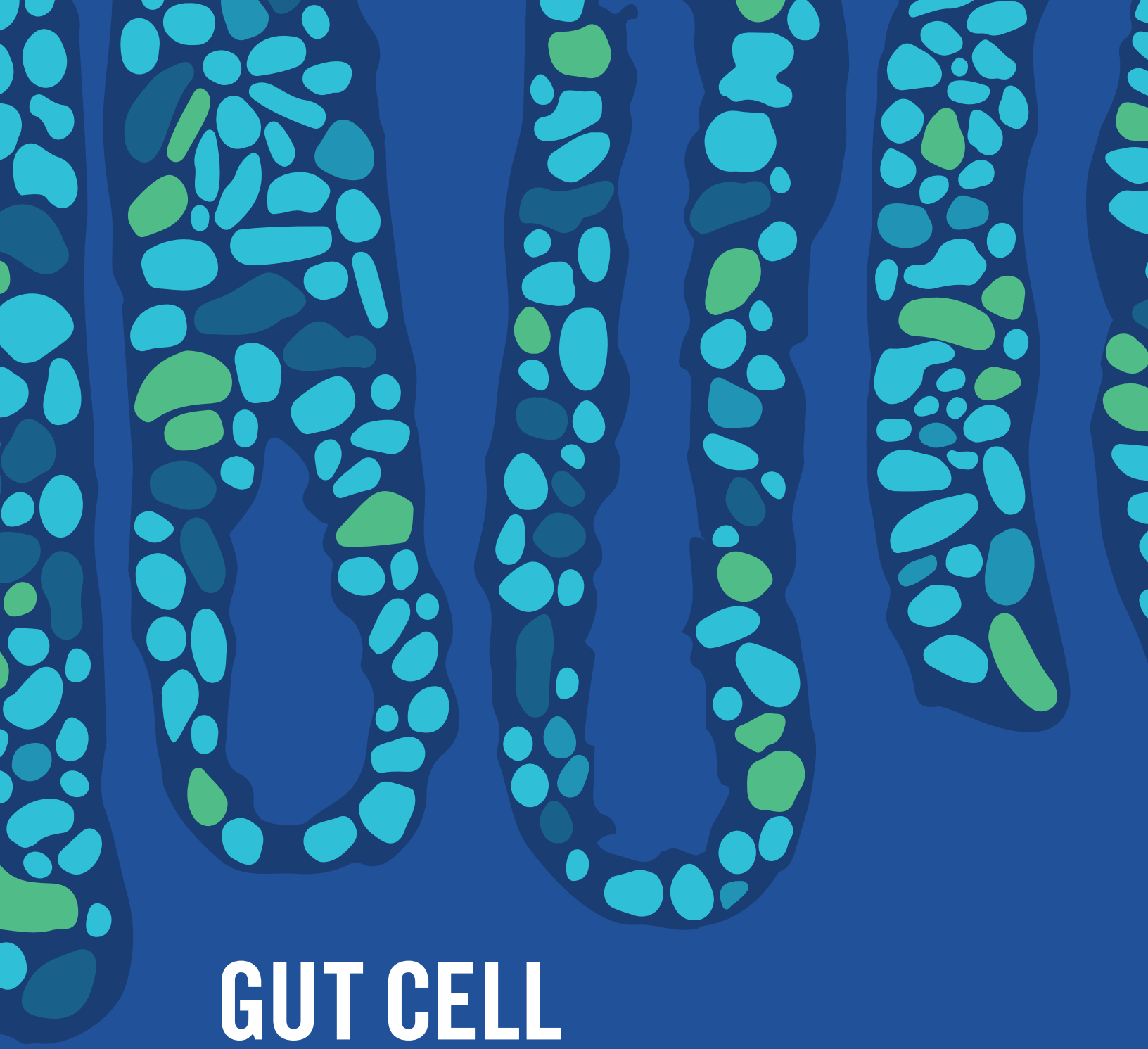


GUT CELL ATLAS

HCA Gut Roadmap

A document produced by the Broad Institute with funding from the Helmsley Charitable Trust

2021

HCA GUT ROADMAP

GUT CELL ATLAS

GOAL

Sufficiently sample the cellular entities in each unique tissue region in order to be able to place a cell in its spatial context given information about its characteristics.

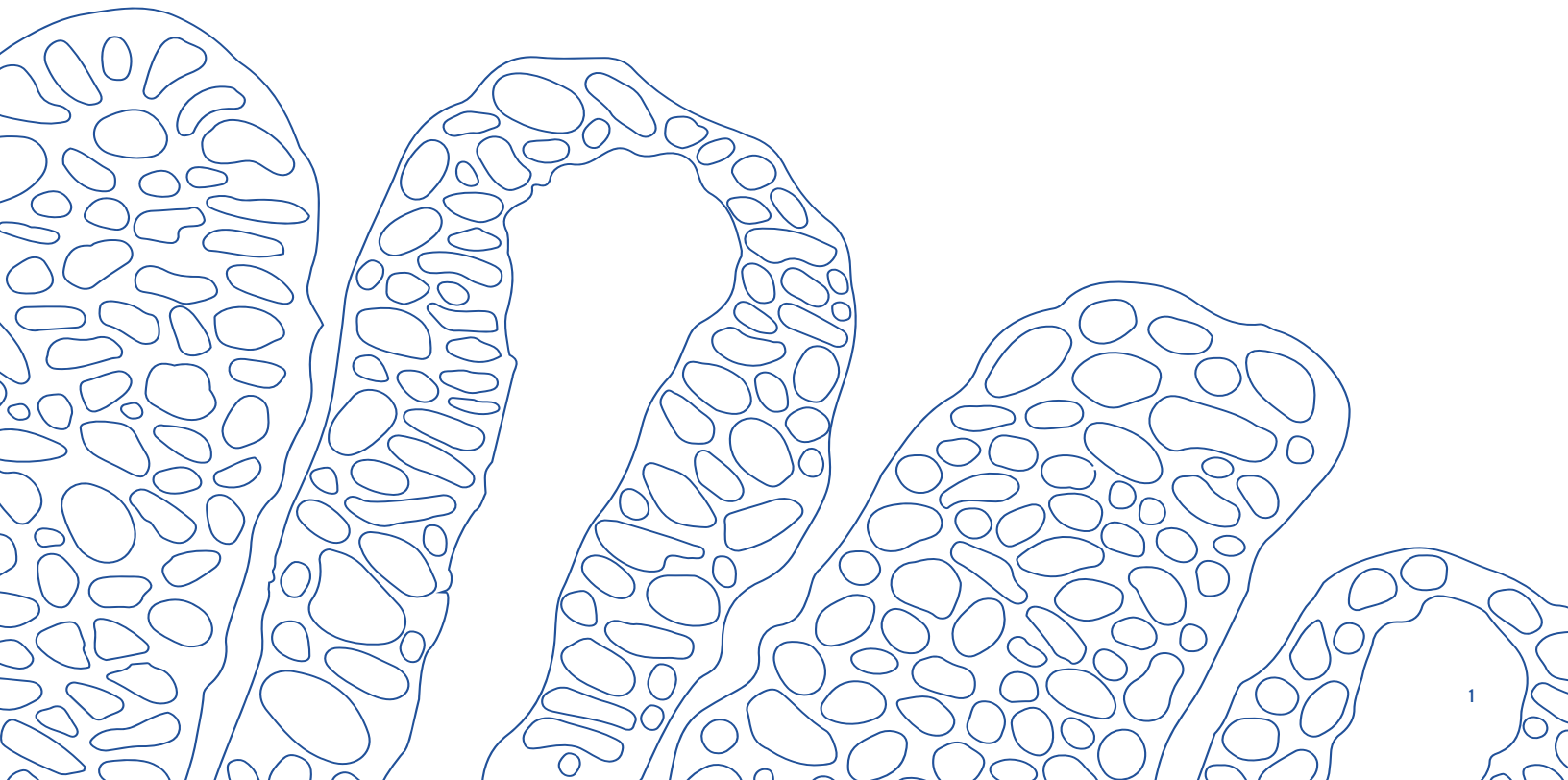
CHALLENGE

When sampling an organ, we must take into consideration:

I. natural variations in donor and organ size and shape;
II. anatomical landmarks and nomenclature, and
III. higher levels of precision sampling, such as a spatial grid.

These are essential for mapping the anatomical location of every sample from each organ in a comparable manner across individuals and sites.

Each part of this roadmap describes a comprehensive plan to sample each organ or system to ensure that the samples collected represent all the tissues' cellular diversity, and are coherent with each other such that the atlas results can be replicated among individuals and profiling sites.



C.1. GASTROINTESTINAL SYSTEM: SMALL INTESTINE AND COLON

We focus for now on two anatomical regions of the GI tract, the small intestine and the colon.

C.1.1. ANATOMICAL AND HISTOLOGICAL BACKGROUND

C.1.1.1. Anatomical organization

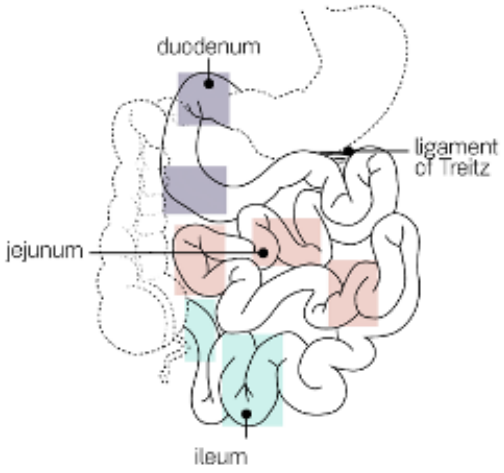
The ***small intestine*** is composed of three regions:

1. Duodenum
2. Jejunum
3. Ileum

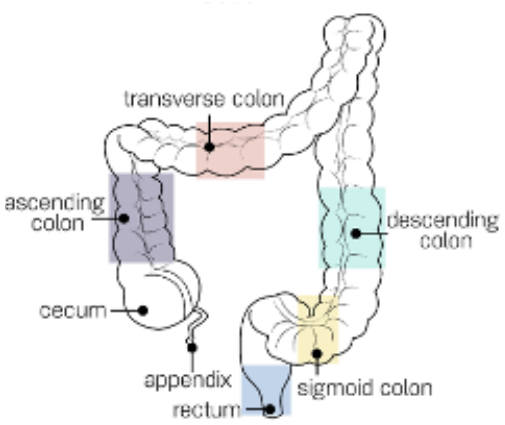
The ***colon*** is divided into six sections:

1. Cecum
2. Ascending colon (flanked by cecum and hepatic flexure)
3. Transverse colon (flanked by hepatic flexure and splenic flexure)
4. Descending colon (flanked by splenic flexure and sigmoid colon)
5. Sigmoid colon
6. Rectum

SMALL INTESTINE



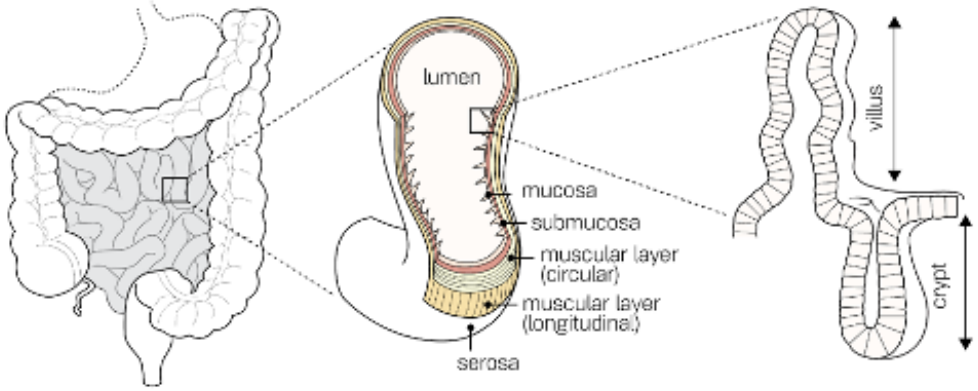
COLON



C.1.1.2. Histological organization

The small intestine and colon each have four histological layers:

1. mucosa (innermost layer; typically lined by columnar epithelium with mucous secreting goblet cells, including Peyer's patches in the small intestine);
2. submucosa (Meissner's plexus of the enteric nervous system and vasculature);
3. muscular layer or muscularis (inner circular layer and an outer longitudinal muscle layer, including myenteric plexus); and
4. serosa (the outermost layer of visceral peritoneum).



Within each histological layer, cells are organized in specific patterns. For example, in the mucosa of the small intestine, cells are organized along a crypt-villus axis, and vary both in terms of cell type and phenotype of a given cell type by their position within this axis (Moor et al., 2018, Haber et al., 2017). Similarly, the colon is organized along a crypt-lumen axis, but villi are absent in this tissue. The crypts are repeated units, such that two adjacent crypts may be indistinguishable. At the same time, within each defined region (e.g. duodenum), cells vary in their position, and, moreover, at longer length scales within one region, there is a gradient from proximal to distal, and cell programs and cell-type frequencies can change. Some differences may occur in disease – for example in the development of colon cancer there are differences between the left and right anatomical sides of the colon, which suggest additional distinctions in regional physiology.

C.1.1.3. Cellular composition

The small intestine and colon are diverse ecosystems of cells and we need to know the cell types and their proportions within the different regions of the tissue. For example, in the tissue's parenchyma, the epithelial cells are further subdivided into enterocytes (absorptive cells), goblet cells (mucin-secretors), Paneth cells (secreting anti-microbial defense peptides), Tuft cells (chemosensory cells), enteroendocrine cells (secreting

diverse hormones), Microfold (M) cells (which serve as a window to the luminal content) and undifferentiated stem cells, which give rise to transit amplifying cells and replenish the differentiated cells. In addition to regional differences, these cells are organized along the crypt-villus axis, for example in the small intestine stem cells and Paneth cells are located deep within crypts, and differentiation occurs as cells move towards villi.

Each of these cell types vary in their prevalence, and some are exceedingly rare, such as M cells. In order to capture all cell types regardless of their prevalence, enrichment, depletion (by molecular and/or histological methods), and spatial methods can be used. Bulk RNA and spatial methods should also be used to verify that no key cell types are missing in single-cell or single-nucleus analysis.

C.1.2. UNIQUE CHALLENGES

General challenges

- At the level of gross anatomy, humans have the same general structure to the GI tract, but its absolute length will vary from person to person based on age/stage of development, body size, diet habits (which may in turn be influenced by ethnicity and culture), other environmental factors such as circadian rhythms, and disease state. Ongoing efforts to build a Common Coordinate Framework (CCF) will help with sampling consistency.
- Although the gut is composed of repetitive units, with many folds and structures that look similar at first glance, there are key differences among the different anatomical and histological regions.
- The GI tract is heterogeneous and is composed of specialized cells spanning a continuous spectrum in an ongoing differentiation process, in addition to undergoing changes in disease states.
- Human sample quality is a challenge due to rapid degradation of intestinal tissues following surgical resection and/or post mortem.
- The heterogeneity in the different layers means the tissue prep needs to be considered carefully, as well as the stages of the preparation, e.g. when working with fresh samples, the epithelial cells are stripped off before the tissue is digested, as they would not survive the digest. This is important to detail in the phase I part of the project.
- The gut is a dynamic tissue, in which the composition and state of cells in one location may change upon external threats or changes to their local environment (such as infections, medications, contact with certain foods, etc.).

- Many of the anatomical landmarks that demarcate different regions of the small intestine – such as the ligament of Treitz (or the suspensory muscle), which serves as a boundary between the duodenum and the jejunum – are difficult to discern postmortem in autopsies. The same is true for some regions of the large bowel/colon, where the only clear landmarks are the cecum, and with it, the ascending colon and rectum.
- Although the sigmoid can be difficult to discern in autopsies, by taking multiple samples, it should be captured.
- In very young pediatric patients and donors, the gut is not yet fully developed.
- Other factors such as diet and circadian rhythm can affect gut composition and shape.
- It will be challenging to capture sufficient "healthy" samples as a base reference.

Disease-specific challenges

- The size and length of tissue, as well as the viability, state, and programs of the cells, change with disease (for example, tissue thickness, the orientation of crypt and villus, and lumen changes identified by MRI/CT) as does cell state and viability.
- Disease distribution, age of onset, duration of disease prior to onset/diagnosis, and different treatment regimens and responses to treatment.

C.1.3 UNIQUE ADVANTAGES AND OPPORTUNITIES

- The gut is rarely used for organ transplantation, and thus there is an opportunity to leverage organ donation programs to collect gut samples from organ donors.
- Colonoscopies are a routine medical procedure especially after the age of 50 for colon cancer screening, and increasingly in younger age groups.
- Upper endoscopies ([o]esophago-gastro-duodenoscopies; OGDs/EGDs) are also routine procedures and provide access to the upper GI tract mucosa, including duodenum up to D4.
- Small bowel endoscopy is an increasingly utilized technology and this could allow access to samples from the small bowel beyond the duodenum.

C.1.4 PROJECT ORGANIZATION

We will take a staged approach and the effort will proceed in three phases for both the healthy and disease atlases.

C.1.4.1. Healthy adults and pediatric samples

Phase 1: Pilot (2-3 data generation centers; several tissue sourcing sites)

The goals are to benchmark methods and sample sources, compare across methods and provide a scaffold for deeper exploration of regions of interest in both healthy and adult and pediatric samples. Collection of samples and profiling do not need to be coupled if frozen samples are processed.

In addition, **gut organoids** should be profiled during the pilot phase to determine how similar they are to tissues from human donors. Importantly, the degree of similarity will likely vary depending on stage at which the organoids are profiled (e.g., passage, level of differentiation, *etc.*). Ideally, organoids should be generated from the same sample that was subjected to scRNAseq.

Phase 2: Depth (deep dive; 2-3 data production centers; multiple tissue sourcing sites)

Once a sampling and analysis framework has been established from the comparative work in **Phase 1**, two to three centers should drive the data production phase, which will use fewer individuals to profile the tissues in greater depth, and guide the questions in **Phase 3**.

Phase 3: Breadth (design specific for genetic variation across individuals; additional specific cohorts (non-disease))

The goal is to **increase the number of individuals**, and use a sampling design/ strategy- specific based on knowledge gained from **Phase 2** (e.g. smaller number of methods and voxels per individual).

This phase can also include **genetic variation analysis and eQTL mapping** using sc/snRNA-Seq (with auxiliary scATAC-Seq and a multiplex spatial method). This will allow linking genetic variation to changes in gene programs and cellular composition, particularly as sample sizes become larger.

C.1.4.2 Disease atlas

Creation of a disease atlas will require a similar three-step approach:

Phase 1: (2-3 data generation centers, multiple tissue sourcing sites)

Benchmark adult and pediatric disease sample sources. Collection of samples and profiling do not need to be coupled if frozen samples are processed, but potential differences between profiling frozen versus fresh tissue must be considered. The goal is to demonstrate technical feasibility across sourcing sites.

Phase 2: Discovery/exploratory cohorts (multiple sites)

Small discovery/exploratory cohorts will be profiled. These cohorts can be much smaller than atlas-building cohorts because their purpose is to uncover main cell

types and states and cell proportions that can then be compared to the reference atlas to determine key changes in disease. They can also be compared to the healthy atlas as well as existing publicly-available datasets.

Phase 3: Hypothesis driven cohorts (a few sites)

A large hypothesis-driven disease cohort, built on the discovery and exploration of **Phase 2**. It will allow relation to genetic variation (as a QTL study) and clinical parameters. This would follow the framework design in **Phase 3** of the healthy atlas.

C.1.5. SAMPLING STRATEGY TO GENERATE A DEEP, VALIDATED AND GENERALIZABLE MAP OF HEALTHY TISSUE

C.1.5.1. Tissue sources

Samples should be drawn from three sources:

1. **Rapid autopsy** allows for the capture of all four tissue layers (mucosa, submucosa, muscularis, and serosa). However, as noted above, some of the landmark structures are not visible after death. Taking two or three samples in each region should help circumvent this problem. We recommend **12** autopsies total, **six** of each gender. These can rely on existing organ donation regional organizations.
2. **Biopsies** provide a snapshot of live tissue from healthy individuals, but only capture the mucosa and submucosa layers, and it is not possible to sample all regions of the small intestine and colon from living individuals in this way. For these layers, biopsies provide confirmation that cell types and compositions present in these samples can be found in the autopsies and that their gene signatures are similar as well. We recommend *at least* **12** biopsies for each of the numbered regions above (C.1.1.1) that can be biopsied, **6** of each gender. Biopsies could leverage new tools (e.g. tethered endoscope that provides positional information).
3. **Resections** containing healthy tissue are an additional sampling opportunity from living individuals, and should be analyzed **as available**. It is unlikely that all regions can be captured this way. By first benchmarking with autopsy and biopsy samples, the region of the resection can be determined. There are caveats for resections: (i) the bowel has been prepped, in a way that it is not in autopsy or biopsy, and (ii) even "uninvolved" tissue comes from sick patients (e.g. colorectal cancer, Crohn's disease). As such samples are generally from the margins of diseased tissue, whether they represent 'healthy' tissue is questionable and must be carefully considered, perhaps through comparison to biopsies. This will be especially true for pediatric samples; visually "normal-looking" samples may in fact not be molecularly "normal." These can also rely on new tools, such as 3D models/ reconstructions identifying areas for resection.

This will allow enough variation within the sampling sources due to technical variations (cell death, timing and etc.) to allow confidence that shared common features have been captured by all three sources. Biopsies are critical to provide a "normal" reference, since this may not be possible in autopsies and resections, but may not be possible in harder-to-reach regions, resulting in low sample numbers for these rarer regions. In this case, longitudinal sampling may be desirable.

In addition, **gut organoids** should be profiled to determine how similar they are to tissues from human donors, bearing in mind that gut organoids only represent the intestinal epithelium.

Where available, **co-located microbiome samples** can be collected from each of these sample types, but this is not a primary aim of the Human Cell Atlas. These sample types will be compared to each other to determine which samples are best to use for deeper profiling efforts in **Phase 2**. A challenge is that the protocols for obtaining mucosal-associated microbiome samples vary widely, and small differences will have a vast impact on microbial composition. It is important to apply the same protocol to obtain comparable results.

C.1.5.2. Where to sample

Small intestine: Sampling should be along the length of the following regions, which are visible regardless of sample source and will give a comprehensive view of the cells in the small intestine. **Current work** (*e.g.*, Smillie et al., Cell, 2019) suggests that the cellular composition is comparable within any one of these regions, and therefore the precise sampling location can vary slightly between individuals.

Overall, **18 samples** from **9 different locations** (**2** per location, to account for gradients) of the three **major regions** of the small intestine per individual (one autopsy or multiple actual individuals if by biopsy), as follows:

Duodenum: Two samplings each from

1. First (superior) part (D1), including duodenal bulb
2. D2/D3
3. Ascending (D4)

Jejunum (when possible, in disease state): **Two samplings** each from

1. Proximal
2. Middle
3. Distal

Ileum: Two samplings each from

1. Immediately proximal to the ileocecal valve
2. 6-12 centimeters proximal to the ileocecal valve

3. Between the two landmarks

   N.B. For biopsies, it may be most practical to sample the terminal ileum, as a distance of 6-12 centimeters during endoscopy cannot be determined reliably and there may not be major histological differences between these sites.

Colon

Overall, **12** samples should be collected per "individual" (equivalent to one autopsy or multiple actual individuals if by biopsy) from the **6** major regions of the colon per individual (**2** samples each to account for any gradients):

1. cecum
2. ascending colon (flanked by cecum and hepatic flexure)
3. transverse colon (ends at splenic flexure)
4. descending colon
5. sigmoid colon
6. rectum

Each sampled anatomical region will be further divided to anatomical tiles (voxels) localized to within a 1mm voxel by an experimental gridding procedure.
The anatomical tiles will be either frozen for subsequent assays and banking, or fixed, slide-mounted, stained, imaged and computationally atlas-mapped, to recover 3D atlas coordinates. We will select 3 random, non-adjacent anatomical tiles for each assay.

Pediatric sampling may need to be more limited, *e.g.* researchers may be allowed to take a maximum of 6-8 research biopsies during endoscopy. Samples will also be smaller as smaller forceps are used.

C.1.5.3. Measured profiles

Samples should be processed with three kinds of methods:

Anatomical-histological:

- CT
- MRI followed by serial histology (H&E) to allow deep registration
- 3D modeling from images (Larry Smarr)

Molecular profiling:

- Single-cell RNA-Seq or Single-nucleus sequencing
- Single-cell ATAC-seq
- Bulk whole exome sequencing
- Bulk RNA-Seq (for annotation)

Spatial profiling:

- One multiplex *in situ* RNA method
- One multiplex *in situ* protein method

Physiological/clinical profiling (meta data):

- To be defined
- Distance (from tethered endoscope)

C.1.5.4. Number and nature of individuals

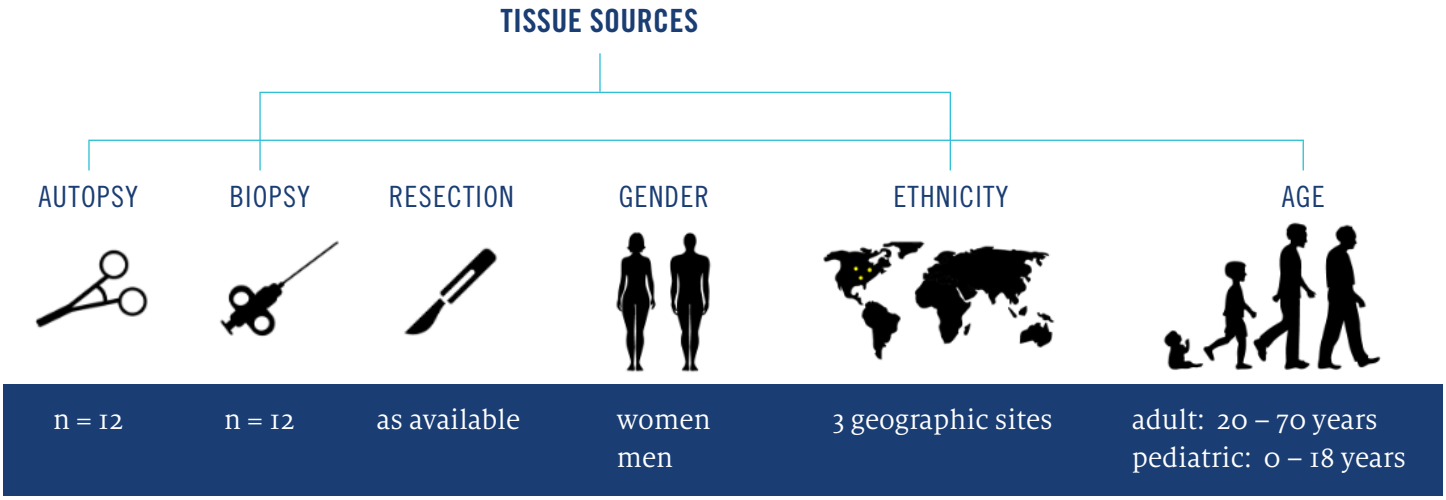
Individuals should span both genders, 3 geographic sites and a range of adult ages (20-70) as well as pediatric samples (from birth to 18 years of age).

ADULT ATLAS:

NUMBER OF DONORS	ADULT GENDER	ADULT AGE (20-50, 50+ YEARS)	ADULT GEOGRAPHICAL LOCATION
8 adults/site *3 sites = 24	12 from each = 24	2 in each age range per gender x 6 groups= 24 total	3 sites

PEDIATRIC ATLAS:

NUMBER OF DONORS	PED GENDER	PED AGE (0-18 YEARS)	PED GEOGRAPHICAL LOCATION
8 peds/site *3 sites = 24	12 from each = 24	For comparison to disease: 2 in each age range per gender x 3 groups (0-6 years, 6-10, 10-18)= 24	3 sites



SUMMARY

To create the first draft atlas of the small intestine and colon, 24 adults (12 autopsies and 12 biopsies for each region) and 24 pediatric donors should be sampled across 15 intestinal regions, 2 samples each, for a total of 30 samples per individual. In total, 60000 cells (with average of 2000 cells/sample) should be analyzed per individual.

ORGAN	REGION	# ADULTS (OR # OF PEDIATRIC DONORS)	# SAMPLES TOTAL	# CELLS
Small intestine	Duodenum, duodenal bulb (first/superior part)	24	48	96000
	Duodenum, second/descending part	24	48	96000
	Duodenum, between regions	24	48	96000
	Jejunum, proximal	24	48	96000
	Jejunum, middle	24	48	96000
	Jejunum, distal* (primarily autopsy)	24	48	96000
	Ileum, immediately proximal to the ileocecal valve	24	48	96000
	Ileum, 6-12 centimeters proximal to the ileocecal valve	24	48	96000
	Ileum, between regions	24	48	96000
Colon	Cecum	24	48	96000
	Ascending colon	24	48	96000
	Transverse colon	24	48	96000
	Descending colon	24	48	96000
	Sigmoid colon	24	48	96000
	Rectum	24	48	96000
Overall				1440000

THE LEONA M. AND HARRY B.
HELMSLEY
CHARITABLE TRUST

HUMAN
CELL
ATLAS