work in the Human Reference Atlas's Common Coordinate Framework: the specimen, biological structure, and spatial ontologies.



### Using the EUI on the GTEx Portal

- How to cross-compare data from GTEx and HuBMAP by using the Exploration User Interface embedded in the GTEx Portal.

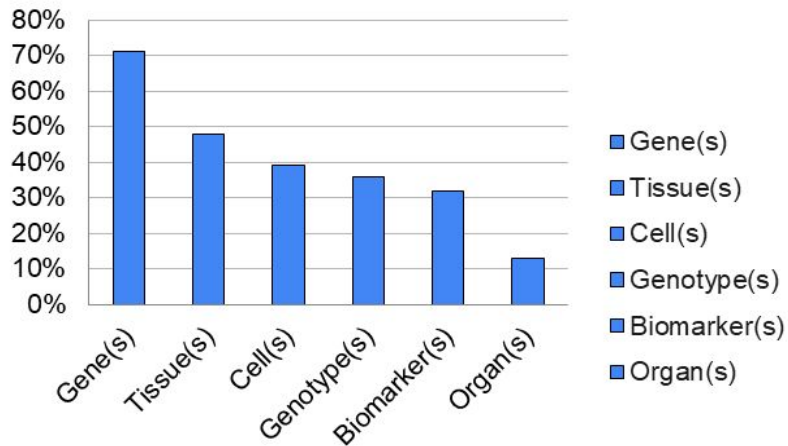
<https://gtexportal.org/home/eui>

# GTEEx / HuBMAP Survey

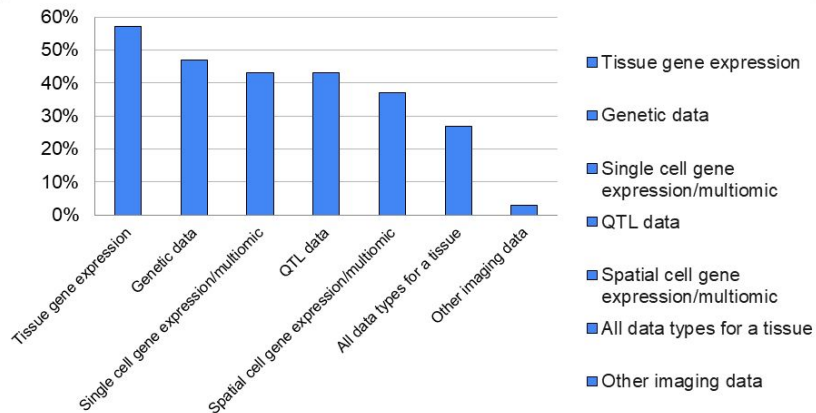
User Survey results (GTEEx Portal) (n=32):

- 100% state that report tutorial video is helpful
- 90% say it is valuable (or will be valuable) to their research to have tissues mapped across studies

**Data type most interested in (searching for):**



**Data type most interested in accessing:**



# GTEEx / HuBMAP Survey

Compare bulk RNAseq from a GTEx tissue site to matching scRNAseq from HuBMAP	Cell data	Gene expression data	Tissue-specific data		
Shared and specific expression data	Cell data	Gene expression data	Tissue-specific data		
QTLs, genotypes by tissue and cell type	Cell data	Gene expression data	Tissue-specific data		Genotypes
Reference for disease studies	Cell data	Gene expression data	Tissue-specific data		
Immune cells across tissues	Cell data	Gene expression data	Tissue-specific data		
Cell type composition between tissues	Cell data	Gene expression data	Tissue-specific data		
Infer cell composition of GTEx tissues	Cell data	Gene expression data	Tissue-specific data		
Protein and single cell data with matching gene expression and methylation data	Cell data	Gene expression data			Protein/methylation
Compare gene expression from both along with clinical data	Cell data	Gene expression data			
Gene expression in Kidney		Gene expression data	Tissue-specific data		
Expression patterns		Gene expression data			
scRNA, scATAC, RNAseq		Gene expression data			
Gene expression		Gene expression data			
Expand GTEx data to include HuBMAP expression data, eventually spatial		Gene expression data		Spatial data	
Methylation data for epigenetics		Gene expression data			Protein/methylation
Compare gene expression and protein levels		Gene expression data			Protein/methylation
RNAseq		Gene expression data			
Spatial transcriptomics				Spatial data	
Organs: kidney, colon, prostate, spleen, lung			Tissue-specific data		

**For mapped tissues what data types were users mostly hoping to access, and what studies are they hoping to do:**

# Next Steps

We would like to expand the HuBMAP - GTEx collaboration to

1. Select CxGM data from at least two rich single cell data sources (GTEx and HuBMAP, and at least one additional published reference resource, and including any additional mapped CFDE single cell data if available) that have been spatially registered using the existing CCF Registration User Interface (RUI, <https://hubmapconsortium.github.io/ccf-ui/rui>) and compile a combined single cell data source for at least two tissue types (e.g., lung and heart).
2. Harmonize and annotate these data using Azimuth, CellTypist, or similar tools.
3. Use these mappings/comparison results to further iteratively enhance the registration process and accuracy, and help prioritize capture of meaningful sample metadata, mapping additional GTEx tissues as test cases.
4. Present results to other teams within the CFDE as well as teams working on major human atlas papers such as Tabula Sapiens and/or data portal teams (e.g., HCA, cellxgene, NeMO) for comments.