

HTAN Spatial Transcriptomics Data Requirements

Author: Darya Pozhidayeva *ISB*

Updated: Jan 30, 2025

Status: Approved ▾

Terminology:

This document uses the following terms from [IEFT RFC 2119](#)

- **MUST / REQUIRED / SHALL:** ✓ (denotes absolute requirement)
- **MUST NOT / SHALL NOT:** ✗ (denotes absolute prohibition)
- **SHOULD / RECOMMENDED:** ☀️ (denotes recommendation)
- **SHOULD NOT / NOT RECOMMENDED:** 🙅 (denotes not recommended)
- **MAY / OPTIONAL:** 🤖 (denotes optional)

Before getting started:

When submitting spatial transcriptomics data in HTAN, it's crucial to understand the unique requirements associated with different platforms. For example, CosMx and Xenium data are submitted as bundled datasets. This means that the submission and metadata annotation will be level-agnostic. Conversely, Visium or GeoMx spatial informatics data isn't expected as a bundled package and will require a level-based approach. For more information, contact your DCC liaison. Each spatial platform's requirements are outlined below.

[CosMx](#)

[Xenium](#)

[GeoMx](#)

[Visium](#)

[Slide-Seq](#)

[MERFISH](#)

CosMx

CosMx Experiment Bundle

- ✓ CosMx Experimental Bundles **MUST** consist of a single biospecimen per zipped file.
- ✓ The zipped experimental submission **MUST** contain a directory for Cell Overlay data.
- ✓ The zipped experimental submission **MUST** contain a directory for Cell Label data.
- ✓ The zipped experimental submission **MUST** contain a directory for Cell Composite data.
- ✓ The zipped experimental submission **MUST** contain a FOV position file.
- ✓ The zipped experimental submission **MUST** contain a matrix count file.
- ✓ The zipped experimental submission **MUST** contain a metadata file.
- 🙋 Submission of all Images (raw or processed) in this level is **OPTIONAL**.
- ✓ You **MUST** adhere to all the rules that apply to imaging submission defined in [Image ingress documentation](#)

Xenium

Xenium Experiment Bundle

- ✓ Xenium Experimental Bundles **MUST** consist of a single biospecimen per zipped file.
- ✓ The zipped experimental submission **MUST** contain a cell group file.
- ✓ The zipped experimental submission **MUST** contain a gene panel file.
- ✓ The zipped experimental submission **MUST** contain a gene group file.
- ✓ The zipped experimental submission **MUST** contain a count matrix file.
- 🙋 Submission of all Images (raw or processed) in this level is **OPTIONAL**.
- ✓ You **MUST** adhere to all the rules that apply to imaging submission defined in [Image ingress documentation](#)

GeoMx

NanoString GeoMx DSP Spatial Transcriptomics Level 1

 Submission of this level is **OPTIONAL**.

- ✓ GeoMx Level 1 Files **MUST** contain .RCC or .DCC file types.
- ✓ GeoMx Level 1 Files **MUST** link the synapse ID of a GeoMx experimental PKC file.
- ✓ GeoMx Level 1 Files **MUST** link the synapse ID of a GeoMx lab worksheet file.

NanoString GeoMx - Images

- ✓ You **MUST** adhere to all the rules that apply to imaging submission defined in [Image ingress documentation](#) and should be submitted under Imaging Level 2.

NanoString GeoMx DSP Spatial Transcriptomics Level 3

 Submission of this level is **OPTIONAL**.

- ✓ GeoMx Level 3 Files **MUST** contain processed data from the NanoString GeoMx DSP Pipeline.
- ✓ GeoMx Level 3 Files **MUST** depend on GeoMx Level 1 and Imaging Level 2.
- ✓ GeoMx Level 3 Files **MUST** link the synapse ID of a GeoMx Segment Annotation Metadata.

NanoString GeoMx DSP ROI DCC/RCC Segment Annotation Metadata

- ✓ Submission of this level is **REQUIRED** for all GeoMx experiment types.
- ✓ This metadata template **MUST** contain ROI identifiers associated with a single HTAN biospecimen.
- ✓ This metadata template **MUST** contain X and Y coordinates for a specified ROI region.
- ✓ A separate metadata template **MUST** be submitted for each DCC or RCC experimental component. For example if an experiment has RCC and DCC files, each must be annotated and submitted.

☀ The RCC and DCC metadata template is based closely on the Segment Summary File as it is exported from the GeoMx Software. Therefore, that file can be submitted so long as the column names match the metadata template (manual adjustment is necessary).

Visium

10x Visium Spatial Transcriptomics - RNA-seq Level 1

- ✓ Visium RNA-seq Level 1 Files **MUST** contain FASTQ files
- ✓ Visium RNA-seq Level 1 Files **MUST** associate each file with a Visium Slide ID and Slide Region
- ✗ Visium RNA-seq Level 1 Files **MUST NOT** contain multiple biospecimen IDs per slide region.

10x Visium Spatial Transcriptomics - RNA-seq Level 2

- ✓ Visium RNA-seq Level 2 Files **MUST** contain BAM files

10x Visium Spatial Transcriptomics - RNA-seq Level 3

- ✓ Visium RNA-seq Level 3 Files **MUST** contain a count matrix file (raw or normalized).

10x Visium Spatial Transcriptomics - RNA-seq Level 4

- ✓ Visium RNA-seq Level 4 Files **MUST** contain an h5ad or TSV file.

10x Visium Spatial Transcriptomics - Auxiliary Files

- ✓ Visium RNA-seq Auxiliary Files **MUST** contain a TIFF image.
- ✓ You **MUST** adhere to all the rules that apply to imaging submission defined in [Image ingress documentation](#)

Slide-Seq

Slide-seq Level 1

- ✓ Slide-seq Level 1 Files **MUST** contain FASTQ files

Slide-seq Level 2

- ✓ Slide-seq Level 2 Files **MUST** contain BAM files

Slide-seq Level 3

- ✓ Slide-seq Level 3 Files **MUST** contain a count matrix file (raw or normalized).
- ✓ Slide-seq Level 3 Files **MUST** contain bead matched and unmatched barcode location files.
- ✓ Bead Barcode Files **MUST** contain X and Y coordinates for bead locations on the Slide-seq puck

MERFISH

Imaging Level 2

- ✓ You **MUST** adhere to all the rules that apply to imaging submission defined in [Image ingress documentation](#)
- ✓ You **MUST** provide a synapse ID for a MERFISH Positions File
- ✓ You **MUST** provide a synapse ID for a MERFISH Codebook File