



Consortium for Advancing Management and Prevention of Cancer in People with HIV Data Commons (AMC-DC)

Data Standardization Plan for Tissue Imaging

Template v2.0; 07APR2021

Version 2.0, 14OCT2026

TABLE OF CONTENTS

1. GLOSSARY	3
2. INTRODUCTION	3
3. RAW DATA FILES AND FORMAT	4
3.1 Raw Data Files	4
3.2 Raw Data Format Description	5
4. DATA PRE-PROCESSING	7
4.1 Data Pre-Processing Steps	7
4.2 Final Processed Data Format	10
5. EXPERIMENTAL INFORMATION AND KEY META DATA DESCRIPTORS	29
6. QUALITY METRICS	32
7. MANIFEST OF FILE COLLECTION TO BE SUBMITTED TO THE ODMC	35
8. ODMC DATA PROCESSING	37
9. APPROVAL	37
10. APPENDIX	38

1. GLOSSARY

Data Formats: A collection of requirements / templates to ensure that the data adheres to a certain standard including minimal required data elements, accepted values, control vocabulary and/or ontologies to be used for certain fields.

Metadata: Collection of data to be collected about the data. This includes information about each field to be collected for a certain assay, experimental information such as instrument, instrument configuration, measurement unit/description, descriptions of the file collection, and other information that is valuable for analysis. Minimal meta data descriptors are a subset of meta data that are considered to be essential.

Operations and Data Management Center (ODMC): provides operational, logistical, communications, contracting, protocol development, site management, data and specimen management, monitoring, regulatory and medical monitoring support to the AMC.

Statistics and Data Analysis Center (SDAC): provides statistical design and analysis of AMC studies.

AMC Network Research Laboratories (NRLs): provide expertise in virology, viral oncology (e.g., KSHV, HPV, EBV), pharmacology, pathology, hematopathology, genomics, and advanced HIV diagnostics for correlative studies in support of AMC clinical trials.

Data Standardization Teams: Develop Data Standardization Plans for each assay type specifying expected data pre-processing, sample quality metrics, data formats, data processing workflows, and experimental meta data descriptors.

2. INTRODUCTION

This document outlines specifications for raw data files/formats, data pre-processing, quality metrics, and experimental metadata descriptors for the following assays: tissue imaging.

Data submitted to the ODMC must follow the standards included in this document. This includes laboratory data generated by AMC network research laboratories (NRLs) and affiliate laboratories.

For each tissue imaging dataset, the following tables will be provided along with a manifest of the files:

1. Raw data in [Protocol ID]_[assay]_[date]_raw_data.tiff.
2. Pre-processing workflow steps in [Protocol ID]_[assay]_[date]_workflow.csv, format details in Table 3.
3. Processed data in [Protocol ID]_[assay]_[date]_exp_data_.csv, format details in Table 5.
4. Image linkage data in [Protocol ID]_[assay]_[date]_image_linkage.csv
5. Experimental metadata in [Protocol ID]_[assay]_[date]_exp_metadata.csv, format details in Table 7.
6. Quality metrics in [Protocol ID]_[assay]_[date]_quality_metrics.csv

Multiple complete sets of data files will be provided if testing occurred for multiple different assay types.

3. RAW DATA FILES AND FORMAT

3.1 Raw Data Files

Table 1A: Raw Data File Existing Standards Overview

File Extension/T ype	Description of File Content	Reference
.tif, .tiff	Tagged Image File Format, a bitmap graphics file format utilizing tagged fields.	https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C70631
.svs	Spectral View Settings format, an electronic file format based on TIFF that is used for storing multiple high-resolution images in a single electronic file.	https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C172214
.jpg, .jpeg	Joint Photographic Experts Group format, a file compression format mostly used for color and greyscale photographs.	https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C48230
.vms, .vmu, .ndpi	Hamamatsu formats, including Virtual Microscope Specimen (VMS), Uncompressed Virtual Microscope Specimen (VMU), and Nanozoomer Digital Pathology Image (NDPI), are proprietary file formats used for storing whole slide images. VMS and VMU formats typically include an index file, multiple image files, and an optimization file (for VMS). NDPI is a single TIFF-like file with custom TIFF tags. VMS and NDPI formats contain JPEG-compressed images, while VMU files use NGR, a custom uncompressed format.	https://openslide.org/formats/hamamatsu/
.mrxs	Multi-file with very complicated proprietary metadata and indexes, a MRXS file is an image file created by MIRAX-compatible microscope digital slide scanners. It stores images of multiple specimen samples from glass slides on a single digitized virtual slide. MRXS files are used to archive and analyze microscope images.	https://openslide.org/formats/mirax/
.dcm, .dicom	Digital Imaging and Communications in Medicine, a standard file format for medical images that contains images along with details about the patient, the scan that generated the image and the characteristics of the image itself.	https://www.dicomstandard.org/using/cds
.csv	Comma separated values format, a text file format that separates data elements with commas.	https://ncithesaurus.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=ncit&code=C182456

Table 1B: Raw Data File Overview

Assay(s)	File Extension/Type	Description of File Content	Reference
Biopsy Morphological Assays Results	.csv	Biopsy data received as processed files, see Table # for details.	n/a
Biopsy immunohistochemical Assays Results	.csv	Immunohistochemistry data received as processed files, see Table # for details.	
Immunohistochemistry (IHC)	.jpg, .jpeg	Compressed IHC image	
Immunohistochemistry (IHC)	.tif, .tiff	IHC raw data	https://learn.canceridc.dev/dicom/dicom-tiff-dual-personality-files
Histopathology slide microscopy images	.jpg, .jpeg	Compressed slide microscopy scanned image file	
Histopathology slide scanned microscopy images	.tif, .tiff	Slide microscopy scanned image file raw data	https://openslide.org/formats/mirax/
De-identified native or converted assay images	.dcm, .dicom	De-identified DICOM files, free from PHI or PII	https://www.dicomstandard.org/ https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9141493/ https://www.oit.va.gov/Services/TRM/ToolPage.aspx?tid=9204

3.2 Raw Data Format Description

Slide images provided to the ODMC will include file names that contain Global Trace specimen number (9-digit GTID)-Stain-Image #. The stain convention will be: h(H&E), p(p16), k (ki67), u (unstained) and o (other).

Table 2: Example of Raw Data File Details

Data File Name	File extension/type	Original Lab accession number	File Description	Additional data	Reference
123456789-h-1	.tif	Ap23-00456	Oropharyngeal carcinoma H&E slide scanned image	Incisional biopsy	
123456789-h-2	.tif	Ap23-01058	Oropharyngeal carcinoma H&E slide scanned image	Incisional biopsy	
123456789-h-3	.tif	Ap23-02486	Oropharyngeal carcinoma p16 slide scanned image	Surgical excisional specimen	
123456789-p-1	.tiff	IH23-00089	Oropharyngeal carcinoma p16 immunoh slide representantive picture	n.a.	
123456789-p-2	.tiff	IH23-0097	Oropharyngeal carcinoma p16 immunoh slide representantive picture	n.a	
123456789-p-3	.tiff	IH23-128	Oropharyngeal carcinoma p16 immunoh slide representantive picture	n.a	

4. DATA PRE-PROCESSING

4.1 Data Pre-Processing Steps

The following data pre-processing steps will be performed. Specific software and database versions will be reported in the [Protocol ID]_[assay]_[date]_workflow.csv table.

Table 3: Example of Data Pre-Processing Steps

Example ID	Process Description	Software	Input File Type	Description of Input File Content	Output File Type	Description of Output File Content
1	Image Quality Assessment	ImageJ/FIHI	.svs	Whole slide image captured from a digital pathology scanner	.svs	Quality-checked whole slide image, flagged for issues if applicable
2	Color Normalization	QuPath (open source software)	.svs	Whole slide image with potential color variation	.svs	Color-normalized image with consistent staining across slides
3	Image Registration	Elastix	.tiff	Series of tissue images from different modalities	.tiff	Registered images aligned to a common coordinate system
4	Artifact Removal	MATLAB	.tiff	Image containing artifacts such as bubbles or folds	.tiff	Cleaned image with artifacts removed
5	Segmentation	CellProfiler	.tiff	Tissue image containing regions of interest (e.g., tumor)	.tiff	Segmented image with regions of interest highlighted
6	Annotation	Aperio ImageScope	.svs	Annotated tissue slide with regions marked by pathologists	.svs	Annotated image with marked regions stored as overlay
7	Downsampling	Python (OpenSlide)	.svs	High-resolution whole slide image	.tiff	Downsampled image for easier handling
8	Format Conversion to TIFF	Bio-Formats	.ndpi	NDPI file with whole slide image and metadata	.tiff	TIFF file with image data for analysis
9	Format Conversion to DICOM	DICOMizer, ImageJ (with DICOM plugin), or custom script	.tiff, .svs, .jpeg	Image file from whole slide imaging or other formats	.dcm (DICOM)	DICOM file with image data and associated metadata converted from original format

Example ID	Process Description	Software	Input File Type	Description of Input File Content	Output File Type	Description of Output File Content
10	DICOM De-identification	DicomCleaner, Amazon HealthLake Imaging, or custom script	.dcm, .dicom	Native or converted assay images with PHI or PII	.dcm, .dicom	De-identified native or converted assay images free of PHI or PII
11	Image analysis of Ki-67	QuantCenter (3D-histech	.mrxs	Full scanned HE slides of invasive squamous carcinoma	.xls	.xls file with percentage of Ki-67 positive cells
12	IHC p16 expression evaluation	Not automated evaluation	.mrxs	Full scanned p16 immunohistochemistry slides of invasive squamous carcinoma	.xls	.xls with positive/negative results
13	Image analysis of CD3 lymphocytes infiltration	QuPath (open source software)	.TIFF	Picture of the highest density area of CD3 tumor infiltration	.xls	.xls with number of CD3 lymphocytes per mm2

Data processing specifications will be reported in the following format. Requirements for collection are noted in the last column titled “Collection Requirement”. Data processing specifications will be submitted in a [.csv] file to the ODMC.

- Filename will be [Protocol ID]_[assay]_[date]_workflow.csv
- The data must have the columns listed exactly as in the “Variable Name” column in the table below. Included values must be included as listed in the “Validation Rules” column in the table below.

Table 4: Variables to be Submitted in [Protocol ID]_[assay]_[date]_workflow.csv

Variable Name	Type	Length	Description	Reference	Validation Rules	Collection Requirement
Protocol	Num	5	Protocol ID	--	Must match ODMC-specified value for the study	Required
Software Version	Num	8, 1	Concentriq v1.0	https://proscia.com/concentriq-platform/	Expected to have format x.x (e.g. 6.1)	Required
Tumor diagnosis	Num	1	0=no tumor 1=intraepithelial squamous lesion high grade 2= invasive squamous carcinoma	http://www...	Must be recorded in a single algorithm	Required
ki-67	percentage	2	Percentage of positive neoplastic cells	http://www....	Must be recorded in the following format: XX%	Required
p16	Num	1	0=negative, 1=questionable, 2=positive	https://....	Must be recorded in a single algorithm	Required

4.2 Final Processed Data Format

Data will be submitted in the following processed data format. Requirements for collection are noted in the last column titled “Collection Requirement”. Processed data will be submitted in a [.csv] file to the ODMC.

- Filename will be [Protocol ID]_[assay]_[date]_exp_data.csv
- The data must have the columns listed exactly as in the “Variable Name” column in the table below. Included values must be included as listed in the “Controlled Vocabulary” and “Validation Rules” columns in the table below.

Table 5: Variables to be Submitted in [Protocol ID]_[assay]_[date]_exp_data.csv

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
Protocol	Num	5	Protocol ID	--	Must match ODMC-specified value for the study	Required
Specimen ID	Num	9	Specimen number, refers to the barcode number printed on each microscopy slide	--	Must be 9 characters long. If a Specimen is being entered as a reference (control) sample, the SN will be labeled as [specified value]	Required
Participant ID	Num	13	AMC study participant ID	N/A	Must match list of enrolled participants from ODMC	Optional

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
Test Date	Date	8	Date of test	--	xx/xx/xxx x (mm/dd/yy yy)	Required
Collection Date	Date	10	Date of biopsy collection	N/A	xx/xx/xxx x (mm/dd/yy yy)	Required
Visit Number	Num	3	Study visit number	--	Match list of available visit numbers for given study	Required

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
Specimen Type	Char	100	Specimen Type When differentiating “Perianal Distal” and “Perianal Proximal”, 2 cm is the cutoff delineating the two types.	SCJ Proximal to SCJ Perianal proximal Perianal distal Cervix Endocervical canal Vagina Vulva Endometrium Oral mucosa GI mucosa, specify site: Respiratory mucosa, specify site: Other mucosa, specify site: Visceral non-mucosa, specify site: Lymph node(s), site not specified Lymph node(s); Specify site(s): Other (specify): Other tissue(s) or organ(s) (specify): Not specified	Must match one of the values listed in the Controlled Vocabulary column. <i>Enter all that apply</i>	Must match one of the values listed in the Controlled Vocabulary column. <i>Enter all that apply.</i>
Specimen Type, Relative Location	Char	25	Location relative to specimen type defined	Anterior, Posterior, Right Lateral, Left Lateral, Other: [specify]	Must match one of the values listed in the Controlled Vocabulary column	

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
Specimen Collection Method	Num	3	Type of tissue acquisition	Biopsy, Excision, Recession, Other	Must match one of the values listed in the Controlled Vocabulary column	Required
Specimen Collection Method, Other (Specify)	Char	50	Procedure of specimen collection if not listed in appendix	n/a	n/a	Required if Specimen Collection Method value is 16 (other)
Specimen Adequacy	Char	50	Specimen adequacy for diagnosis.	Adequate, Suboptimal: [with rationale], Insufficient for evaluation: [with rationale]	Must match one of the values listed in the Controlled Vocabulary column	Required
Tumor Site, Anatomical Location	Char	50	Anatomical location from which tissue was removed	n/a	n/a	Required

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
Laterality and Directionality	Num	1	Biopsy site or tumor location	1, Left superior (anterior) quadrant (12 to 3 o'clock) 2, Left inferior (posterior) quadrant (3 to 6 o'clock) 3, Right inferior (posterior) quadrant (6 to 9 o'clock) 4, Right superior (anterior) quadrant (9 to 12 o'clock) 5, Other 6, Cannot be determined	Must match one of the values listed in the Controlled Vocabulary column	Optional
Laterality and Directionality, Specify	Char	50	Provide additional details for Laterality and Directionality, if necessary.	n/a	n/a	Required if Laterality and Directionality values are 5 (Other) or 6 (Cannot be determined)
Histologic Type	Num	3	A description of a tumor based on how abnormal the cancer cells and tissue look under a microscope and how quickly the cancer cells are likely to grow and spread.	see code list in appendix	Must match one of the values listed in the Controlled Vocabulary column	Required
ICC Classification	Char	20	International Consensus Classification	n/a	n/a	Optional for Lymphoma histologies
Histologic Type, Other (Specify)	Char	200	Enter details of histology if "Other, specify" was Selected for Histological Type	n/a	n/a	Required if Histologic Type Value is 142 (Other)

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
IHC p16	Num	1	Immunophenotyping marker. Commonly used for Anal and Gynecologic Cancer assessments.	1, Positive (block-like staining) 2, Negative (patchy or non-block staining) 3, Negative (no staining) 4, Not available	Must match one of the values listed in the Controlled Vocabulary column	Required for Anal and Gynecologic Cancer assessments.
HHV-8 (LANA-1)	Num	1	Immunophenotyping marker. Commonly used for Kaposi Sarcoma Assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Kaposi Sarcoma Assessments
BCL2	Num	1	Immunophenotyping marker. Commonly used for Mature B-Cell non-Hodgkin's lymphoma (large and small cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Mature B-Cell non-Hodgkin's lymphoma (large and small cell subtypes) assessments

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
BCL6	Num	1	Immunophenotyping marker. Commonly used for Nodular lymphocyte predominant Hodgkin's Lymphoma and Mature B-Cell non-Hodgkin's lymphoma (large cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Nodular lymphocyte predominant Hodgkin's Lymphoma and Mature B-Cell non-Hodgkin's lymphoma (large cell subtypes)
BOB.1	Num	1	Immunophenotyping marker. Commonly used for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma
CD10	Num	1	Immunophenotyping marker. Commonly used for Mature B-Cell non-Hodgkin's lymphoma (large and small cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Mature B-Cell non-Hodgkin's lymphoma (large and small cell subtypes)

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
CD15	Num	1	Immunophenotyping marker. Commonly used for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma assessments
CD20	Num	1	Immunophenotyping marker. Commonly used for Classic Hodgkin's Lymphoma, Nodular lymphocyte predominant Hodgkin's Lymphoma, and Mature B-Cell non-Hodgkin's lymphoma (large and small cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Classic Hodgkin's Lymphoma, Nodular lymphocyte predominant Hodgkin's Lymphoma, and Mature B-Cell non-Hodgkin's lymphoma (large and small cell subtypes) assessment
CD23	Num	1	Immunophenotyping marker. Commonly used for Mature B-Cell non-Hodgkin's lymphoma (small cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Mature B-Cell non-Hodgkin's lymphoma (small cell subtypes)

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
CD30	Num	1	Immunophenotyping marker. Commonly used for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma assessments
CD31	Num	1	Immunophenotyping marker. Commonly used for Kaposi Sarcoma Assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Kaposi Sarcoma Assessments
CD34	Num	1	Immunophenotyping marker. Commonly used for Kaposi Sarcoma Assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Kaposi Sarcoma Assessments

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
CD45	Num	1	Immunophenotyping marker. Commonly used for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin's Lymphoma assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin's Lymphoma assessments
CD57	Num	1	Immunophenotyping marker. Commonly used for Nodular lymphocyte predominant Hodgkin's Lymphoma assessments	0, Not Increased 1, Increased	Must match one of the values listed in the Controlled Vocabulary column	Required for Nodular lymphocyte predominant Hodgkin's Lymphoma assessments
CD5	Num	1	Immunophenotyping marker. Commonly used for Mature B-Cell non-Hodgkin's lymphoma (small cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Mature B-Cell non-Hodgkin's lymphoma (small cell subtypes) assessments

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
CD79a	Num	1	Immunophenotyping marker. Commonly used for Nodular lymphocyte predominant Hodgkin lymphoma and Mature B-Cell non-Hodgkin's lymphoma (large cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Nodular lymphocyte predominant Hodgkin lymphoma and Mature B-Cell non-Hodgkin's lymphoma (large cell subtypes) assessments
CD138	Num	1	Immunophenotyping marker. Commonly used for Mature B-Cell non-Hodgkin's lymphoma (large cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Mature B-Cell non-Hodgkin's lymphoma (large cell subtypes) assessments
Cyclin D1/SOX11	Num	1	Immunophenotyping marker. Commonly used for Mature B-Cell non-Hodgkin's lymphoma (small cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Mature B-Cell non-Hodgkin's lymphoma (small cell subtypes) assessments

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
EBER	Num	1	Immunophenotyping marker. Commonly used for Kaposi Sarcoma, Nodular lymphocyte predominant Hodgkin lymphoma, Mature B-Cell non-Hodgkin's lymphoma (large and small cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Kaposi Sarcoma, Nodular lymphocyte predominant Hodgkin lymphoma, Mature B-Cell non-Hodgkin's lymphoma (large and small cell subtypes) assessments
EMA	Num	1	Immunophenotyping marker. Commonly used for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma assessment
Kappa	Num	1	Immunohistochemistry marker. Commonly used in lymphoma assessments.	0, Negative <i>1, Positive</i>	Must match one of the values listed in the Controlled Vocabulary column	Optional

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
Ki-67	NUM	3	Ki-67 tumor cell proliferation assay. Often used for Mature B-Cell non-Hodgkin's lymphoma (large and small cell subtypes) and Anal Cancer Assessments. A	0 - 100 (%) For example, if cell proliferation is 44%, "44" would be entered.	Must match one of the values listed in the Controlled Vocabulary column	Required for Mature B-Cell non-Hodgkin's lymphoma (large and small cell subtypes) and Anal Cancer
Lambda	Num	1	Immunohistochemistry marker. Commonly used in lymphoma assessments.	0, Negative 1, Positive	Must match one of the values listed in the Controlled Vocabulary column	Required
LEF1	Num	1	Immunophenotyping marker. Commonly used for Mature B-Cell non-Hodgkin's lymphoma (small cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Mature B-Cell non-Hodgkin's lymphoma (small cell subtypes) assessments

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
MUM1	Num	1	Immunophenotyping marker. Commonly used for Mature B-Cell non-Hodgkin's lymphoma (large cell subtypes) assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Mature B-Cell non-Hodgkin's lymphoma (large cell subtypes) assessments
MYC	Num	1	Immunophenotyping marker.	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Optional
OCT-2	Num	1	Immunophenotyping marker. Commonly used for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma assessments	0, negative 1, positive 2, Not Available	Num	Required for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma assessments

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
PAX5	Num	1	Immunophenotyping marker. Commonly used for Classic Hodgkin's Lymphoma, Nodular lymphocyte predominant Hodgkin lymphoma, and Mature B-Cell non-Hodgkin's lymphoma (large cell subtypes)assessments	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Required for Classic Hodgkin's Lymphoma and Nodular lymphocyte predominant Hodgkin lymphoma assessments
TFH marker location	Char	20	Location of TFH marker for IHC assays.	n/a	n/a	Optional
PDRM/ BLIMP1	Num	1	Immunophenotyping marker.	0, negative 1, positive 2, Not Available	Must match one of the values listed in the Controlled Vocabulary column	Optional
Additional IHC	Char	200	Additional Immunohistochemistry assays performed.	n/a	Must list the target of the assay (e.g. CD5) and results(i.e. positive, negative. Example: "CD5: positive"	Optional

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
Histologic Grade	Char	2	Invasive cancer histologic grade.	G1, Well differentiated; G2, Moderately Differentiated; G3, Poorly Differentiated; GX, cannot be assessed NA	Must match one of the values listed in the Controlled Vocabulary column	Required for Anal and Cervical Cancer Assays
Histologic Grade-specify	Char	50	Invasive cancer histologic grade additional information for GX and not applicable Histologic Grade values.	n/a	n/a	Required if Histologic Grade is GX or Not applicable
Stromal Invasion Depth	Char	50	Depth of stromal invasion in millimeters, as applicable. If stromal invasion depth cannot be determined, provide rationale.	n/a	Must be 2 characters long (e.g. 02 for 2 mm)	Required if available for Gynecologic biopsies
Stromal Invasion Laterality	Char	50	Horizontal extent of stromal invasion in millimeters, as applicable. If stromal invasion laterality cannot be determined, provide rationale.	n/a	Must be 2 characters long (e.g. 02 for 2 mm)	Required if available for Gynecologic biopsies

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
Lymph-Vascular Invasion	Num	1	Identified lymph-vascular invasion, as applicable.	<i>0 = Negative, 1 = Positive</i>	Must match one of the values listed in the Controlled Vocabulary column	Required for Gynecological biopsies
Negative Invasive Carcinoma Margin St distance	Num	2	If all margins negative for invasive carcinoma, specify distance.	<i>Distance (mm)</i>	Must be 2 characters long (e.g. 02 for 2 mm)	One Invasive Carcinoma Margin required per donor for Gynecologic biopsies
Positive Invasive Carcinoma Margin Location	Char	20	If margin positive for invasive carcinoma, specify which margin.	<i>N/A</i>	n/a	One Invasive Carcinoma Margin required per donor for Gynecologic biopsies
Invasive Carcinoma Margin, Not applicable or Cannot be Determined	Char	50	If not applicable or margin status cannot be determined, explain.	<i>N/A</i>	n/a	One Invasive Carcinoma Margin required per donor for Gynecologic biopsies
Negative HSIL or AIS Margin Distance	Num	2	Margin status in invasive HSIL or AIS.	<i>Distance (mm)</i>	Must be 2 characters long (e.g. 02 for 2 mm)	One HSIL or AIS Margin required per donor for Gynecologic biopsies
Positive HSIL or AIS Margin Location	Char	20	If margin positive for invasive carcinoma, specify which margin. [include instructions about	<i>N/A</i>	n/a	One HSIL or AIS Margin required per donor for Gynecologic biopsies

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
HSIL or AIS Margin, Not applicable or Cannot be Determined	Char	50	If not applicable or margin status cannot be determined, explain.	<i>N/A</i>	n/a	One HSIL or AIS Margin required per donor for Gynecologic biopsies
Other Ancillary Studies	Char	200	Result Comment Text or Cancellation Reason, where applicable	--	--	Optional
Additional Pathologic Findings	Char	200	Result Comment Text or Cancellation Reason, where applicable	--	--	Optional
Pre-Biopsy Treatment Type	Char	50	Treatment performed pre-biopsy.	Chemotherapy, Radiation Therapy, Immunotherapy, Targeted Therapy, Other: [elaborate]	Must match one of the values listed in the Controlled Vocabulary column	Optional
Comments	Char	200	Result Comment Text or Cancellation Reason, where applicable	--	--	Optional

Specimen and data file linkages will be reported in the following format. Requirements for collection are noted in the last column titled “Collection Requirement”. Specimen linkages will be submitted in a [.csv] file to the ODMC.

- Filename will be [Protocol ID]_[assay]_[date]_specimen_linkage.csv
- The data must have the columns listed exactly as in the “Variable Name” column in the table below. Included values must be included as listed in the “Controlled Vocabulary” and “Validation Rules” columns in the table below.

Table 6: Variables to be Submitted in [Protocol ID]_[assay]_[date]_specimen_linkage.csv

Variable Name	Type	Length	Description	Controlled Vocabulary	Validation Rules	Collection Requirement
Specimen ID	Num	9	Specimen number, refers to the barcode number printed on each microscopy slide	--	Must be 9 characters long. If a Specimen is being entered as a reference (control) sample, the SN will be labeled as [specified value]	Required
File Name	Char	256	Name of the file	See details in each section	<i>See details in each section</i>	Required
File Type	Char	256	Type of the file	Raw Data, Interim-Processed Data, Final-Processed Data	Must match one of the values listed in the Controlled Vocabulary column	Required

5. EXPERIMENTAL INFORMATION AND KEY META DATA DESCRIPTORS

Experimental meta data descriptors will be collected and reported in the following format. Requirements for collection are noted in the last column titled “Collection Requirement”. Experimental meta data descriptors will be submitted in a [.csv] file to the ODMC.

- Filename will be [Protocol ID]_[assay]_[date]_exp_metadata.csv
- The data must have the columns listed exactly as in the “Variable Name” column. Included values must be included as listed in the “Controlled Vocabulary” and “Validation Rules” columns in the table below.

Table 7: Variables to be Submitted in [Protocol ID]_[assay]_[date]_exp_metadata.csv

Variable Name	Type	Length	Description	Controlled Vocabulary	Existing Standard/Link	Validation Rules	Collection Requirement
Protocol	Num	5	Protocol	--	--	Must match ODMC-specified value for the study	Required
Assay Type	Char	50	Assay type Name	Immunohistochemistry	--	Must match the value listed in the Controlled Vocabulary column	Required
Assay	Char	50	Assay Name	p16	--	Must match one of the values listed in the Controlled Vocabulary column	Required

Variable Name	Type	Length	Description	Controlled Vocabulary	Existing Standard/Link	Validation Rules	Collection Requirement
Lab	Char	25	Testing Laboratory Name	see appendix for Lab codelist. If your site is not in the codelist, provide lab name as text.	--	Must match one of the values listed in the Controlled Vocabulary column	Required
Lab Institution	Char	50	Testing Laboratory Institution	see appendix for Lab codelist. If your site is not in the codelist, provide lab name as text.	--	Must match one of the values listed in the Controlled Vocabulary column	Required
PI	Char	25	Testing Laboratory Principal Investigator	see appendix for PI codelist. If your site is not in the codelist, provide lab name as text.	--	Must match one of the values listed in the Controlled Vocabulary column	Required for university-based laboratories
Specimen Preparation	Char	50	Specimen Preparation	Frozen tissue, paraffin embedded tissue, Other: [elaborate]	--	Must match one of the values listed in the Controlled Vocabulary column	Required

Variable Name	Type	Length	Description	Controlled Vocabulary	Existing Standard/Link	Validation Rules	Collection Requirement
Modality	Char	50	Method of capturing image	Slide Microscopy, External-camera Photography, Secondary Capture, Other: [modality not listed]	https://portal.imaging.datacommons.cancer.gov/explore/		Required
Manufacturer	Char	50	Manufacturer of imaging equipment	Leica, Phillips, Keyence, Hamamatsu, Scpio, Other: [elaborate]	https://portal.imaging.datacommons.cancer.gov/explore/		Required
Manufacturer Model name	Char	50	Model of imaging equipment		https://portal.imaging.datacommons.cancer.gov/explore/		Required
Objective Lens Power	Num	2	Magnification factor	1, (if no magnification used), 5, 20, 40	https://portal.imaging.datacommons.cancer.gov/explore/	Must match one of the values listed in the Controlled Vocabulary column	Required
Samples per Pixel	Num	1	Number of samples per pixel in image	1, 3	https://portal.imaging.datacommons.cancer.gov/explore/	Must match one of the values listed in the Controlled Vocabulary column	Required
Illumination Type	Char	50	Method of illumination during imaging	Brightfield, Epifluorescence, None	https://portal.imaging.datacommons.cancer.gov/explore/	Must match one of the values listed in the Controlled Vocabulary column	Required

Variable Name	Type	Length	Description	Controlled Vocabulary	Existing Standard/Link	Validation Rules	Collection Requirement
Max total Pixel Matrix Columns	Num	7	Number of columns of pixels in an image		https://portal.imaging.datacommons.cancer.gov/explorer/		Required
Max total Pixel Matrix Rows	Num	7	Number of rows of pixels in an image	[potential ranges of rows, similar to imaging data commons]	https://portal.imaging.datacommons.cancer.gov/explorer/		Required
Minimum Pixel Spacing	Num	6	Distance between pixels. Calculated by the following formula: [Row pixel spacing]/[Column pixel spacing]=Pixel Spacing		https://dicom.innolitics.com/ciods/rt-dose/image-plane/00280030		Optional

6. QUALITY METRICS

Sample quality metrics will be collected and reported in the following format. Requirements for collection are noted in the last column titled “Collection Requirement”. Sample quality metrics will be submitted in a [.csv] file to the ODMC.

- Filename will be [Protocol ID]_[assay]_[date]_quality_metrics.csv
- The data must have the columns listed exactly as in the “Variable Name” column in the table below. Included values must be included as listed in the “Validation Rules” column in the table below.

Table 8: Variables to be Submitted in [Protocol ID]_[assay]_[date]_quality_metrics.csv

Variable Name	Type	Length	Description	Reference	Validation Rules	Collection Requirement
Protocol	Num	5	Protocol ID	--	Must match ODMC-specified value for the study	Required
Specimen ID	Num	9	Specimen number, refers to the GlobalTrace barcode number printed on each cryovial	N/A	Must be 9 characters long. If a Specimen is being entered as a reference (control) sample, the SN will be labeled as [specified value]	Required
Image Laterality	Num	1	Laterality of tumor site.	https://documents.cap.org/documents/Cervix.Bx_4.4.2.0.REL_CAPCP.pdf	1, Left superior (anterior) quadrant (12 to 3 o'clock) __2,__ Left inferior (posterior) quadrant (3 to 6 o'clock) __3, Right inferior (posterior) quadrant (6 to 9 o'clock) __4, Right superior (anterior) quadrant (9 to 12 o'clock) 5, Other 6, Cannot be Determined	Required
Image Laterality, Other or Cannot be Determined	Char	50	Elaborate if Image Laterality Variable value is 5 or 6			Required if Image Laterality variable value is 5 or 6

Variable Name	Type	Length	Description	Reference	Validation Rules	Collection Requirement
Burned In Annotation	Num	1	Binary enumerated value (1 for Yes, 0 for No). Indicates whether or not image contains sufficient burned in annotation to identify the patient and date the image was acquired.		Must be a binary value; 1 for Yes, 0 for No	

7. MANIFEST OF FILE COLLECTION TO BE SUBMITTED TO THE ODMC

The manifest of the file collection describes the collection of files that are expected along with the transferred files to ensure the integrity of the collection. The following files are expected to be listed in the manifest, at minimum: [Protocol ID]_[assay]_[date]_workflow.csv, [Protocol ID]_[assay]_[date]_exp_data.csv, [Protocol ID]_[assay]_[date]_image_linkage.csv, [Protocol ID]_[assay]_[date]_exp_metadata.csv, [Protocol ID]_[assay]_[date]_quality_metrics.csv, [Protocol ID]_[assay]_[date]_exp_manifest.csv.

Manifests will be submitted in a [.csv] to the ODMC. Requirements for collection are noted in the last column titled “Collection Requirement”.

- Filename will be [Protocol ID]_[assay]_[date]_exp_manifest.csv
- The data must have the columns listed exactly as in the “Variable Name” column. Included values must be included as listed in the “Controlled Vocabulary” and “Validation Rules” columns in the table below.

Table 9: Variables to be Submitted in [Protocol ID]_[assay]_[date]_exp_manifest.csv

Variable Name	Type	Length	Description	Controlled Vocabulary	Reference	Validation Rules	Collection Requirement
Protocol	Num	5	Protocol ID	--	N/A	Must match ODMC-specified value for the study	Required
File Name	Char	256	Name of the file	--	N/A	--	Required
File Type	Char	256	Type of the file	Raw, Interim-Processed, Final-Processed, Meta-Data, Quality-Metrics	N/A	Must match one of the values listed in the Controlled Vocabulary column	Required
sha256sum	Char	256	sha256 Checksum	--	See Data Preparation Instructions available at https://amcoperations.com/committee/amc-data-commons	--	Required
Timestamp	yyyy-mm-dd hh:mm		Data / Time of file creation	--	N/A	--	Required

Assay results will be reviewed for accuracy and compliance with specified data standards by the laboratory staff prior to transfer per the laboratory's procedures and SOPs.

8. ODMC DATA PROCESSING

No image or data processing will be performed by the ODMC for tissue imaging datasets.

9. APPROVAL

Data Standardization Plan Approval	
By signing this data transfer plan, I approve of the included specifications.	
Laboratory Resource Committee Chair:	
Signature: <u>Jeffrey Bethony</u>	Electronically signed by: Jeffrey Bethony Reason: I am Approving this document Date: Oct 14, 2026 16:31:06 ADT
<u>Jeffrey Bethony, PhD</u>	<u>10/14/2026</u> Date

10. APPENDIX

Common Anal Cancer Histologies	
1	Histologic type cannot be assessed.
2	Squamous cell carcinoma
3	Verrucous squamous cell carcinoma
4	Adenocarcinoma
5	Large cell neuroendocrine carcinoma
6	Small cell neuroendocrine carcinoma 3 CAP Approved Anus.Bx.TDE4.2.0.0.RELCAPCP
7	Mixed neuroendocrine-non-neuroendocrine neoplasm
8	Other histologic type not listed (specify):
9	Carcinoma, type cannot be determined: +Histologic Type Comment:
Common Cervical Cancer Histologies	
10	Benign and/or squamous metaplasia, includes reactive changes.
11	Atypia: Squamous, glandular
12	LSIL, Condyloma/HPV
13	LSIL, IN1
14	HSIL, IN2
15	HSIL, IN3
16	HSIL with features suspicious for invasion
17	Adenocarcinoma in situ
18	SISCCA
19	At least SISCCA
20	Squamous cell carcinoma
21	Adenocarcinoma
Common Kaposi Sarcoma Histologies	
22	Patch-stage KS
23	Plaque-stage KS
24	Nodular KS
25	Anaplastic KS
26	Lymphangioma-like KS
27	Other histologic type not listed (specify):
28	Histologic type cannot be assessed
Common Lymphoma Histologies	
<u>Hodgkin Lymphoma</u>	
29	Hodgkin lymphoma, histologic subtype cannot be determined
30	Classic Hodgkin lymphoma, histologic subtype cannot be determined
31	Nodular lymphocyte predominant Hodgkin lymphoma
32	Nodular sclerosis classic Hodgkin lymphoma
33	Mixed cellularity classic Hodgkin lymphoma
34	Lymphocyte-rich classic Hodgkin lymphoma
35	Lymphocyte-depleted classic Hodgkin lymphoma

Precursor Lymphoid Neoplasms	
36	B lymphoblastic leukemia/lymphoma, not otherwise specified (NOS)
37	B lymphoblastic leukemia/lymphoma with t(9;22)(q34;q11.2); BCR::ABL1
38	B lymphoblastic leukemia/lymphoma with t(v;11q23); KMT2A rearranged
39	B lymphoblastic leukemia/lymphoma with t(12;21)(p13;q22); ETV6::RUNX1
40	B lymphoblastic leukemia/lymphoma with hyperdiploidy
41	B lymphoblastic leukemia/lymphoma with hypodiploidy (hypodiploid acute lymphoblastic leukemia/lymphoma [ALL])
42	B lymphoblastic leukemia/lymphoma with t(5;14)(q31;q32); IL3::IGH
43	B lymphoblastic leukemia/lymphoma with t(1;19)(q23;p13.3); (TCF3-PBX1)
44	T lymphoblastic leukemia/lymphoma
Mature B-Cell Neoplasms	
45	B-cell lymphoma, subtype cannot be determined (Note: not a category within the WHO classification)
46	Chronic lymphocytic leukemia/small lymphocytic lymphoma
47	Splenic marginal zone lymphoma
48	Hairy cell leukemia
49	Splenic B-cell lymphoma/leukemia, unclassifiable with prominent nucleoli
50	Splenic diffuse red pulp small B-cell lymphoma
51	Hairy cell leukemia-variant
52	Lymphoplasmacytic lymphoma
53	Cold agglutinin disease
54	IgM monoclonal gammopathy of undetermined significance
55	Non- IgM monoclonal gammopathy of undetermined significance
56	Monoclonal gammopathy of renal significance
57	Gamma heavy chain disease
58	Mu heavy chain disease
59	Alpha heavy chain disease
60	Plasma cell myeloma / multiple myeloma
61	Solitary plasmacytoma of bone
62	Plasmacytoma
63	Plasma cell neoplasms with associated paraneoplastic syndrome
64	Extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue (MALT lymphoma)
65	Nodal marginal zone lymphoma
66	Pediatric nodal marginal zone lymphoma
67	In situ follicular B-cell neoplasm
68	Follicular lymphoma
69	Pediatric follicular lymphoma
70	Duodenal-type follicular lymphoma

71	Primary cutaneous follicle center lymphoma
72	In situ mantle cell neoplasm
73	Mantle cell lymphoma
74	Leukemic non-nodal mantle cell lymphoma
75	Transformations of indolent B-cell lymphomas
76	Diffuse large B-cell lymphoma (DLBCL), NOS
77	T cell/histiocyte-rich large B-cell lymphoma
78	Primary cutaneous DLBCL, leg type
79	Epstein-Barr virus (EBV)-positive DLBCL
80	DLBCL associated with chronic inflammation
81	Fibrin associated large B cell lymphoma
82	Fluid overload-associated large B cell lymphoma
83	Large B cell lymphoma with IRF4 rearrangement
84	Primary large B cell lymphoma of immune privileged sites
85	Lymphomatoid granulomatosis
86	High grade B cell lymphoma with 11q aberrations
87	Primary mediastinal large B-cell lymphoma
88	Intravascular large B-cell lymphoma
89	Anaplastic lymphoma kinase (ALK)-positive large B-cell lymphoma
90	Plasmablastic lymphoma
91	Diffuse large B-cell lymphoma / high grade B-cell lymphoma with MYC and BCL2 rearrangements
92	High grade lymphoma, NOS
93	Burkitt lymphoma
94	KSHV/HHV8-positive diffuse large B-cell lymphoma
95	Primary effusion lymphoma
96	Mature B-Cell Neoplasms ICC classification (optional):
Mature T- and NK-Cell Neoplasms	
97	T-cell lymphoma, subtype cannot be determined
98	T-cell prolymphocytic leukemia
99	T-cell large granular lymphocytic leukemia
100	NK large granular lymphocytic leukemia cells
101	Aggressive NK-cell leukemia
102	Systemic EBV-positive T-cell lymphoma of childhood
103	Hydroa vacciniform lymphoproliferative disorder
104	Severe mosquito bite allergy
105	Systemic chronic active EBV disease
106	Adult T-cell leukemia/lymphoma
107	Extranodal NK/T-cell lymphoma, nasal type
108	EBV positive nodal T- and NK-cell lymphoma
109	Enteropathy-associated T-cell lymphoma
110	Hepatosplenic T-cell lymphoma
111	Subcutaneous panniculitis-like T-cell lymphoma
112	Primary cutaneous CD30+ large cell lymphoma: Primary cutaneous anaplastic large cell lymphoma
113	Primary cutaneous CD30+ large cell lymphoma: Lymphomatoid papulosis

114	Primary cutaneous gamma-delta T-cell lymphoma
115	Primary cutaneous CD8-positive aggressive epidermotropic cytotoxic T-cell lymphoma
116	Primary cutaneous acral CD8-positive T-cell lymphoproliferative disorder
117	Primary cutaneous CD4-positive small/medium T-cell lymphoproliferative disorder
118	Indolent T cell lymphoma of the gastrointestinal tract
119	Indolent NK cell lymphoproliferative disorder of the gastrointestinal tract
120	Monomorphic epitheliotropic intestinal T-cell lymphoma
121	Intestinal T cell lymphoma, NOS
122	Peripheral T-cell lymphoma, NOS
123	Nodal TFH lymphoma, Angioimmunoblastic – type
124	Nodal TFH lymphoma, follicular-type
125	Nodal TFH lymphoma, NOS
126	ALK-positive anaplastic large cell lymphoma
127	ALK-negative anaplastic large cell lymphoma,
128	Breast implant associated anaplastic large cell lymphoma
Histiocytic and Dendritic Cell Neoplasms	
129	Histiocytic sarcoma
130	Langerhans cell histiocytosis
131	Langerhans cell sarcoma
132	Interdigitating dendritic cell sarcoma
133	Follicular dendritic cell sarcoma
134	Fibroblastic reticular cell tumor
135	Indeterminate dendritic cell tumor
136	Disseminated juvenile xanthogranuloma
Lymphoid proliferations and lymphomas associated with immune deficiency and dysregulation	
137	Hyperplasias arising in immune deficiency / dysregulation
138	Polymorphic lymphoproliferative disorders arising in immune deficiency / dysregulation
139	EBV positive mucocutaneous ulcer
140	Lymphomas arising in immune deficiency / dysregulation
141	Inborn error of immunity-associated lymphoid proliferations and lymphomas
142	Other (specify):

Lab and Lab Institution

1	University of California, San Diego (UCSD)
2	Miami Cancer Institute at Baptist Health, Inc
3	University of Illinois
4	Boston Medical Center
5	Johns Hopkins University (JHU)
6	Washington University
7	NYP/Weill-Cornell Medical College
8	Mt Sinai

9	Montefiore Medical Center-Einstein Campus
10	Ohio State University
11	Pennsylvania Hospital
12	MD Anderson (Thomas Street at Quentin Mease Health Center)
13	Virginia Mason Medical Center (VMMC)

PI Name

1	Erin Reid
2	Marco Ruiz
3	Paul Rubinstein
4	Elizabeth Stier
5	Richard Ambinder
6	Lee Ratner
7	Grant Ellsworth
8	Joseph Sparano
9	Rafi Kabarriti
10	Robert Baiocchi
11	David Henry
12	Elizabeth Chiao
13	David Aboulafia