TCIA-NLST-LSS Pathology Data Dictionary

11/6/2013

Table of Contents

1. Background information

This document gives a detailed description of Lung Screening Study (LSS) pathology database tables that are related to pathology images from a subset of lung cancer patients who participated in the National Lung Screening Trial (NLST). The following paragraphs contain relevant background information.

Pathology Collection Process

The NLST pathology collection consists of biospecimens, images, and data for research. Near the conclusion of NLST activities, all available donor blocks of preserved lung tissue from NLST lung cancer patients were requested from pathology labs (Figure 1.1). A thin section was cut from each donor tissue block, stained with hematoxylin and eosin (H & E), and imaged using a digital microscope (Aperio ScanScope) (Figure 1.2). The pathology images were obtained as an intermediate step in TMA construction. A pathologist marked the outlines of regions of interest (ROI), where the outline's color indicated the tissue type / histology (black = invasive carcinoma, purple = normal lung tissue, etc.) (Figure 1.3). He also recorded more detailed histology information on data collection forms at both tumor-wide and ROI-specific levels (including the 2004 WHO classification of lung cancers). The detailed information from the data collection forms was extracted and placed into the NLST database tables named with the prefix LSSPath. The histology data from the pathologist's review sometimes disagrees with the histology recorded at lung cancer diagnosis. Tissue cores were sampled from these donor blocks and placed into tissue microarrays (TMAs) and into Eppendorf tubes for later analysis by the research community (Figure 1.4). The request process for the tissue specimens is under development as of October 2013.

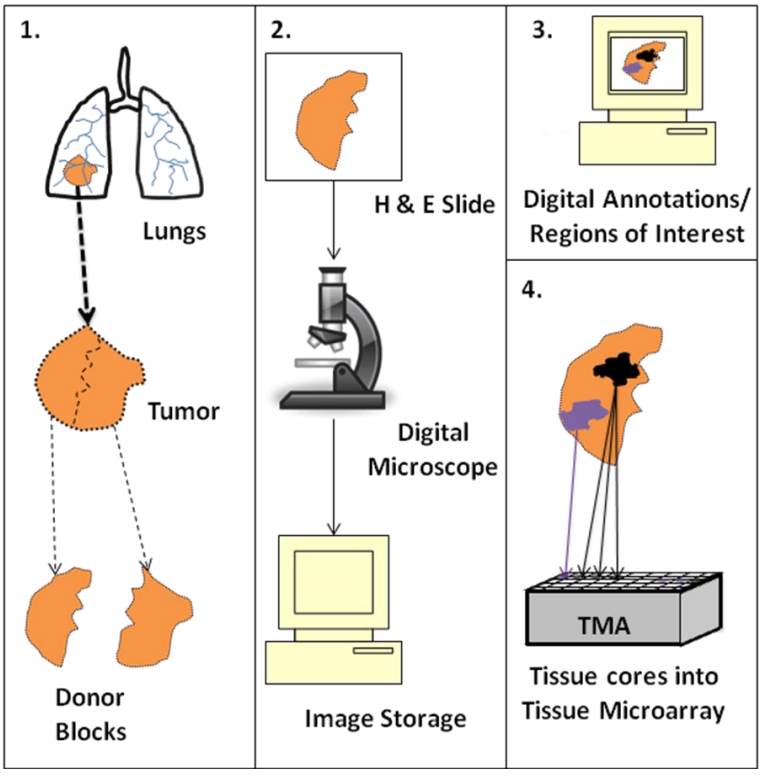

Figure 1: NLST Pathology Collection Process

1. Lung Tumors are removed from lungs (as part of diagnosis and treatment) and divided into Donor Blocks.
2. A thin H & E slide from each donor block is scanned by a digital microscope, and image data is stored on a computer.
3. A pathologist adds digital annotations of regions of interest (ROI) on the image. Each annotation color represents a particular histology.
4. Small tissue cores (0.6 mm in diameter) are punched from regions of interest on donor blocks and placed in a tissue microarray (TMA), a wax block with a gridded layout that holds ~500 cores.

Access to Pathology Images and Data

The Cancer Imaging Archive (TCIA) hosts the LSS subset of the pathology images and data. The NLST was conducted by two separate networks of screening centers: LSS (10 centers, 34,612 participants) and ACRIN (American College of Radiology Imaging Network; 23 centers, 18,840 participants). The LSS pathology collection obtained tissue, images, and data from 463 lung cancer patients (out of 1,284 total lung

1. Background information

cancer patients in LSS). From those 463 patients, 1,254 donor blocks were obtained; almost all of these have images available for viewing. 2,522 regions of interest were annotated, and 7,596 tissue cores ("TMACores") were sampled and placed into the tissue microarray blocks. Note that nine of the 463 participants had two primary lung tumors; only one tumor's data and images are accessible through TCIA for these nine participants. ACRIN pathology images and data are not available through TCIA.

Access to the LSS pathology images and data for the 463 patients is provided through the TCIA-NLST Database Query Tool (QT) and its associated pathology image viewer (caMicroscope). In order to view the images, one must run a query that includes the variable PATHOLOGY_IMAGE from the LSSPathDonorBlock table and then click on a "ViewPathologyImage" button in the Results area. The Query Tool User's Manual provides more information on how to use the Query Tool and about the LSS Pathology images/data. It can be downloaded by selecting it from the Query Tool's Help menu.

Permission is required for access to NLST data and images. To obtain permission, submit a request through the National Cancer Institute's (NCI) Cancer Data Access System (CDAS) at https://biometry.nci.nih.gov/cdas/nlst/requests/instructions/. The Query Tool is only accessible to users who have obtained permission.

The QT enables viewing but not downloading of the pathology images. Copies of the images (in Aperio SVS format) may be obtained by submitting a request through CDAS for the pathology images. Once NCI has approved, you will be contacted about shipping an external hard drive for the transfer of the images to the TCIA group at Washington University in St. Louis. If you have questions about the process, send an e-mail to cdas@imsweb.com.

More Information

- Query Tool User's Manual. Accessible from the Query Tool's Help menu.
- LSS Manual of Operations and Procedures (MOOP), Chapter 12 (p. 769). (https://biometry.nci.nih.gov/shared/cdas-static/forms/nlst/lss/LSS%20MOOP.pdf).
- "Tissue sampling and digital histologic imaging for the National Lung Screening Trial." Abstract from 2010 ASCO Annual Meeting (http://meeting.ascopubs.org/cgi/content/abstract/28/15_suppl/e12032).

2. LSSPathTumor Table Variables

The LSSPathTumor database table contains one row per LSS lung cancer patient with pathology data. All such participants have lung cancer. The variables describe tumor-wide characteristics.

Section	Variable	Label	Description	Format Text
2.1: Identifiers	**pid**	Participant Identifier		Numeric (######)
2.1: Identifiers	**study_yr**	Study year associated with first confirmed lung cancer		Numeric
2.1: Identifiers	**tumor_ID**	Tumor identifier	ID of original tumor	char, 7 (AA#####)
2.1: Identifiers	**lc_order**	Order of this tumor among all tumors diagnosed	For participants with multiple primary lung cancers, this variable indicates which cancer is being described.	1="1st lung cancer"
2.2: Tumor Information	**case_topography_descrip**	Topography Description	Character Field for Tumor Location(s) (from original path report)	Char, 64
2.2: Tumor Information	**case_side**	Side of body where tumor is located	Side of body on which tumor is located. From original path report.	1="Right" 2="Left"
2.2: Tumor Information	**case_any_tmacore**	Any tissue from this tumor in the tissue microarrays?	Indicator for TMA Cores	0="No" 1="Yes"
2.2: Tumor Information	**case_tmacore_cnt**	Number of tissue cores from this tumor in the tissue microarrays	TMA Core Count	Numeric
2.2: Tumor Information	**case_who_class**	WHO classification of tumor	Raw WHO Classification (2004) in Character Format. Usually 6 characters, ####/# Occasionally contains 2 separate classifications for the same tumor.	char, 14
2.2: Tumor Information	**case_highest_inv_grade**	Highest grade observed in tumor		0="Blank" 1="Well Differentiated" 2="Moderately" 3="Poorly" 4="Undifferentiated" 99="Missing"
2.2: Tumor Information	**case_likely_metastases**	Case likely to have metastases?		0="None" 1="Probable" 2="Unlikely" 9="Cannot Determine"
2.2: Tumor Information	**case_longest_inv_diam**	Longest diameter of invasive component of tumor	From central pathology review. Sometimes multiple entries are entered, comma-delimited. Two entries implies two WHO Classes given, but not vice-versa.	char, 8

3. LSSPathDonorBlock Table Variables

The LSSPathDonorBlock database table contains one row per donor block (piece of lung tumor). Each pathology image corresponds to a donor block, though a few donor blocks do not have images available. In order to view the pathology images, one must include the variable PATHOLOGY_IMAGE in a query.

Section	Variable	Label	Description	Format Text
3.1: Identifiers	**pid**	Participant Identifier		Numeric (######)
3.1: Identifiers	**pathology_image**	Pathology Image ID (Same as tissue_block_id)	This identifier is used to create a ViewPathologyImage button to allow viewing of the pathology images with an image viewer. Format: AA##### ####	char, 15 null ="Missing"
3.1: Identifiers	**study_yr**	Study year associated with first confirmed lung cancer		Numeric
3.1: Identifiers	**tissue_block_id**	Identifier of donor tissue block	ID of original Tissue Block	char, 12 (AA##### ####)
3.1: Identifiers	**tumor_ID**	Tumor identifier	ID of original tumor	char, 7 (AA#####)
3.1: Identifiers	**co_slide_id**	Colorado Slide Identifier for the Pathology Images	Identifier for images of pathology slides. As assigned by annotating path lab and transcribed by TMA construction path lab.	char, 11 (NLSI########)
3.2: Block Region Information	**block_annotated**	Were ROIs annotated on the block?	Were regions of interest annotated for this block? By protocol, digital annotations were made on an image of an H&E slide of each tissue block. Each annotation color was used for a specific histology.	0="No" 1="Yes"
3.3: Block TMA Cores	**block_has_tmacore**	Any tissue from this donor block in the TMAs?	Indicator for whether or not Tissue block has TMA cores or not.	0="No" 1="Yes"
3.3: Block TMA Cores	**block_tmacore_orig_cnt**	Count of cores in the TMAs from this block	Number of TMA Cores before any were used	Numeric

4. LSSPathRegOfInterest Table Variables

The LSSPathRegOfInterest database table contains one row per region of interest (ROI) noted on donor block images by the reviewing pathologist. The variables describe the tissue characteristics within each region of interest.

Section	Variable	Label	Description	Format Text
4.1: Identifiers	**pid**	Participant Identifier		Numeric (######)
4.1: Identifiers	**study_yr**	Study year associated with first confirmed lung cancer		Numeric
4.1: Identifiers	**tissue_block_id**	Identifier of donor tissue block	ID of original Tissue Block	char, 12 (AA##### ####)
4.1: Identifiers	**roi_number**	Number of ROI on donor block	Each Region of Interest within a tissue block is assigned a number. This is used for linking TMA cores to the roi.	Numeric
4.1: Identifiers	**tumor_ID**	Tumor identifier	ID of original tumor	char, 7 (AA#####)
4.2: Region Summary	**roi_label_color**	Annotation color (=> histology)	Color used by pathologist to mark the regions of interest of each tissue block. Each color corresponds to a particular histology (see roi_histology_cat).	"AQUA"="Aqua (Metastasis)" "BLACK"="Black (Invasive Carcinoma)" "BLUE"="Blue (Carcinoma In Situ)" "BROWN"="Brown (Pre-Malignant)" "GOLD"="Gold (Other)" "GREEN"="Green (Distal Bronchiole)" "ORANGE"="Orange (Normal Lymph Node)" "OTHER"="Other" "PURPLE"="Purple (Normal Lung)" "RED"="Red (Invasive Type II)" "YELLOW"="Yellow (Proximal Bronchus)"
4.2: Region Summary	**roi_histology_cat**	Histology category (=> annotation color)		1="Metastasis" 2="Invasive Carcinoma" 3="Carcinoma In Situ" 4="Pre-Malignant" 5="Distal Bronchiole" 6="Proximal Bronchus" 7="Normal Lymph Node" 8="Normal Lung" 9="Other"
4.2: Region Summary	**roi_histology_subtype**	Histology: Text description	For cancers, it includes a text description of the WHO classification. For non-tumor lung tissue, see roi_histology_cat for further distinctions (i.e. distal bronchiole vs. proximal bronchus vs. other normal lung tissue).	Char, 65
4.2: Region Summary	**roi_features_comments**	Misc. comments about ROI	Miscellaneous comments	Char, 107 null ="Missing"
4.3: Region Cancer Variables	**roi_who_class**	WHO classification of tumor tissue in ROI	Based on 2004 WHO classification of lung tumors. Includes a 4-digit morphology code and a 1-digit behavior code (#####/#).	Char, 6 null ="Missing"

4. LSSPathRegOfInterest Table Variables

Section	Variable	Label	Description	Format Text
4.3: Region Cancer Variables	**roi_who_class_mor**	ICD-O-3 morphology code from WHO classification	Whole number with 4 digits.	Numeric null ="Missing"
4.3: Region Cancer Variables	**roi_who_class_beh**	ICD-O-3 behavior code from WHO classification		null ="Missing" 2="In-Situ" 3="Invasive"
4.3: Region Cancer Variables	**roi_highest_grade**	Highest grade observed in ROI		0="Blank" 1="Well Differentiated" 2="Moderately" 3="Poorly" 4="Undifferentiated"
4.3: Region Cancer Variables	**roi_metastases**	Metastases: likely from tumor tissue in this ROI?		Missing None Not Applicable 0="No" 1="Probable" 2="Unlikely" 99="Can't Determine"
4.3: Region Cancer Variables	**roi_tumor_cell_percent**	Percent tumor cells within ROI		Numeric null ="Missing"
4.3: Region Cancer Variables	**roi_inv_tumor_dim**	Size of Invasive Component of Tumor (mm)	In millimeters	Numeric null ="Missing"
4.3: Region Cancer Variables	**roi_lymph_inv**	Lymphatic invasion in ROI?		Null ="Missing" Not Applicable 0="No" 1="Yes"
4.3: Region Cancer Variables	**roi_vessel_inv**	Blood vessel invasion in ROI?		Null ="Missing" Not Applicable 0="No" 1="Yes"
4.3: Region Cancer Variables	**roi_inflamm_cell_percent**	Percent inflammatory cells within ROI		Numeric null ="Missing"
4.4: Region Non-Tumor Histologies	**roi_nthist_normal_lung**	Has Normal lung tissue in ROI		0="No" 1="Yes"
4.4: Region Non-Tumor Histologies	**roi_nthist_granuloma**	Has Granuloma in ROI		0="No" 1="Yes"
4.4: Region Non-Tumor Histologies	**roi_nthist_hemorrhage**	Has Hemorrhage in ROI		0="No" 1="Yes"
4.4: Region Non-Tumor Histologies	**roi_nthist_pneumonia**	Has Pneumonia in ROI		0="No" 1="Yes"
4.4: Region Non-Tumor Histologies	**roi_nthist_necrosis**	Has Necrosis in ROI		0="No" 1="Yes"
4.4: Region Non-Tumor Histologies	**roi_nthist_infarction**	Has Infarction in ROI		0="No" 1="Yes"
4.4: Region Non-Tumor Histologies	**roi_nthist_emphysema**	Has Emphysema in ROI		0="No" 1="Yes"
4.4: Region Non-Tumor Histologies	**roi_nthist_fibrosis**	Has Fibrosis in ROI		0="No" 1="Yes"

4. LSSPathRegOfInterest Table Variables

Section	Variable	Label	Description	Format Text
4.4: Region Non-Tumor Histologies	**roi_nthist_pre_neoplastic_tiss**	Has Pre Neoplastic tissue in ROI		0="No" 1="Yes"
4.4: Region Non-Tumor Histologies	**roi_nthist_chronic_bronchitis**	Has Chronic bronchitis in ROI		0="No" 1="Yes"
4.4: Region Non-Tumor Histologies	**roi_nthist_other_specify**	Has Other (SPECIFY) non-tumor histology in ROI		0="No" 1="Yes"
4.4: Region Non-Tumor Histologies	**roi_nthist_other_specify_text**	Has Other Condition - Verbatim text		Char, 57 null ="Missing"
4.5: Region Pre-Malignant Histologies	**roi_nontumor_premalig**	Pre-malignant tissue type		1="Squamous Carcinoma in-Situ" 2="Squamous Dysplasia, MILD" 3="Squamous Dysplasia, Moderate" 4="Squamous Dysplasia, SEVERE" 5="Atypical Adenomatous Hyperplasia (AAH)" 7="Reserve Cell Hyperplasia" 8="Carcinoid Tumorlet" 88="Other" 99="N/A"

5. LSSPathTMACore Table Variables

The LSSPathTMACore database table contains one row per tissue core used in the tissue microarrays (TMAs). These "TMA cores" were sampled from regions of interest on donor blocks. Some variables are identifiers to link back to the source tumor, donor block, and region of interest. Other variables indicate which of 18 TMA blocks this tissue core was placed in and the core's position within that TMA block.

Thin sections from the tissue microarrays will be made available to researchers in the future. This database table is included in the Query Tool for future use by researchers who have obtained those sections. No images of TMAs are accessible through the Query Tool.

Section	Variable	Label	Description	Format Text
5.1: Identifiers	**pid**	Participant Identifier		Numeric (######)
5.1: Identifiers	**study_yr**	Study year associated with first confirmed lung cancer		Numeric
5.1: Identifiers	**tissue_block_id**	Identifier of donor tissue block	ID of original Tissue Block	char, 12 (AA##### ####)
5.1: Identifiers	**roi_number**	Number of ROI on donor block	Each Region of Interest within a tissue block is assigned a number. This is used for linking TMA cores to the roi.	Numeric
5.1: Identifiers	**tmacore_id**	TMA Core Identifier	Format: AA##### ####	char, 14
5.1: Identifiers	**tma_block_id**	Identifier of TMA block in which core is arrayed	Format: AA##### ####	char, 14
5.1: Identifiers	**tumor_ID**	Tumor identifier	ID of original tumor	char, 7
5.1: Identifiers	**tma_block**	TMA Block Name	Label for each tissue microarray block. Note that for each letter, two TMA blocks (with identical layouts) exist. These duplicate blocks are the Alpha and Beta blocks, distinguishable by using the alpha_beta variable.	"A NEO"="A NEO" "A"="A" "B"="B" "C"="C" "D"="D" "E"="E" "F"="F" "G"="G" "H"="H"
5.1: Identifiers	**alpha_beta**	In Alpha or Beta TMA block?	Is the TMA tissue core from the Alpha or Beta block? For each TMA layout (as distinguished by the TMA_BLOCK variable), there were two duplicate blocks constructed, the Alpha and the Beta blocks.	1="Alpha" 2="Beta"
5.2: TMAcore Sample Info.	**quadrant**	Micro Array quadrant		1="Upper Left" 2="Lower Left" 3="Upper Right" 4="Lower Right"
5.2: TMAcore Sample Info.	**micro_array_col**	Micro Array column	= 2 to 15 for quadrants 1 & 2, 22 to 35 for quadrants 3 & 4. Columns 1 & 21 are orientation columns containing non-NLST tissue.	numeric
5.2: TMAcore Sample Info.	**micro_array_row**	Micro Array row	= 1 to 9 for quadrants 1 & 3, 21 to 29 for quadrants 2 & 4	numeric