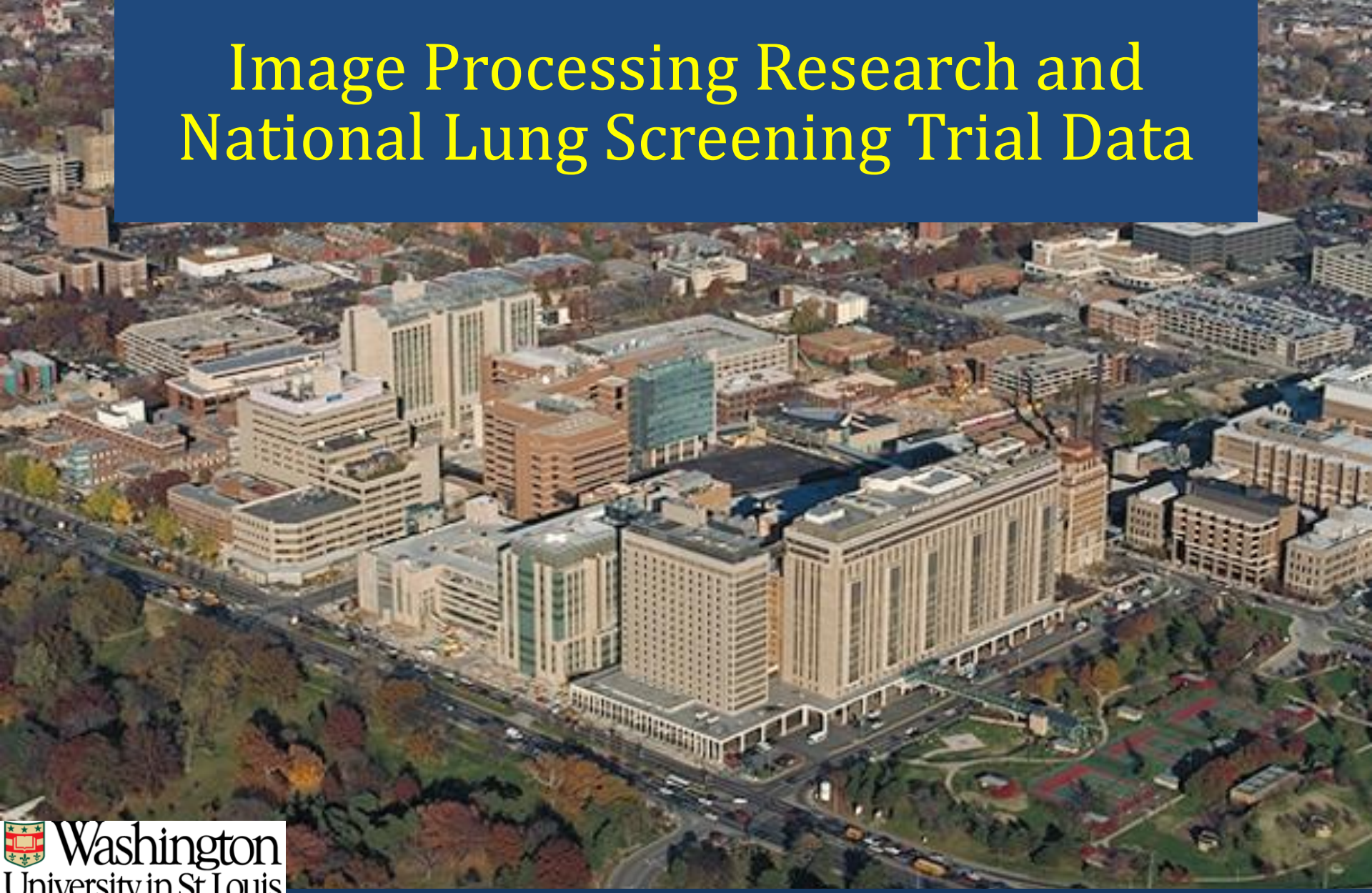
Image Processing Research and National Lung Screening Trial Data



Washington University in St. Louis
SCHOOL OF MEDICINE
MIR Mallinckrodt Institute of Radiology

Lung Cancer Workshop X

David S. Gierada, M.D.

NLST Data and Imaging Research

- Data and obtaining access
- Current and potential uses
- Limitations and needs

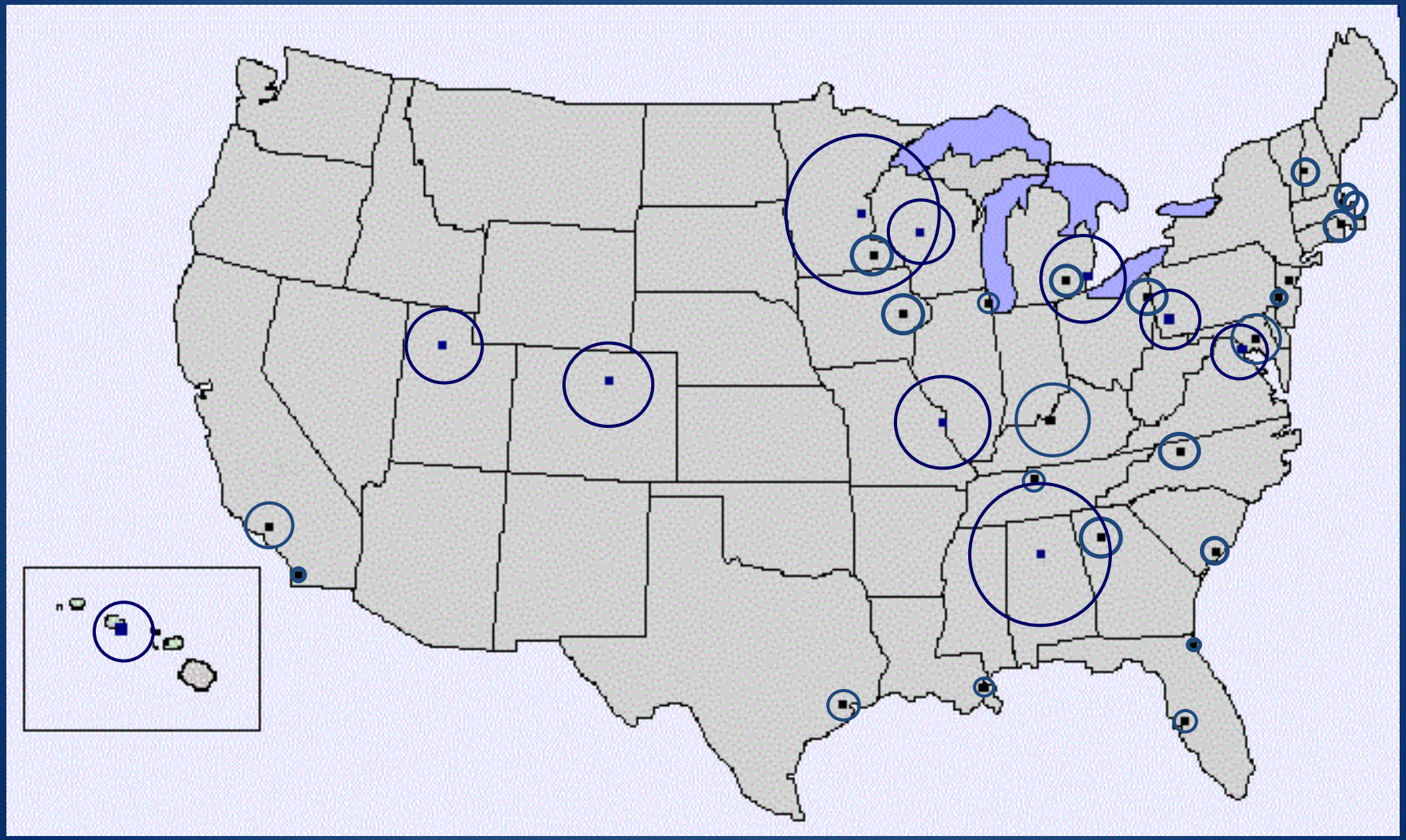
National Lung Screening Trial 2002-09

- 53,454 randomized to CT or CXR screening
  - 55-74 years at enrollment
  - ≥ 30 pack years
- 3 annual screening rounds (T0, T1, T2)
- Positive CT screen: 4 mm nodule
- Median follow-up 6.5 years

National Lung Screening Trial 2002-09

- CT arm results
  - 20% fewer lung cancer deaths (P=0.004)
  - 6.7 % fewer total deaths (P=0.02)
- 24% false positive CT rate
- 1060 lung cancers (4% of CT arm)
- 50% Stage I
- 356 lung cancer deaths

33 Screening Centers
23 ACRIN (~1/3), 10 LSS (~2/3)



NLST Data
Public Availability

- Released for investigational use November 2012
- All screening CTs (no CXRs)
- Participant-level demographics and medical history, annual screening results, follow-up data
- Biospecimens

Obtaining NLST Data

- Apply through the NCI Cancer Data Archive System (CDAS) website (https://biometry.nci.nih.gov/cdas/)
- Submit brief description of project for NCI approval
- List users
- Complete Data Transfer Agreement (DTA) to accept terms of use
  - Requires signature of sponsoring institution official
  - Approval is project-specific

CDAS Cancer Data Access System

NLST Data
PLCO Data
Questions/Comments

Welcome to the Cancer Data Access System

The Cancer Data Access System (CDAS) is a submission and tracking system for the use of data from the National Lung Screening Trial (NLST) and the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial.

Interested investigators can register with CDAS and request access to data from either screening trial. All requests are reviewed by NCI trial leadership. Upon approval, investigators will be granted access to the requested data for a limited period.

CDAS provides extensive documentation for each trial including a summary of the trial, a description of the data collected, and a searchable list of research projects and publications.

NLST

The National Lung Screening Trial compared two ways of detecting lung cancer: low-dose helical computed tomography (CT) and standard chest X-ray. Both chest X-rays and low-dose helical CT scans have been used to find lung cancer early, but the effects of these screening techniques on lung cancer mortality rates had not been determined. NLST enrolled approximately 54,000 current or former heavy smokers from 33 sites and coordinating centers across the United States.

Information about NLST and available data and documentation

NLST
National Lung Screening Trial
NATIONAL CANCER INSTITUTE

PLCO

The Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial is a large scale, randomized study to determine whether certain screening tests will reduce the number of deaths from these cancers. PLCO is being conducted at ten sites, geographically and demographically disparate, around the U.S. The Trial enrolled approximately 155,000 male and female participants between the ages of 55 and 74 from 1993 to 2001.

Information about PLCO and available data and documentation

NATIONAL CANCER INSTITUTE
PLCO
Prostate, Lung, Colorectal, & Ovarian
CANCER SCREENING TRIAL

CDAS Cancer Data Access System

NLST

The National Lung Screening Trial (NLST) was a randomized controlled trial to determine whether screening for lung cancer with low-dose helical computed tomography (CT) reduces mortality from lung cancer in high-risk individuals relative to screening with chest radiography. Approximately 54,000 participants were enrolled between August 2002 and April 2004.

NLST was administered by two separate groups: the Lung Screening Study group (LSS), and the American College of Radiology Imaging Network (ACRIN). The protocols and data collection practices of these groups were tightly harmonized. A joint ACRIN-LSS database was created and is available for download from this website. It contains detailed trial data on events occurring through December 31, 2009. No further data collection is ongoing.

This website is organized into several sections as follows.

- Learn about NLST
  This describes the trial, explaining how and what data were collected. The main findings of the trial and counts of cancers can be found here.

- Datasets
  This page has detailed documentation of the NLST data available on this website. Each dataset has a data dictionary and SAS format code. The participant dataset additionally has a user guide.

- Publications/Projects
  Search through published and ongoing studies on NLST data.

- Submit a Request
  Learn how to access the available NLST data including the screening, cancer, mortality and questionnaire data. You must submit a request to access the data. All requests are reviewed by NCI. If your request is approved, you will be required to complete a Data Transfer Agreement before you will be granted access to the data.

- Questions/Comments
  If you have any questions about the data or access to it, please contact us.

Other NLST Resources

Data Transfer Agreement

Terms of Access

In consideration of NCI providing the DATA to the RECIPIENT, the RECIPIENT hereby agrees to the following terms and conditions:

Will only use data as described in Research Plan

The DATA will be used only for purposes described in the Research Plan. Such use shall be permitted only during the thirty-six (36) months following the execution of this agreement. No other use of the DATA is permitted unless specifically allowed by the written consent of the NCI and accompanied by a new project application.

Will not share data with unauthorized entities within and outside of my institution

RECIPIENT's access to the DATA shall be limited to the minimum number of individuals necessary to achieve the purpose stated in the Research Plan. RECIPIENT will advise such individuals of the terms and conditions

Will not identify or contact patients or physicians

The RECIPIENT will ensure that its investigator will not, and shall assure that his/her collaborators will not, attempt to learn the identity of any human sources, their physicians, or collection sites for the DATA, and shall ensure that in the event that any such identities are discovered then the RECIPIENT will ensure that its investigator will not, and shall assure that his/her collaborators will not, contact these individuals or institutions. The RECIPIENT hereby agrees it shall require its investigator(s) to notify collaborators of these policies.

Will not claim ownership of the data or misrepresent access as an endorsement

The RECIPIENT will not claim, infer, or imply endorsement of its activities under the Research Plan by the U.S. Government, DHHS, NIH, NCI, or NCI employees. The DATA are the property of the NCI and are made available as a service to the research community.

Data delivered pursuant to this Agreement is understood to be experimental in nature. THE NCI MAKES NO REPRESENTATIONS AND EXTENDS NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED. THERE ARE NO EXPRESS OR IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE, OR THAT THE USE OF THE DATA WILL NOT INFRINGE ANY PATENT, COPYRIGHT, TRADEMARK, OR OTHER PROPRIETARY RIGHTS.

Will allow my full Research Plan to be displayed on this web site

The NCI is hereby granted the right to publicly display information describing the Research Plan, including titles and summaries thereof, as well as names and contact information for personnel conducting the Research Plan.

Will include acknowledgment in publications

RECIPIENT will acknowledge the source of the DATA in all publications and presentations by including language identical to or substantially similar to the following: "Authors thank the National Cancer Institute and the American College of Radiology Imaging Network for access to the National Lung Screening Trial database. The interpretation and reporting of these data are the sole responsibility of the authors."

Will notify web site of publications

RECIPIENT shall submit a description of each publication resulting from the use of the DATA to the following website: http://biometry.nci.nih.gov/cdas and agree that the NCI may publicly display such information for NCI purposes.

Will abide by applicable law

RECIPIENT will use the DATA in compliance with all applicable local, State, and/or Federal laws and regulations including those for protections of human subjects.

CT Data Characteristics

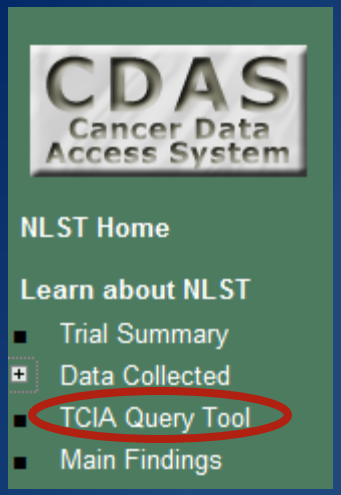
- ∽ 75,000 annual exams in >25,000 subjects
- Accessed through The Cancer Imaging Archive (TCIA)
- Technical parameters within protocol-defined ranges
  - Minimum 4-detector scanners (all vendors)
  - 120-140 KeV
  - 20-60 effective mAs
  - 1-2.5 mm section thickness
  - Pitch 1.25-2.0
  - Single smooth and/or sharp kernel per model

Non-image Data

- Demographics: Age, sex, race/ethnicity, education, marital status, height, weight
- Smoking history: Pk yrs, age started, ppd, total years smoked, age quit, current status
- Medical history: Lung diseases, cardiovascular diseases, HTN, DM, cancers
- Occupational history
- Family history of cancer

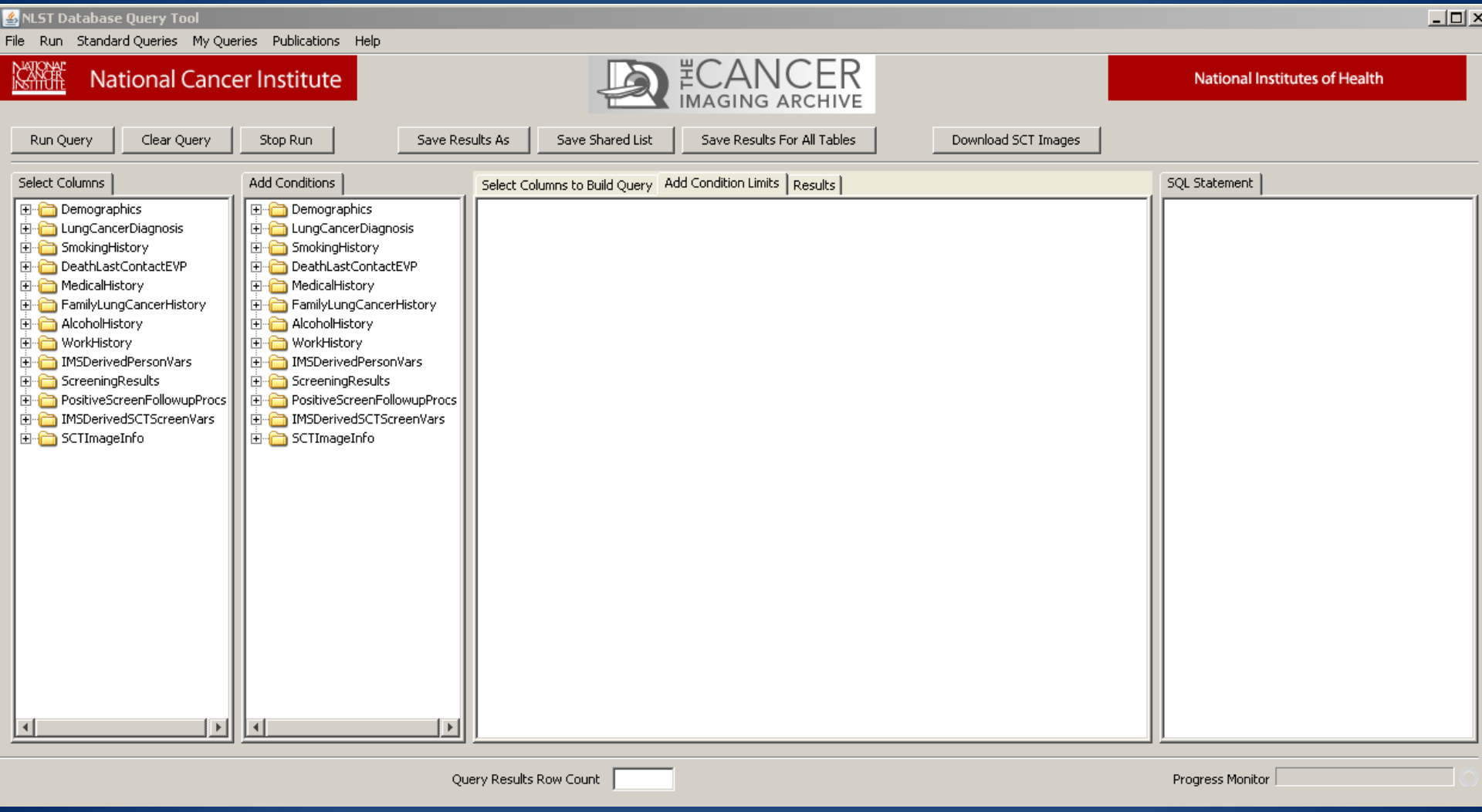
Non-image Data

- Screening: result, recorded nodules
- Diagnostics: imaging, biopsies
- Lung cancer: location, histology, stage
- Treatment: type(s), time course
- Progression: occurrence, site, second primary
- Mortality: occurrence, cause



Searching the NLST Database
TCIA-NLST Query Tool

- Web-based application for searching NLST database
- Enables selection of NLST CT scans and associated participant data meeting desired criteria
  - E.g. demographics, risk factors, screening result, CT scanner model/technical factors, lung cancer status, etc.
- Enables downloading of CT scans in DICOM format and associated participant data in spreadsheet format

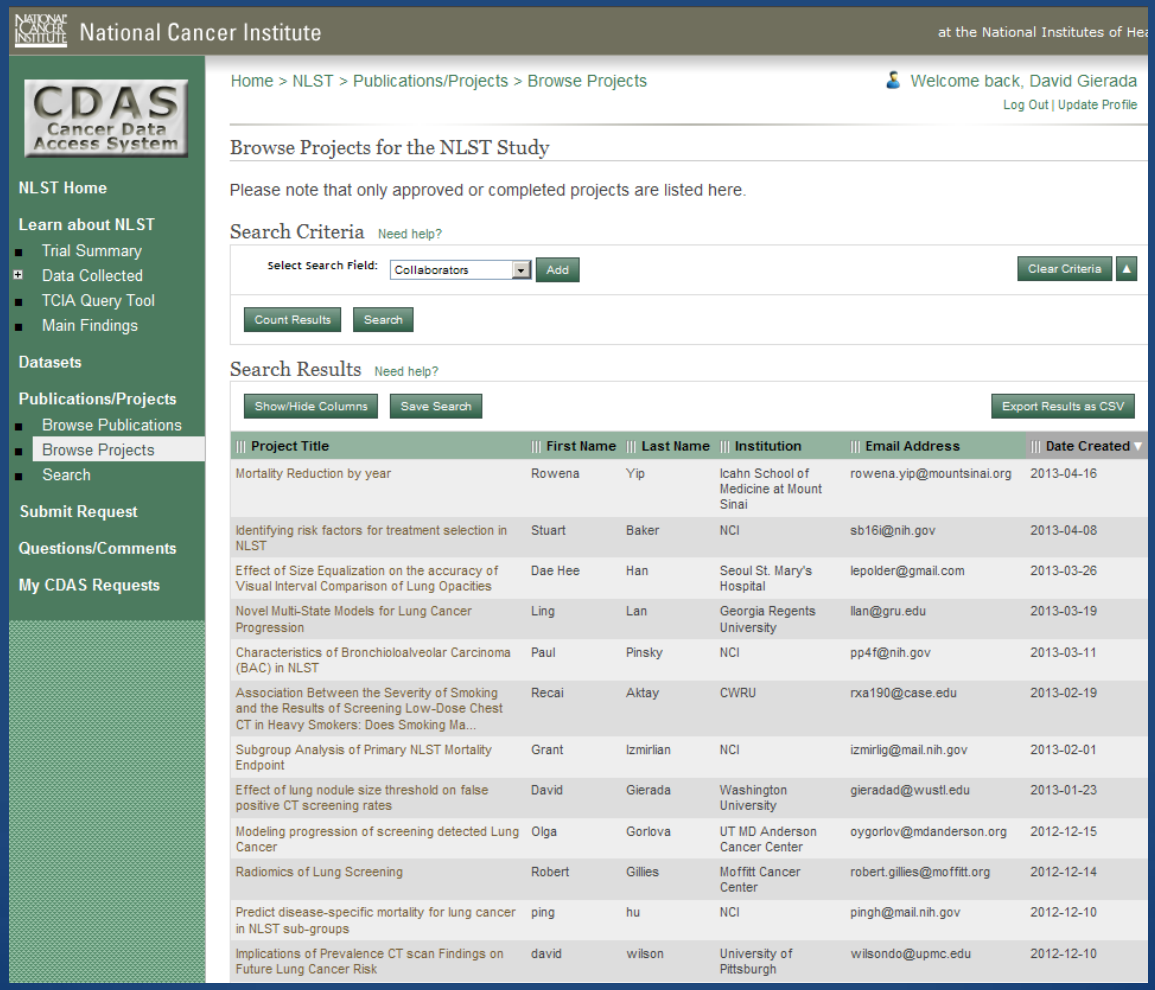

NLST Database Query Tool
File
Run
Standard Queries
My Queries
Publications
Help
National Cancer Institute
THE CANCER IMAGING ARCHIVE
National Institutes of Health
Run Query
Clear Query
Stop Run
Save Results As
Save Shared List
Save Results For All Tables
Download SCT Images
Select Columns
Add Conditions
Select Columns to Build Query
Add Condition Limits
Results
SQL Statement
Demographics
LungCancerDiagnosis
SmokingHistory
DeathLastContactEVP
MedicalHistory
FamilyLungCancerHistory
AlcoholHistory
WorkHistory
IMSDerivedPersonVars
ScreeningResults
PositiveScreenFollowupProcs
IMSDerivedSCTScreenVars
SCTImageInfo
Query Results Row Count
Progress Monitor

Biospecimens

- Blood, urine, sputum on ACRIN subset (~10,000)
  - http://www.acrin.org/ACRIN-NLSTBIOREPOSITORY.aspx
- Tissue slides and microarrays for most lung cancers
  - Collected, not yet available
- Separate application process

NLST Data Usage

Approved studies and dates on CDAS website



Project Title	First Name	Last Name	Institution	Email Address	Date Created
Mortality Reduction by year	Rowena	Yip	Icahn School of Medicine at Mount Sinai	rowena.yip@mountsinai.org	2013-04-16
Identifying risk factors for treatment selection in NLST	Stuart	Baker	NCI	sb16i@nih.gov	2013-04-08
Effect of Size Equalization on the accuracy of Visual Interval Comparison of Lung Opacities	Dae Hee	Han	Seoul St. Mary's Hospital	lepolder@gmail.com	2013-03-26
Novel Multi-State Models for Lung Cancer Progression	Ling	Lan	Georgia Regents University	llan@gru.edu	2013-03-19
Characteristics of Bronchioloalveolar Carcinoma (BAC) in NLST	Paul	Pinsky	NCI	pp4f@nih.gov	2013-03-11
Association Between the Severity of Smoking and the Results of Screening Low-Dose Chest CT in Heavy Smokers: Does Smoking Ma...	Recai	Aktay	CWRU	rxa190@case.edu	2013-02-19
Subgroup Analysis of Primary NLST Mortality Endpoint	Grant	Izmirlian	NCI	izmirlig@mail.nih.gov	2013-02-01
Effect of lung nodule size threshold on false positive CT screening rates	David	Gierada	Washington University	gieradad@wustl.edu	2013-01-23
Modeling progression of screening detected Lung Cancer	Olga	Gorlova	UT MD Anderson Cancer Center	oygorlov@mdanderson.org	2012-12-15
Radiomics of Lung Screening	Robert	Gillies	Moffitt Cancer Center	robert.gillies@moffitt.org	2012-12-14
Predict disease-specific mortality for lung cancer in NLST sub-groups	ping	hu	NCI	pingh@mail.nih.gov	2012-12-10
Implications of Prevalence CT scan Findings on Future Lung Cancer Risk	david	wilson	University of Pittsburgh	wilsondo@upmc.edu	2012-12-10

NLST Data Usage

- Epidemiologic/risk prediction
- CT-focused studies
  - Subjective visual reviews
  - Quantitative assessment of comorbidities +/- their relation to cancer
- Little related to quantitative assessment of nodules

Quantitative Image Analysis Studies

"Non-invasive Characterization of Peripheral Pulmonary Nodules of the Lung Adenocarcinoma Spectrum and Correlation with Clinical Outcomes using CANARY (Computer-Aided Nodule Assessment and Risk Yield)" (Mayo Clinic)

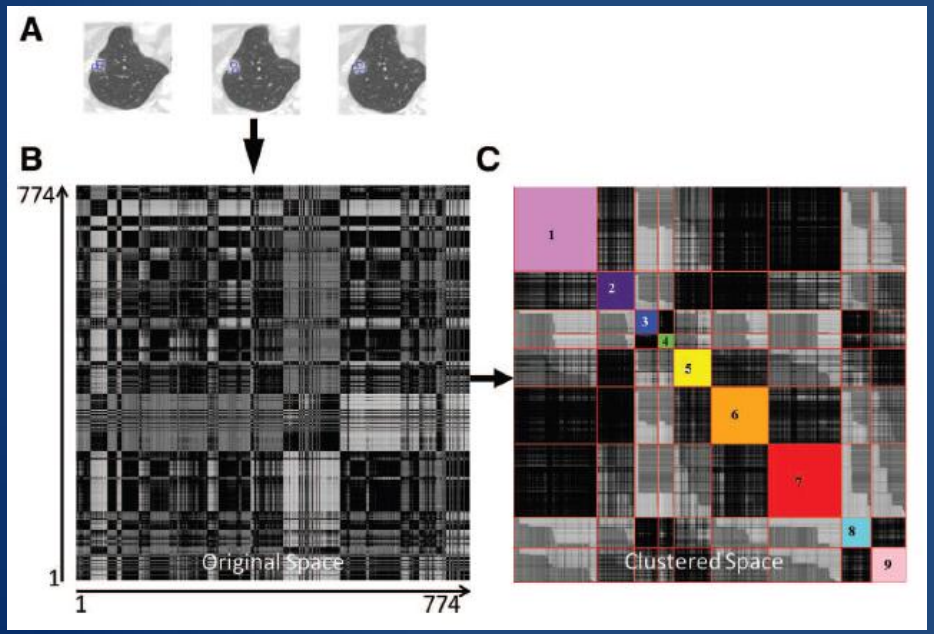
- Validation of algorithm predicting aggressiveness of adenocarcinoma
- Attenuation-based voxel clustering analysis to characterize tissue regions as invasive or noninvasive
  - High correlation with histology findings and survival

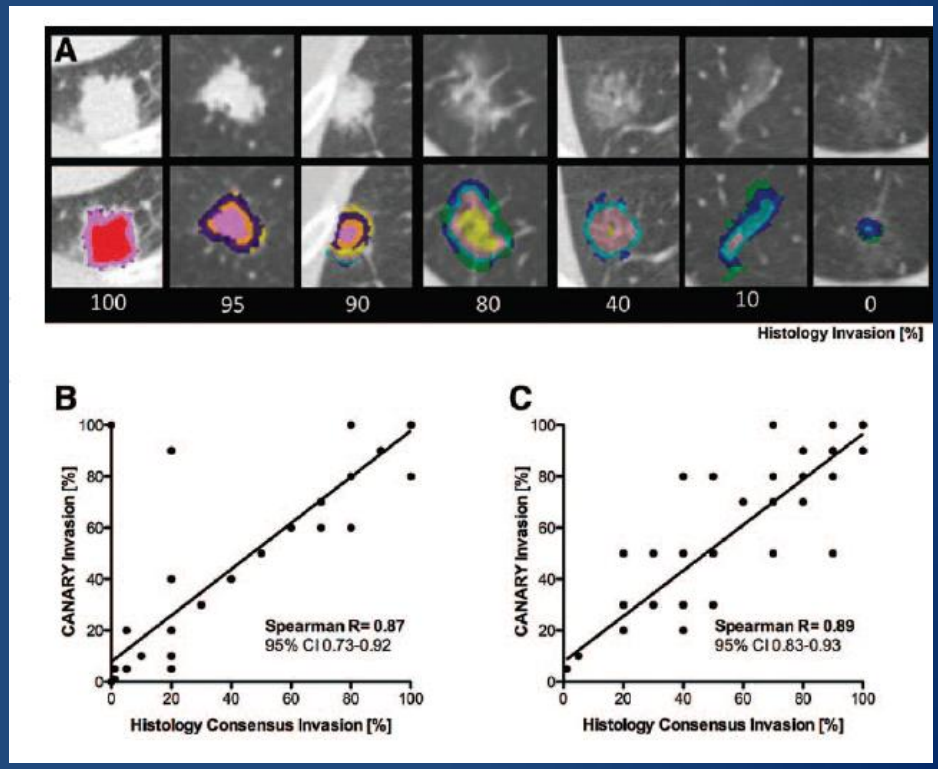
J Thorac Oncol. 2013 Apr;8(4):452-60. doi: 10.1097/JTO.0b013e3182843721.

Noninvasive characterization of the histopathologic features of pulmonary nodules of the lung adenocarcinoma spectrum using computer-aided nodule assessment and risk yield (CANARY)--a pilot study.

Maldonado F, Boland JM, Raghunath S, Aubry MC, Bartholmai BJ, Deandrade M, Hartman TE, Karwoski RA, Rajagopalan S, Sykes AM, Yang P, Yi ES, Robb RA, Peikert T.

Department of Medicine, Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, Minnesota 55905, USA.





Quantitative Image Analysis Studies

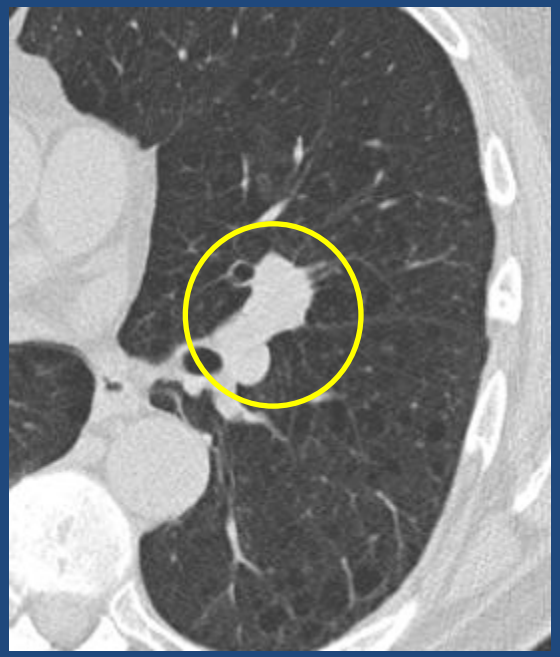
"Radiomics of Lung Screening" (Moffitt Cancer Center)

- Search for quantitative features of lungs associated with a low risk for lung cancer
- Quantitative descriptors of bronchiolar tree and terminal bronchiolar architecture

Image Processing Studies
Fundamental Issues

- Nodule segmentation algorithm development and performance assessment/comparison

Subsolid



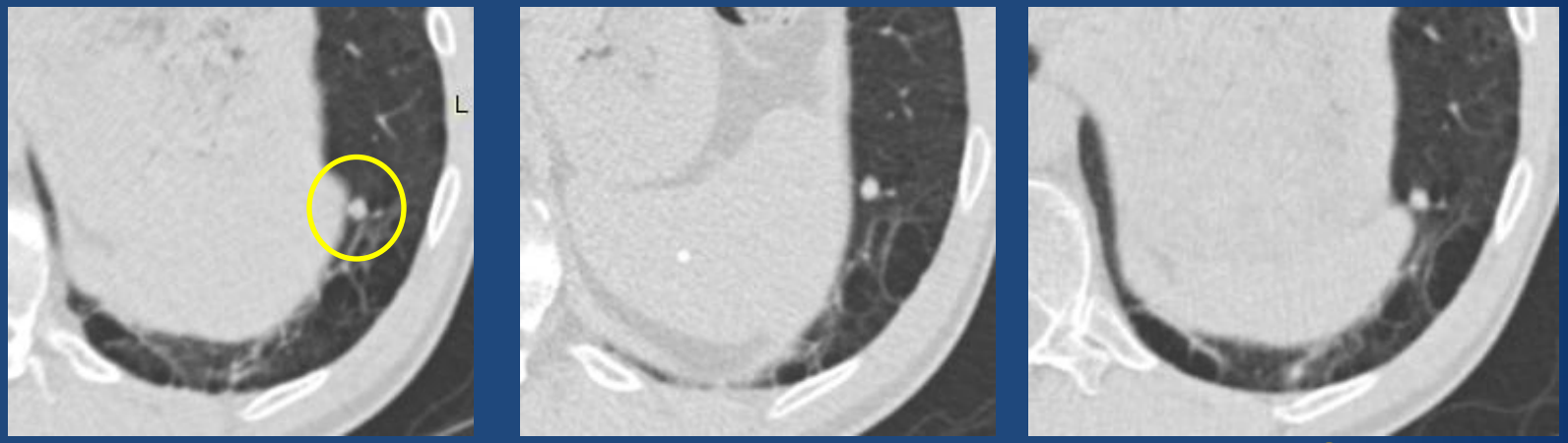
Vessel contact

Lung margins

Image Processing Studies
Fundamental Issues

- Determining minimum measurable change

Stable



T0 T1 T2

RUL Ca

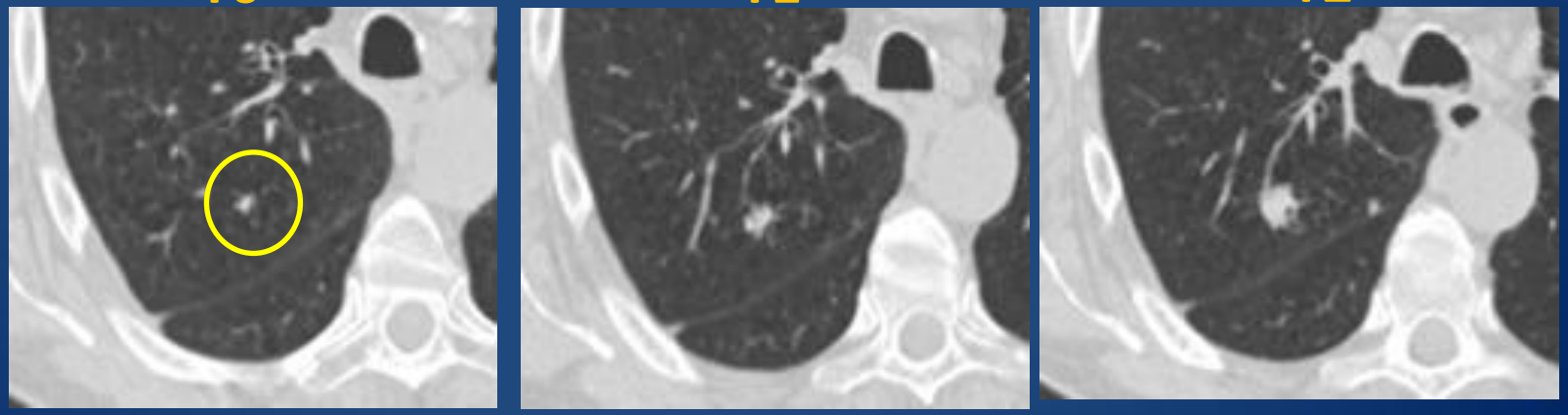
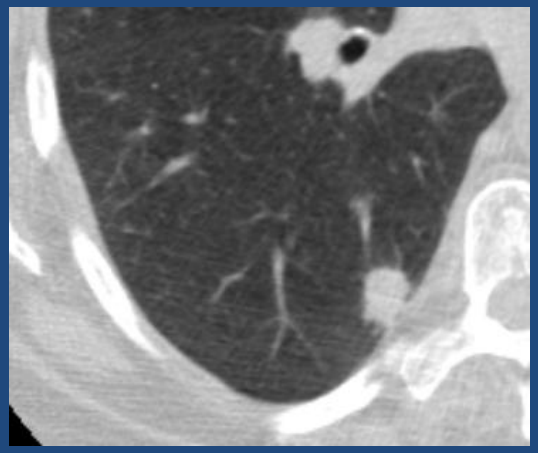


Image Processing Studies
Fundamental Issues

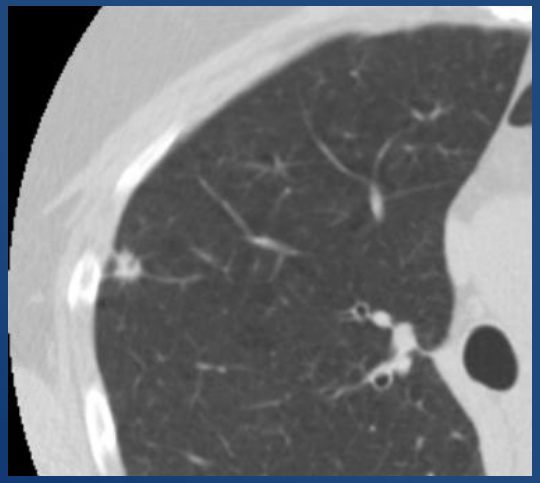
- Tissue characterization

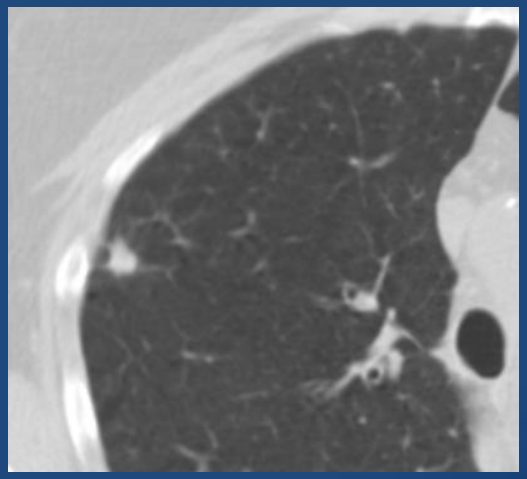
RLL Ca

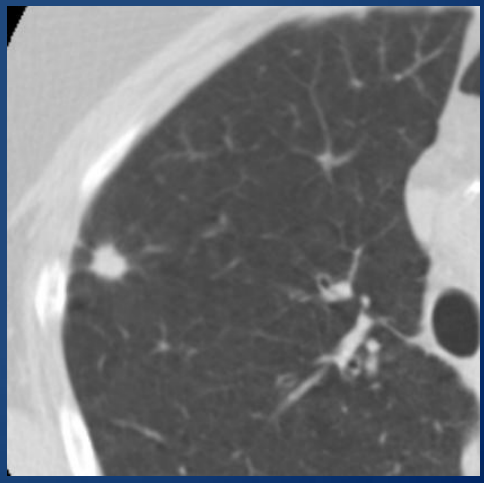


T0 T1 T2

RUL Ca







NLST Database Limitations

- Slice thickness
- Number of cancers scanned
  - 649/1060 cancers found before followup period
  - Stratified analyses (e.g. size, histology, stage, development vs. validation) reduce available sample sizes further
- Limited serial imaging (annual only)
- No nodule-level pathology correlation
- Technology will become dated over time

Other Databases

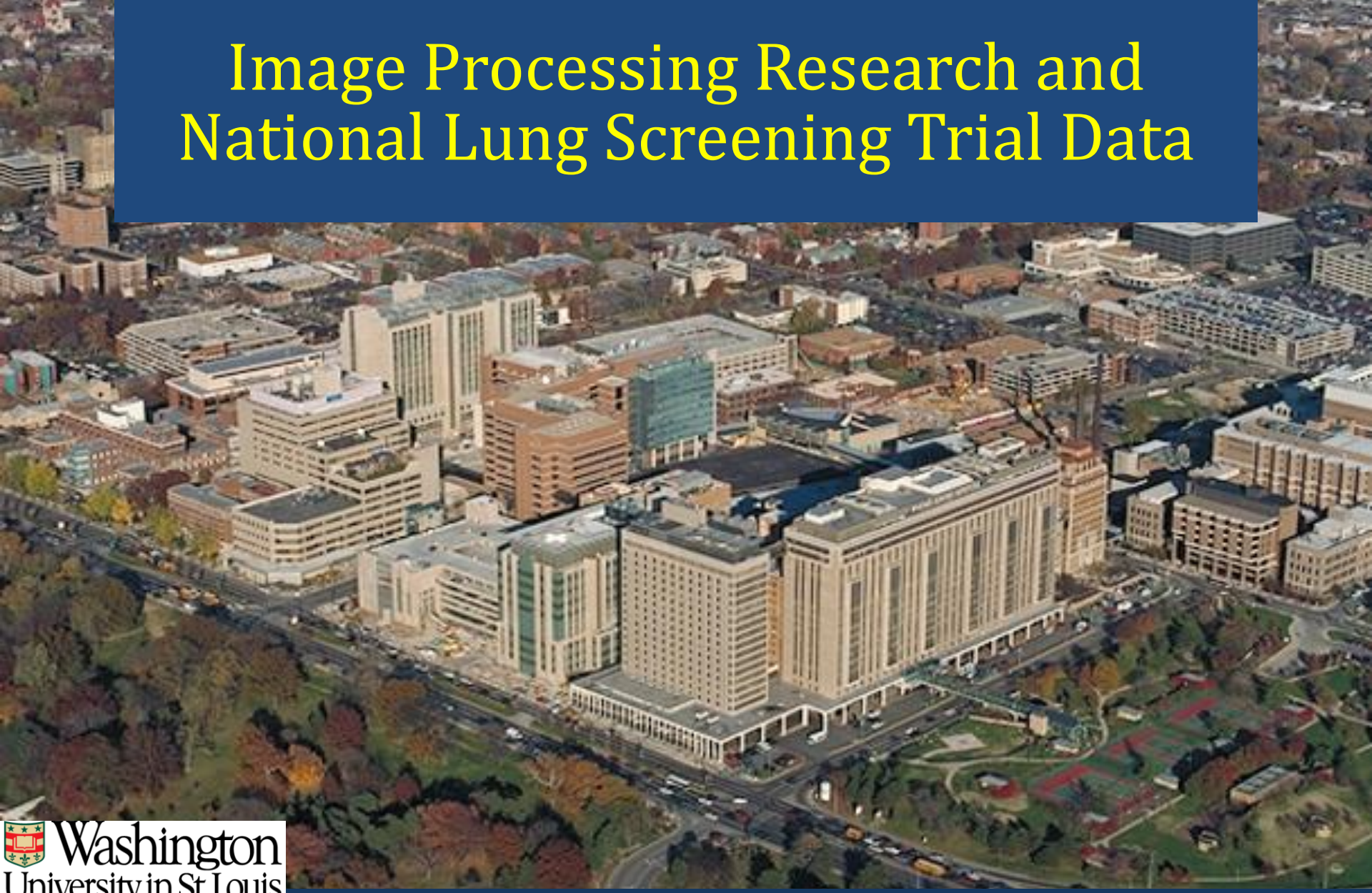
- LIDC-Lung Image Database Consortium*
  - 1,010 cases of nodules marked by radiologists (268 with path)
- RIDER-Reference Image Database to Evaluate Therapy Response*
  - 32 cases of longitudinal studies after therapy
- TCGA-LUAD-The Cancer Genome Atlas Lung AdenoCa*
  - 38 cases of CTs associated with genomic and path data
- I-ELCAP-Public Research Access Database
  - 50 low-dose CTs with marked nodule locations

*In TCIA

Image Processing Research in CT Screening
Future Needs for Progress

- Sustainable, growing image archive(s)
  - Voluntary registry?
- Demand/incentive for better tools
  - Widespread screening implementation?
- Funding
  - Federal and/or private

Image Processing Research and National Lung Screening Trial Data



Washington University in St. Louis
SCHOOL OF MEDICINE
MIR Mallinckrodt Institute of Radiology

Lung Cancer Workshop X

David S. Gierada, M.D.