THE CANCER IMAGING ARCHIVE (TCIA): AN INFORMATION RESOURCE TO ENABLE OPEN SCIENCE

Fred Prior1, Ken Clark1, Paul Commean1, John Freymann2, Carl Jaffe3, Justin Kirby3, Stephen Moore3, Kirk Smith1, Lawrence Tarbox3, Bruce Vendt1, and Guillermo Marquez1

1Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO 63110
2Clinical Research Directorate / CMRP SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702
3Boston University School of Medicine, Boston MA 02118
4National Cancer Institute, Bethesda, MD 20892

ABSTRACT: Reusable, publicly available data is a pillar of open science. The Cancer Imaging Archive (TCIA) is an open image archive service supporting cancer research. TCIA collects, de-identifies, curates and manages rich collections of oncology image data. Image data sets have been contributed by 28 institutions and additional image collections are underway. Since June of 2011, more than 2,700 users have registered to search and access data from this freely available resource. TCIA encourages and supports cancer-related open science communities by hosting and managing the image archive, providing project wiki space and searchable metadata repositories. The success of TCIA is measured by the number of active research projects it enables and the number of scientific publications and presentations that are produced using data from TCIA collections.

Big Data, Open Science, Open Data

- The Big Data Challenge: The volume of scientific data doubles each year; how can this information be effectively managed?
- Open Science: Using Open Source software to capture and manage Open Data to encourage and support research and education
- Open Data: Well curated information that is freely accessible by the research community
- The Cancer Imaging Archive (TCIA): An Open Data repository of cancer images and related clinical data created under contract to NCI for the purpose of enabling Open-Science cancer research

TCIA: https://cancerimagingarchive.net

TCIA by the Numbers

- Collections: 60
- Subjects: 30,000
- Imaging Exams: > 79,000
- Images: > 26,000,000

Image Counts by Anatomical Region, Modality, Cancer Type

<table>
<thead>
<tr>
<th>Anatomical Region</th>
<th>CT*</th>
<th>DX*</th>
<th>MG*</th>
<th>MR*</th>
<th>PT*</th>
<th>RT**</th>
<th>Cancer Type(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>8,446</td>
<td>1</td>
<td>80,738</td>
<td>6,186</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>50,000</td>
<td>4</td>
<td>250,000</td>
<td>10,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervix</td>
<td>20,000</td>
<td>1</td>
<td>100,000</td>
<td>5,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td>50,000</td>
<td>4</td>
<td>250,000</td>
<td>10,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung/Chest</td>
<td>2,500,000</td>
<td>50</td>
<td>125,000</td>
<td>5,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ovary</td>
<td>5,000</td>
<td>1</td>
<td>25,000</td>
<td>1,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate</td>
<td>10,000</td>
<td>5</td>
<td>50,000</td>
<td>2,500</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Imaging Modalities: CT = computed tomography, DX = digital x-ray, MG = mammography, MR = magnetic resonance imaging, PT = positron emission tomography.
** RT = radiation treatment planning (includes VCON objects for planning, dosage, report)
Note: Phantom images available but not listed here.

Contributed Images and Data

- Contributed Clinical Data
- De-identified Pathology Images
- Curated Images
- Associated Image
- Project Metadata
- Clinical Data
- Image Metadata
- Annotations
- Project Metadata

TCIA-Enabled Research

- TCIA provides the international research community with free access to previously unavailable, or prohibitively costly research data

TCIA Users by Month (cumulative)

Downloads by Month

Active Research Communities & TCIA

<table>
<thead>
<tr>
<th>Community</th>
<th>Collaborative Projects</th>
<th>Active Researchers</th>
<th>TCIA Collections Utilized</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCGA Glioma Phenotype Group</td>
<td>11</td>
<td>&gt;20</td>
<td>TCGA-GBM, TCGA-LSG</td>
</tr>
<tr>
<td>TCGA Breast Phenotype Group</td>
<td>4</td>
<td>&gt;12</td>
<td>TCGA-BRCA</td>
</tr>
<tr>
<td>TCGA Renal Phenotype Group</td>
<td>8</td>
<td>&gt;13</td>
<td>TCGA-KIRC</td>
</tr>
<tr>
<td>Quantitative Imaging Network</td>
<td>&gt;16</td>
<td>&gt;190</td>
<td>QIN Breast, QIN Phantom, QIN Lung, QIN Prostate</td>
</tr>
<tr>
<td>National Lung Screening Trial Related Groups</td>
<td>8</td>
<td>25</td>
<td>NLST</td>
</tr>
</tbody>
</table>

Metric of Success

The success of TCIA is measured by:
- the number of active research projects it currently enables (>40);
- the number of peer reviewed publications (29), presentations (9), Doctoral/Master’s Thesis (2), Educational Exhibits (4), and Posters (21) that have been produced using data from TCIA collections.

References