

Corpus Whole-body reference for human biology

Tissue-level molecular profiles, matched across organs within each donor and linked to clinical history.

Ambitious Bio generates whole-body molecular surveys at population scale for engineers in the life sciences. These surveys include solid tissue from about 30 sites per donor on average today (design target: more than 50), covering systems, organs, structures and cell types, plus matched biofluids and medical records. Corpus nests these measurements within each donor and scales across donors to maximize statistical power.

Matched biofluids and clinical histories provide a bridge between systems that are difficult or impossible to sample in living people and measurements routinely collected from living donors and over time –including EHR narratives and billing codes, blood-based measurements and imaging. These 'accessible' records and samples offer a limited window into the body when used in isolation. Corpus provides a deeper empirical foundation for interpreting them: how the distribution of molecules circulating in blood relates to their distribution and abundance in each solid tissue, and how these relationships vary across people.

Corpus is designed to span the full spectrum from health to disease; donors so far are healthy, and enrollment of diseased donors begins in 2027. Healthy tissue provides a baseline for understanding illness, which is defined relative to that baseline, and a view of the rest of the body that any treatment entering trials or clinical use must preserve. It also reaches beyond conditions routinely sampled by biopsy or surgery to diffuse, systemic, chronic and subacute diseases that rarely yield solid tissue during treatment, including conditions with large therapeutic markets.

MEASUREMENTS

Available today: Orbitrap Astral mass-spectrometry proteomics (~10,000-12,000 proteins per tissue). Beginning in 2027: 30× germline whole-genome sequencing per donor; 120× somatic whole-genome sequencing per tissue; short-read bulk RNA sequencing; long-read RNA sequencing for full-length transcript isoform profiling; and H&E tissue images, with single-cell and spatial transcriptomics on selected tissues.

Projected annual scale	2027 volume	2028 target
Donors	2,500	10,000
Tissue samples	130,000	520,000
Mass spectra	52-65 billion	208-260 billion
Protein quantifications	1.3-1.56 billion	5.2-6.24 billion
Full-length transcript reads	~1.3 trillion	~5.2 trillion
Compressed data (estimated)	~40,000 TB	~160,000 TB

Projections assume 52 tissue samples per donor.

Practical applications

Targets and safety. Map target expression across organs, compare healthy and diseased tissue, and prioritize tissues for validation and safety studies.

Biomarker development. Relate signals in biofluids to molecular abundance in specific tissues, giving blood-based measurements organ-level context.

New uses for medicines. Investigate shared molecular patterns across diseases and organs to identify additional indications for existing treatments.

Model training

Rules of gene regulation. Learn how genetic variation relates to RNA expression and protein abundance, and how these relationships differ by tissue.

From blood to organs. Learn relationships between circulating molecules and their possible tissue sources, distribution and abundance.

Health, aging and disease. Model how molecular relationships vary across ages and disease states, using healthy tissue as a baseline.

Tissue and donor data

Measurements are matched within each tissue, across organs of the same donor and, where available, to clinical history.

Data type	Measurement	Level
Proteomics	Orbitrap Astral mass spectrometry. Approximately 10,000-12,000 proteins per tissue; depth varies by tissue.	Tissue
Somatic genome	120× whole-genome sequencing.	Tissue
Short-read RNA	Bulk RNA sequencing.	Tissue
Long-read RNA	Full-length transcript isoform profiling.	Tissue
Histopathology	H&E-stained tissue images.	Tissue
Germline genome	30× whole-genome sequencing, linked to the donor's tissue profiles.	Donor
Clinical history	Partial medical records today; full EHR linkage and behavioral histories begin in 2027. Clinical records include Epic and Cerner.	Donor
Single-cell / spatial	10x Chromium single-cell RNA profiling. Spatial profiling is transitioning from Xenium/Visium to Atera.	Selected tissues

The table describes program scope. Proteomics is available today; other modalities begin in 2027. Tissue coverage and modalities are specified per cohort and release.

Also available

Adaptive sampling. Our collection rights allow additional tissue sampling and data generation, so questions raised by models can guide new experiments.

Translational reference data. Beginning in 2027, we will also offer reference data from the same material profiled across multiple generations of platforms, providing an empirical basis for aligning new data with existing datasets and the literature. We will also profile animal and human tissues on matched platforms to support translation between animal studies and human biology.

Linkage and supporting information

Profiles carry matched donor and specimen identifiers, tissue-site information, collection and processing metadata, assay protocols and quality-control information. This preserves each measurement's tissue and donor context.

Cohort definition

Cohorts can be used broadly or selected by tissue, modality, age, disease and treatment history, or cause of death. Inclusion and exclusion criteria, definitions and available counts are specified for each release.

We generate and own the experimental data, with full commercial rights to license and sublicense the dataset for research and model training. Collection is continuous.