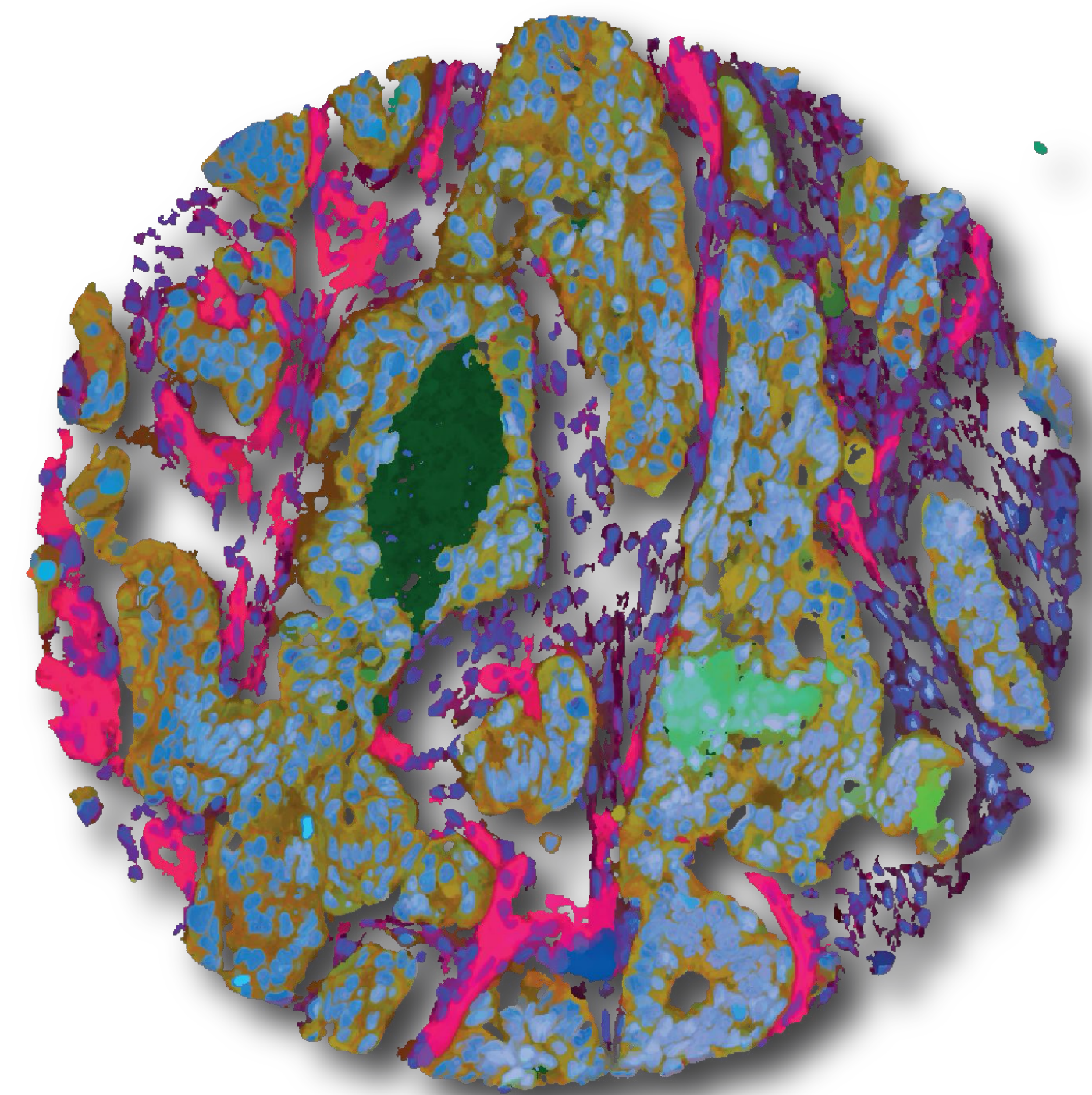


Building the Human Tumor Atlas Network as a rich multimodal source for the next era of data-driven cancer research



Adam Taylor¹, Jennifer Altruder², Kristen Anton³, Ino de Bruijn⁴, Ashley Clayton¹, David Gibbs⁵, Aditi Gopalan¹, Yamina Ktariya⁵, Alex Lash², Dar'ya Pozhidayeva⁵, Ethan Cerami⁵, Nikolaus Schultz⁴, Vésteinn Thorsson⁵

(1) Sage Bionetworks. (2) Dana-Farber Cancer Institute. (3) White River Computing. (4) Memorial Sloan Kettering Cancer Center. (5) Institute for Systems Biology

 Sage Bionetworks

 Memorial Sloan Kettering Cancer Center

 ISB

 Dana-Farber Cancer Institute

data.humantumoratlas.org

adam.taylor@sagebase.org

Mapping Cancer Transitions Across Space and Time

The Goal: Constructing comprehensive, 3D atlases of dynamic cancer transitions, ranging from precancerous lesions to primary tumors and metastatic progression.

Multimodal Integration: Combining high-resolution spatial transcriptomics and proteomics with single-cell sequencing to characterize the tumor microenvironment (TME) in depth

Diverse Scope: Encompassing 20 organ sites and 64 cancer types, the network generates a highly heterogeneous dataset that captures the architectural heterogeneity of cancer.

The Integration Challenge: The HTAN DCC harmonizes data into a unified resource, enabling cross-study analysis of cellular neighborhoods and therapeutic responses.

Twenty programs. One Data Coordinating Centre

HTAN at a glance

The HTAN data portal is the gateway to data from thousands of cancer patients profiled by multi-omic assays behind new insights in cancer progression.

20 atlases

20 organs

59 cancers

2.4k cases

WES/WGS
scRNA-seq
scATAC-seq
Visium
Xenium
CODEX
MIBI & IMC
MERFISH
MxIF / mIF
GeoMX
CosMX

10.5k specimens

31 assays

221k files

0.3B cells

Insights from HTAN Centers

HTAN's Research Centers use multimodal profiling to reveal the spatial and molecular drivers of cancer progression. Recent high-impact from HTAN Phase 1 include:

Distinct Paths to Malignancy: Defined how pre-malignant molecular programs and microenvironmental shifts dictate the route from polyp to CRC. (Chen et al., 2021, Cell)

Spatially Organized Immune Hubs: Discovered critical immune hubs between T cells and myeloid cells that organize anti-tumor immunity in the CRC TME. (Pelka et al., 2021, Cell)

Plasticity Drives Resistance: Revealed how therapy selects for spatially restricted, highly plastic cells to drive resistance in pancreatic cancer. (Cui Zhou et al., 2022, Nat Genet)

Immunoediting at the Boundary: Mapped the spatial landscape of primary melanoma to show how the immune system actively edits tumor evolution by eliminating clones at the tumor-stroma interface. (Nirmal et al., 2022, Cancer Discov)

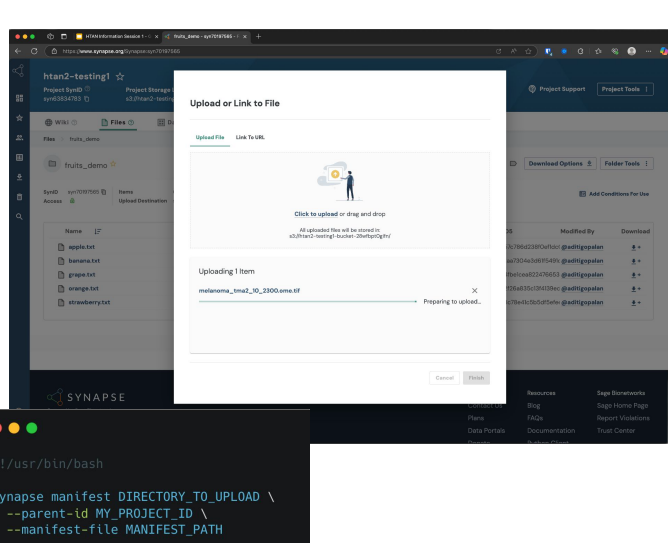
A 3D View of Cancer Transitions: Constructed a volumetric 3D atlas of colorectal cancer, revealing how cell-cell interactions and tumor states shift across the tumor's three-dimensional architecture. (Lin et al., 2023, Cell)

The HTAN data flow

1

HTAN members upload data to Synapse.org projects





2

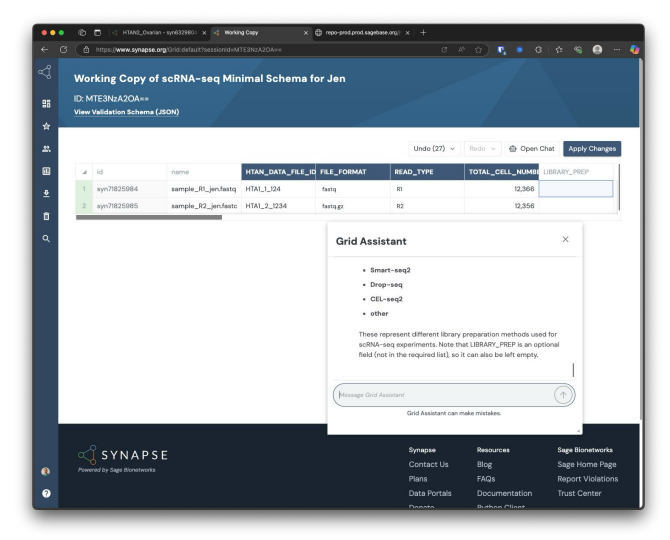
Metadata templates are populated with the help of a AI assistant and validated against the HTAN data model











3

Released data is listed on the portal, shared through the Cancer Research Data Commons General Commons, Synapse, and Google Bigquery tables through ISB-CGC









Tooling for delightful data discovery

The HTAN portal implements a range of tools to enhance data discovery and exploration.



Miniature thumbnails provide rich and informative immediate data preview for mIF & H&E images directly in the portal



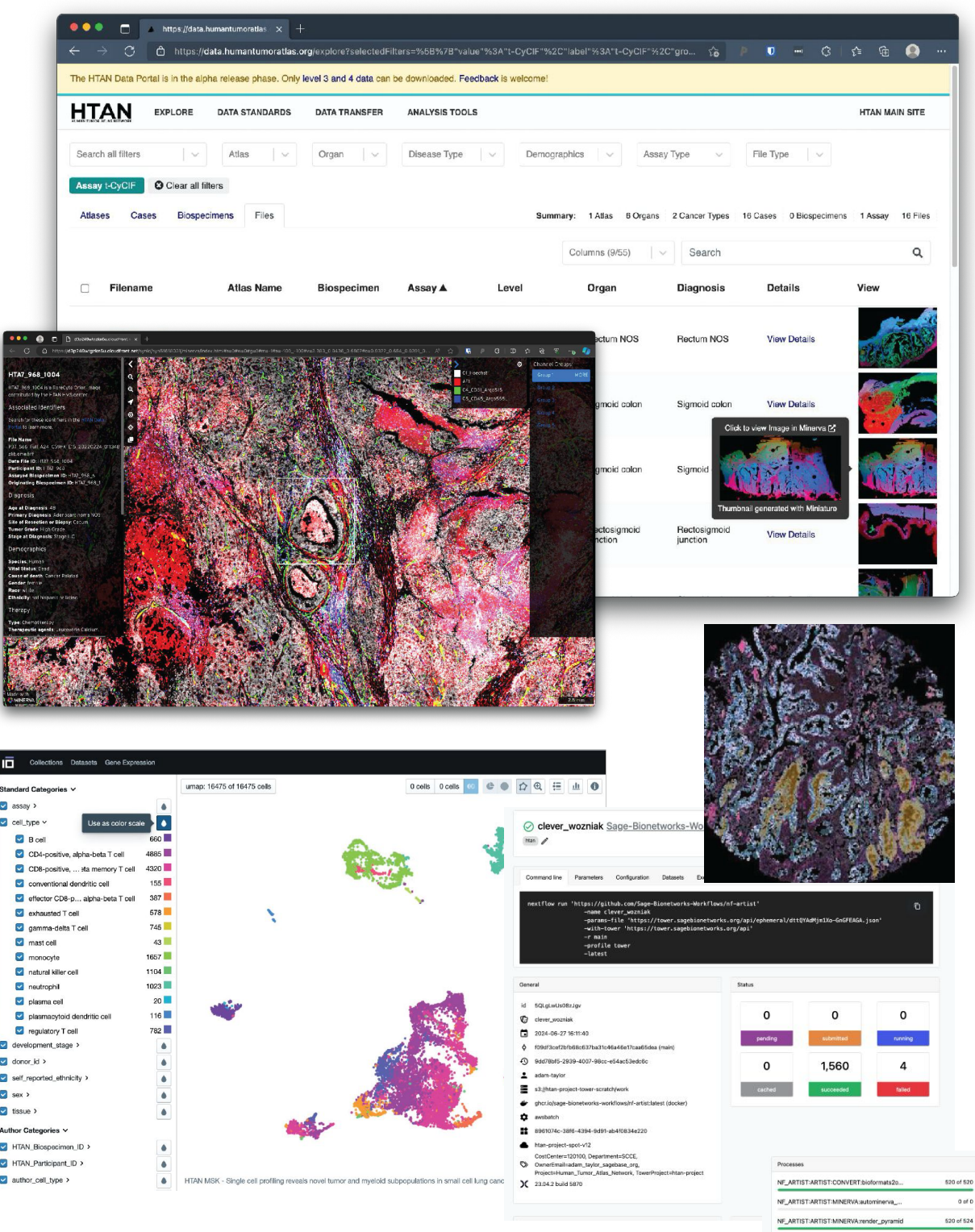
Minerva stories for interactive, guided and metadata-rich inspection of full-resolution mIF and H&E images



Curated CellxGene Discover datasets for exploration of single-cell and spatial transcriptomics data



Google BigQuery tables enable complex queries on clinical, biospecimen and file metadata



How will you use HTAN data in your research?

Train Foundation Models: Use high-resolution spatial and multi-omic datasets as ground truth for digital pathology and genomics models.

Validate Predictive Algorithms: Leverage diverse, multi-center cohorts as independent validation sets for clinical and diagnostic tools.

Discover Spatial Biomarkers: Correlate specific cellular neighborhoods and tissue architectures with disease progression and patient outcomes.

Benchmark Computational Tools: Access harmonized data to test and refine new methods for cell segmentation and multi-omic integration.

Find out more

De Bruijn et al. (2025). Sharing Data from the Human Tumor Atlas Network through Standards, Infrastructure, and Community Engagement. Nature Methods.

sagebionetworks.org

humantumoratlas.org

Acknowledgements

We thank all HTAN programs and patients, alongside NCI program officers and NCI CRDC staff. The HTAN DCC is funded by the National Cancer Institute (U24 CA233243).