

Quantitative Imaging Network: Data Sharing and Competitive Algorithm Validation Leveraging The Cancer Imaging Archive¹

Jayashree Kalpathy-Cramer*, John Blake Freymann[†], Justin Stephen Kirby[†], Paul Eugene Kinahan[‡] and Fred William Prior[§]

*Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA; [†]Clinical Research Directorate/Clinical Monitoring Research Program (CMRP), Leidos Biomedical Research Inc, Frederick National Laboratory for Cancer Research, Frederick, MD; [‡]Department of Radiology, University of Washington, Seattle, WA; [§]Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, MO

Abstract

The Quantitative Imaging Network (QIN), supported by the National Cancer Institute, is designed to promote research and development of quantitative imaging methods and candidate biomarkers for the measurement of tumor response in clinical trial settings. An integral aspect of the QIN mission is to facilitate collaborative activities that seek to develop best practices for the analysis of cancer imaging data. The QIN working groups and teams are developing new algorithms for image analysis and novel biomarkers for the assessment of response to therapy. To validate these algorithms and biomarkers and translate them into clinical practice, algorithms need to be compared and evaluated on large and diverse data sets. Analysis competitions, or “challenges,” are being conducted within the QIN as a means to accomplish this goal. The QIN has demonstrated, through its leveraging of The Cancer Imaging Archive (TCIA), that data sharing of clinical images across multiple sites is feasible and that it can enable and support these challenges. In addition to Digital Imaging and Communications in Medicine (DICOM) imaging data, many TCIA collections provide linked clinical, pathology, and “ground truth” data generated by readers that could be used for further challenges. The TCIA-QIN partnership is a successful model that provides resources for multisite sharing of clinical imaging data and the implementation of challenges to support algorithm and biomarker validation.

Translational Oncology (2014) 7, 147–152

Introduction

The Quantitative Imaging Network (QIN) [1], supported by the National Cancer Institute (NCI), is designed to promote research and development of quantitative imaging methods and candidate biomarkers for the measurement of tumor response to therapies in clinical trial settings. Current projects focus on development and adaptation of quantitative image analysis algorithms and software, image acquisition protocols, and application of these methods in current and planned clinical trials. Each QIN team is multidisciplinary and includes oncologists, clinical and basic imaging scientists, and frequently industrial partners to help promote the adoption of the newly developed quantitative imaging methods.

To date, 17 QIN teams have been selected through the NIH peer review process. Four working groups addressing common, cross-cutting issues have been established, including data collection, image

Address all correspondence to: Jayashree Kalpathy-Cramer, PhD, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, 149 13th St, Charlestown, MA 01940. E-mail: kalpathy@nmr.mgh.harvard.edu

¹This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health (NIH), under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. J.K.-C. is funded in part by the NIH grants U01CA154602 and R00LM009889 and a contract ST13-4130. P.E.K. is funded in part by the NIH grant U01CA148131 and Contract 24XS036-004.

Received 16 December 2013; Revised 17 March 2014; Accepted 19 March 2014

Copyright © 2014 Neoplasia Press, Inc. All rights reserved 1944-7124/14/\$25.00
DOI 10.1593/do.13862

analysis, informatics, and clinical trial design. These working groups are staffed by members of the QIN teams and overseen by an Executive Committee comprising NCI staff and QIN site Principal Investigators. A Coordinating Committee that includes the working group chairs and cochairs as well as NCI staff facilitates the exchange of information among the working groups and coordinates tasks that cross the boundaries of the working groups.

An integral aspect of the QIN mission is to facilitate collaborative activities that seek to develop best practices for the analysis of cancer imaging data. The working groups and QIN teams are continually developing new algorithms for image analysis and novel biomarkers for the assessment of response to therapy in a number of cancers. To validate these algorithms and biomarkers and translate them into clinical practice, they need to be compared and evaluated on common data sets that are both large and diverse. Appropriate data of adequate quality and provenance can be difficult to come by, thus data sharing has become an important component of QIN activities.

To clarify rules of engagement and to encourage meaningful data sharing, QIN adopted a data sharing policy in November 2013. The spirit of this policy is one of collaboration and flexibility intended to introduce a minimal amount of oversight and/or committee work to QIN members. The QIN is committed to providing commercial and academic investigators an opportunity to access data collected as part of QIN studies for purposes that are consistent with the missions of the QIN and the NCI. All QIN members, associate/affiliate members, external collaborators, and companies are made aware of the guiding principles of the QIN Resource Sharing Policy. The objective of the data sharing policy is to help maximize the effectiveness of the QIN by fostering an environment of collaboration and sharing while addressing concerns of data being used without consent either by a member of the QIN or an external collaborator. Concerns about inappropriate data use of these shared data could hinder the multicenter collaborations within the QIN.

The guiding principles of the QIN data sharing policy are as follows [2]:

1. Fairness, collegiality, and cooperation in the joint pursuit of scientific advancement. The QIN encourages use of resources generated within the QIN consistent with the missions of the QIN and the NCI.
2. The QIN has a responsibility to ensure that the use of QIN resources is ethical and scientifically sound.
3. Data will be shared in a manner that allows good use to be made of them. This includes, for example, proper documentation, indexing, or curation/vetting of data where appropriate.
4. Appropriate attribution and acknowledgement for QIN resources will be provided.
5. QIN data and images typically will not be released to individuals or companies before the publication of the project's primary aim manuscript.
6. Data sharing will not burden the QIN's resources such as to impede its ability to pursue its primary research.
7. Investigators interested in asking research questions of data collected as part of QIN projects are encouraged to do so as a collaborative effort within the QIN structure.
8. Investigators interested in using QIN data must agree to adhere to the QIN publication policy.

To facilitate data sharing, the QIN increasingly relies on the facilities of The Cancer Imaging Archive (TCIA) as a resource, which addresses many

of the principles set forth in the data sharing policy. Most importantly, the TCIA provides a mechanism for access-controlled data sharing and extensive deidentification services, which comply with Health Insurance Portability and Accountability Act (HIPAA) regulations. This drastically reduces the burden on QIN sites when sharing their data.

In December 2010, Washington University was awarded a contract to build and manage a full-featured cancer imaging archive service that would support NCI-funded research activities and the cancer research community at large. TCIA is a service that provides a public repository of cancer images and related clinical data. It has been showcased as an example of a high-quality big data resource [3] and was created for the express purpose of enabling open-science research [4]. Currently, more than 26 million radiologic images and several thousand pathology images are contained in this repository. TCIA supports more than 40 active cancer research teams with data and collaboration resources [2]. In addition, a global research community uses TCIA data in a wide array of cancer research efforts (e.g., [5–7]). TCIA was selected by the Bioinformatics and Data Sharing working group as the official repository for sharing QIN data. A number of QIN sites (Vanderbilt University, Oregon Health and Science University (OHSU), University of Washington, Moffitt Cancer Center/MAASTRO Clinic, Brigham and Women's Hospital, University of Iowa, and University of Pittsburgh) currently use TCIA to host and manage images as both public and private collections.

Analysis competitions or "challenges" are being recognized as a practical means of engaging the community in identifying best algorithms or approaches to solve a given problem [8–10]. Globally, challenges are being used to develop solutions to issues that range from climate change, cybersecurity, self-driving cars, and autonomous robots to developing low-cost biomarkers for tuberculosis [11–13]. The computer science and medical imaging communities have held series of challenges, typically held in conjunction with conferences such as Medical Image Computing and Computer Assisted Intervention (MICCAI) or Institute of Electrical and Electronics Engineers (IEEE) International Symposium on Biomedical Imaging (ISBI) [14] to allow algorithm developers of medical image analysis software to compete and compare the performance of algorithms on common data sets.

Challenges, some of which have spurred major advancement of their fields, are now being conducted within the QIN as a means to stimulate the development of image analysis algorithms and novel biomarkers in the wider community. A key component of such challenges, especially in the domain of medical imaging, is access to data. These data need to be shared in a privacy-protected/HIPAA-compliant manner, of a sufficiently large sample size to demonstrate statistical significance, encompass a broad spectrum of presentations of the target disease, and be of sufficient quality to allow its reuse in retrospective research. Publicly accessible archives such as the TCIA that host well-curated clinical or research data in a HIPAA-compliant fashion can be a tremendous resource to the algorithm development community and can be used in challenges to allow algorithm developers to compare the performance of their algorithms against other algorithms on the same data.

In this article, we will review the role currently played by TCIA in support of NCI's QIN and describe a few of the imaging challenge competitions being conducted in the QIN that are conducted using data from TCIA.

Materials and Methods

TCIA encourages and supports cancer-related open-science communities by deidentifying, hosting, and managing image collections and

providing searchable metadata repositories to facilitate collaborative research [4]. To assure the collections managed by TCIA are of high quality and value to the scientific community, NCI staff work directly with potential data providers to evaluate new resources. Part of this process, in compliance with Washington University Institutional Review Board (IRB) protocols, is to validate that proper informed consent was obtained or other appropriate steps are taken in compliance with US and international laws governing human subject research. Once a data set has been accepted by NCI, the TCIA team works with the submitter to facilitate information upload to TCIA. The submitter is responsible for identifying appropriate data for submission and describing this data to TCIA staff. The description includes imaging protocol, modality, number of data sets, and information for meaningful series descriptions.

The TCIA team has defined standard operating procedures for image data acquisition, deidentification, and curation that adhere to the HIPAA Privacy Rule and leverage the Digital Imaging and Communications in Medicine (DICOM) standards for deidentification outlined in the Attribute Confidentiality Profile (DICOM PS 3.15: Appendix E) [15]. TCIA curation focuses primarily on removal of all protected health information while retaining scientifically meaningful standard and private data elements. This process begins with provision of Radiological Society of North America (RSNA; Oak Brook, IL) Clinical Trials Processor [15,16] software that properly deidentifies the information, again in compliance with Washington University IRB protocols. Data submission experts from TCIA staff provide training and support on the use of that software and feedback on the status of the upload and curation process. Once information has been uploaded to TCIA's intake servers, it is curated to assure the anticipated number of patients, studies, images, and expected modalities were received. TCIA curation includes a review of each image to identify gross artifacts. Data that are questioned are placed into quarantine and reviewed with the image submitter and NCI staff. A final inspection to ensure proper deidentification is performed before the data are moved to the public TCIA servers for dissemination [17]. The methods and tools used in this extensive deidentification and curation process are also shared with the wider research community in the form of a deidentification knowledge base, which is available from the TCIA web site (<http://cancerimagingarchive.net/>).

TCIA operates as a system of federated software and data repositories with all information linked using common subject identifiers. This suite of tools includes open-source applications such as the National Biomedical Imaging Archive [18], Clinical Trials Processor [15], AIME Data Service [19], and Confluence wiki [20] to manage images and associated image annotations and markup and provide wiki functionality. Additional software has been created specifically for extending TCIA's ability to support associated clinical data and to aid in the deidentification and curation processes. TCIA has also developed an open-source private cloud infrastructure with clustered deployments of these tools for increased performance and reliability [4].

The mission of the Image Analysis Working Group (WG) of the QIN includes efforts to "provide guidance, coordination, consensus building, and awareness regarding the development of algorithms and methods for quantitatively analysis." The WG consists of Dynamic Contrast-Enhanced Magnetic Resonance Imaging (MRI; DCE-MRI) and the Positron Emission Tomography-Computer Tomography (PET/CT) subgroups. These groups have organized challenges to facilitate the comparison of algorithms being developed by the different member sites on common data sets. An important component of these challenges is

the sharing of a common data set to the participants. The main requirements for data sets that can be used for these challenges include images that are 1) publicly available and shareable, 2) deidentified (for data that might contain protected health information), 3) of a sufficiently large sample size, and 4) of suitable quality and diversity. As images in the TCIA have undergone a rigorous deidentification and curation process, using them in challenges greatly reduces the burden on the organizers to select, deidentify, and curate images for use in challenges. Furthermore, images from a set of different collections can be used to ensure sufficient sample size and diversity of image appearance and acquisition protocols. The download manager and "shared list" feature of TCIA support the easy dissemination of the images.

A lung nodule segmentation challenge, conducted under the auspices of the PET/CT working group, provided a data set of 52 nodules from 41 CT studies, all currently available in TCIA. These included 10 nodules each from the Lung Image Database Consortium (LIDC) [21] and Reference Image Database to Evaluate Therapy Response (RIDER) [22] collections as well as 10 nodules each from the Stanford and Moffitt collections that were shared as part the QIN data-sharing plan. Additionally, 12 nodules in a phantom (single volume), scanned at Columbia University (New York, NY) were also shared. Participants provided segmentations created using their automatic or semiautomatic segmentation algorithms. As described in [23], repeatability (and bias in the case of the phantoms) as well as a variety of performance metrics were calculated for the images submitted by the participants.

In the DCE-MRI challenge, data from two visits of 10 patients each were made available by the OHSU QIN team. These data were generated as part of a clinical trial of breast cancer therapy conducted at OHSU and consisted of images from a baseline scan as well as a scan after the first round of neoadjuvant chemotherapy. The goal of the challenge was to evaluate the ability of the different software packages for the analysis of DCE-MRI data to separate responders from nonresponders, as seen in pathology [24]. These data were shared through TCIA.

The TCIA has also supported other image analysis challenges including the Multimodal Brain Tumor Segmentation (BraTS) challenge [25] held at MICCAI 2013 as well as the prostate segmentation challenge at ISBI (see [26]). The goal of the BraTS challenge was to gauge the current state of the art in automated brain tumor segmentation and compare performance between different methods by comparing them to human delineations generated by expert radiologists and neuro-oncologists. The segmentations were to be performed on multimodal MR imaging consisting of T_1 precontrast, T_1 postcontrast, fluid attenuated inversion recovery (FLAIR), and T_2 images. For each case, each tumor voxel was labeled using one of four labels (enhancing, necrosis, edema, and nonenhancing tumor), although not all tumor images have all four classes present. The performance of the algorithms was evaluated by comparing the overlap between the algorithm-generated labels and the ground truth consisting of the human-generated labels. For the 2013 challenge, for the leaderboard and the on-site challenge phases, cases with all four modalities present were randomly selected from the TCIA The Cancer Genome Atlas-glioblastoma multiforme (TCGA-GBM) data set. After preprocessing steps consisting of registration and skull stripping, the images were made available to the participants.

Results

TCIA supports a large and growing collection of images of a variety of modalities and anatomic sites as shown in Table 1.

Table 1. Data Archived in TCIA.

Anatomic Region	No. of Subjects with One or More Imaging Studies by Modality							Cancer Type(s)
	CT	DX	MG	MR	PET	RT	NM	
Brain	14			466	5			Glioma and glioblastoma multiforme
Breast	30		22	237	60			Invasive carcinoma
Colon	825							Adenocarcinoma
Extremity				12				Sarcoma
Head/neck	118				114	96		Squamous cell carcinoma
Kidney	183			63				Clear cell carcinoma
								Papillary cell carcinoma
Liver	10			1				Hepatocellular carcinoma
Lung/chest	1594	237			222		1	Carcinoid
								Adenocarcinoma
								Squamous cell carcinoma
								Bronchioloalveolar carcinoma
								Large cell carcinoma
								Non-small cell carcinoma
								Small cell carcinoma
Ovary	60			1				Ovarian serous cystadenocarcinoma
Prostate	8			207	8			Adenocarcinoma

DX indicates digital radiography; MG, mammography; RT, radiotherapy; NM, nuclear medicine.

QIN teams currently use TCIA to manage shared image collections and support challenges. In some instances, the collections are available to the public, although some remain restricted because they represent ongoing, unpublished research or the test sets for challenges.

Two public collections include data sets on head and neck cancer and non-small cell lung cancer. The QIN-HeadNeck collection consists of 138 PET/CT patient data sets provided by University of Iowa. This collection is a set of patients with head and neck cancer, each of whom has had multiple PET/CT 18F-fluorodeoxyglucose (FDG) scans before and after therapy and with follow-up scans where clinically indicated. The NSCLC Radiogenomics collection contains images from patients with non-small cell lung cancer imaged before surgical excision with both thin-section CT and whole-body PET/CT scans acquired under IRB approval from Stanford University and the Veterans Administration Palo Alto Health Care System. The imaging data for the first installment of 26 cases are available in TCIA, whereas the microarray data acquired from the excised samples are available on the National Center for Biotechnology Information Gene Expression Omnibus (Bethesda, MD) [27,28].

The remainder of the QIN data sets currently have restricted access. These cover brain, breast, head and neck, lung, prostate, and sarcoma cancer types. Additionally, several sites have contributed various types of phantom data. These have been used to support the QIN challenges and other projects, which involved multi-institutional analysis of data. For example, the QIN-Prostate collection provided by Brigham and Women's Hospital contains 22 patient cases of multiparametric MRI images collected for the purposes of detection and/or staging of prostate cancer. The MRI parameters include T_1 - and T_2 -weighted sequences as well as diffusion-weighted and DCE-MRI. The collection has been used to provide clinical image data for the development and evaluation of quantitative methods for prostate cancer characterization using multiparametric MRI and for a DCE-MRI arterial input function comparison study between Brigham and Women's Hospital and Vanderbilt University [29].

The TCIA has provided access to high quality, well-curated data sets that have facilitated the organization of five image analysis challenges, within the QIN and at meetings such as ISBI and MICCAI. As part of these challenges, expert human reader generated segmentations have been created and added to TCIA. These segmentations

can be used by algorithm developers to validate their algorithms on an individual basis or as the source material for future challenges.

Discussion

The usefulness of information repositories is enhanced by the availability of programmatic interfaces that enable analysis and visualization applications to query and retrieve images without human intervention. Such native access can give researchers the ability to create data "mashups" [30] and extend their image analysis or machine learning algorithms to directly mine data from information repositories. A new middleware platform called Project Bindaas [31] facilitates creating web service-based interfaces that allow data providers to share data stored in databases using a popular standard for developing web services called Representational State Transfer [32]. Developers can use the Representational State Transfer interface with most modern languages to rapidly create and deploy applications that can consume data contained in the underlying database. Project Bindaas (Emory University, Atlanta, GA) has been used to develop a set of services that are used to manage Annotation and Imaging Markup objects generated by researchers that are making use of TCIA imaging collections [33]. Using these middleware tools, a programmatic interface has been added to TCIA and is being adopted by some QIN research teams to integrate access to TCIA data directly into their imaging software.

A number of research groups are investigating platforms such as HUBzero [34], Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC), and NITRC-Computational Environment (NITRC-CE) [35,36] for sharing algorithms and software for quantitative image analysis. The programmatic interface to TCIA allows collaborative groups to use these and other cloud computing environments to directly bring in data from TCIA, apply these shared tools in a transparent manner, and directly compare the results of the different algorithms or analysis techniques.

Members of the cancer imaging community have created derived data such as segmentations, markups, and image descriptors for images in TCIA during the course of their research [7,33]. TCIA maintains a dedicated server to store and make available Annotation and Imaging Markup markups. Additionally, when challenges use the data in TCIA, typically human annotations by experts are generated as the ground

truth. Such data, e.g., from the ISBI challenges, have been contributed back to TCIA and are stored on the associated challenge wiki pages. They serve as a valuable resource to validate algorithms both within and outside the context of challenges. Importantly, image segmentation challenges such as the lung nodule segmentation challenge organized within the QIN and the MICCAI-BRaTS challenge have demonstrated the success of automated algorithms in generating such derived data. In the future, we will explore making use of these validated algorithms to generate derived data and add further value to the data in TCIA.

As TCIA evolves and grows, its processes must be continually optimized to ensure that new collections of increasing size and complexity can be brought online in a timely and cost-efficient manner. At some point, it becomes inefficient for researchers to download large data sets to their local computing environments due to limitations of network throughput and storage requirements. It would then become more efficient to colocate high-performance computing [37] with large-capacity information resources such as TCIA. Extending TCIA with a user interface that enables researchers to easily and efficiently select data and algorithms and launch collocated deep computing jobs that return only the analysis results will greatly enhance the utility of TCIA for the cancer research community.

In summary, the TCIA is a valuable resource for the QIN and the larger cancer imaging community. TCIA's rich data sets are generally hard to obtain for computer scientists. Members of the QIN will continue to share high-value data sets in a HIPAA-compliant manner into a well-curated and searchable environment. These data sets can include high-quality images, imaging metadata, as well as other clinical data. Such data can facilitate the validation of imaging biomarkers and support reproducible research by providing an avenue to share the data used in publications or for challenges.

The challenges conducted by QIN would not have been possible without a data-sharing mechanism like TCIA. The TCIA-QIN challenges provide a model and resources for future challenges. In addition to DICOM imaging data, many TCIA collections provide linked clinical, pathology, and even ground truth segmentation data generated by human readers, which could be used for additional challenges. Results from future challenges could readily be made available to serve as benchmarks against which further algorithms could be tested.

References

- Clarke LP, Croft BS, Nordstrom R, Zhang H, Kelloff G, and Tatum J (2009). Quantitative imaging for evaluation of response to cancer therapy. *Transl Oncol* **2**, 195–197.
- Quantitative Imaging Network Collections. Available at: <https://wiki.cancerimagingarchive.net/x/wwEy>. Accessed January 13, 2014.
- WhiteHouse (2012). Big data fact sheet: Big Data Across the Federal Government. Available at: http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf.
- Prior F, Clark K, Commean P, Freymann J, Jaffe C, Kirby J, Moore S, Smith K, Tarbox L, Vendt B, et al. (2013). TCIA: an information resource to enable open science. *Conf Proc IEEE Eng Med Biol Soc* **2013**, 1282–1285.
- Hunter L (2013). *Radiomics of NSCLC: Quantitative CT Image Feature Characterization and Tumor Shrinkage Prediction*. University of Texas, Houston, TX. Available at: http://digitalcommons.library.tmc.edu/cgi/viewcontent.cgi?article=1365&context=utgsbs_dissertations..
- Sivakumar S and Chandrasekar C (2013). Lung nodule detection using fuzzy clustering and support vector machines. *Int J Eng Technol* **5**, 179–185.
- Zinn PO, Majadan B, Sathyan P, Singh SK, Majumder S, Jolesz FA, and Colen RR (2011). Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme. *PLoS One* **6**, e25451.
- Stine DD (2009). *Federally Funded Innovation Inducement Prizes*. DIANE Publishing. Available at: <http://www.fas.org/spp/crs/misc/R40677.pdf>.
- General Services Administration. A partnership between the public and the government to solve important challenges. 2014. Available at: <https://challenge.gov/>. Accessed January 13, 2014.
- Kaggle. Kaggle, the leading platform for predictive modeling competitions. 2014. Available at: <http://www.kaggle.com/competitions>. The DARPA robotics challenge. Available at: <http://www.theroboticschallenge.org/>. Accessed March 30, 2014.
- The DARPA robotics challenge. Available at: <http://www.theroboticschallenge.org/>. Accessed March 30, 2014.
- Ozguner U, Stiller C, and Redmill K (2007). Systems for safety and autonomous behavior in cars: the DARPA Grand Challenge experience. *Proc IEEE* **95**, 397–412.
- Parida SK and Kaufmann SH (2010). The quest for biomarkers in tuberculosis. *Drug Discov Today* **15**, 148–157.
- Lakhani KR, Boudreau KJ, Loh PR, Backstrom L, Baldwin C, Lonstein E, Lydon M, MacCormack A, Arnaout RA, and Guinan EC (2013). Prize-based contests can provide solutions to computational biology problems. *Nat Biotechnol* **31**, 108–111.
- Freymann JB, Kirby JS, Perry JH, Clunie DA, and Jaffe CC (2012). Image data sharing for biomedical research—meeting HIPAA requirements for de-identification. *J Digit Imaging* **25**, 14–24.
- Radiological Society of North America. The RSNA Clinical Trial Processor. 2012. Available at: http://mirwiki.rsna.org/index.php?title=CTP-The_RSNA_Clinical_Trial_Processor: Radiological Society of North America. Accessed January 13, 2014.
- Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, and Pringle M (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* **26**, 1045–1057.
- National Cancer Institute's Center for Biomedical Informatics and Information Technology. National Biomedical Imaging Archive. 2011. Available at: <https://imaging.nci.nih.gov/ncia/login.jsf>.
- Channin DS, Mongkolwat P, Kleper V, Sepukar K, and Rubin DL (2010). The caBIG Annotation and Image Markup project. *J Digit Imaging* **23**, 217–225.
- Atlassian. Confluence. 2014. Available at: <http://www.atlassian.com/software/confluence>. Accessed January 13, 2014.
- McNitt-Gray MF, Armato SG III, Meyer CR, Reeves AP, McLennan G, Pais RC, Freymann J, Brown MS, Engelman RM, and Bland PH (2007). The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation. *Acad Radiol* **14**, 1464.
- Armato SG III, Meyer CR, McNitt-Gray MF, McLennan G, Reeves A, Croft BY, and Clarke LP (2008). RIDER Research Group The Reference Image Database to Evaluate Response to therapy in lung cancer (RIDER) project: a resource for the development of change-analysis software. *Clin Pharmacol Ther* **84**, 448–456.
- Kalpathy-Cramer J, Zhao B, Goldof D, Gu Y, Wang X, Gillies R, Yang H, Tan Y, and Napel S (2013). A platform for the comparison of lung nodule segmentation algorithms: methods and preliminary results. In *Radiological Society of North America (RSNA) 99th Scientific Assembly and Annual Meeting, Chicago, IL, December, 2013*.
- Huang W, Li X, Chen Y, Li X, Chang M-C, Oborski MJ, Malyarenko DI, Muzi M, Jajamovich GH, Fedorov A, et al. (2014). Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge. *Transl Oncol* **7**, 153–166.
- Menze B, Reyes M, Jakab A, Gerstner E, Kirby J, and Farahani K (2013). MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS). In *Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2013*. Available at: http://martinos.org/qtim/miccai2013/proc_brats_2013.pdf. Accessed January 31, 2014.
- The Cancer Imaging Archive. NCI-ISBI 2013 Challenge - Automated Segmentation of Prostate Structures. 2014. Available at: <https://wiki.cancerimagingarchive.net/display/Public/NCI-ISBI+2013+Challenge+++Automated+Segmentation+of+Prostate+Structures>. Accessed January 13, 2014.
- The Cancer Imaging Archive. NSCLC Radiogenomics. Available at: <https://wiki.cancerimagingarchive.net/display/Public/NSCLC+Radiogenomics>. Accessed January 13, 2014.
- Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, Rubin DL, Napel S, and Plevritis SK (2012). Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. *Radiology* **264**, 387–396.
- Fedorov A, Fluckiger J, Ayers GD, Li X, Gupta SN, Tempny C, Mulkern R, Yankeelov TE, and Fennessy FM (2014). A comparison of two methods for

- estimating DCE-MRI parameters *via* individual and cohort based AIFs in prostate cancer: a step towards practical implementation. *Magn Reson Imaging* **32**(4), 321–329.
- [30] Makki SK and Sangtani J (2008). Data mashups & their applications in enterprises. In SK Makki, and J Sangtan (Eds.), *Internet and Web Applications and Services, 2008. ICIW'08. Third International Conference on: IEEE* pp. 445–450.
- [31] Sharma A and Saghar YN (2013). Project Bindaas. Available at: <http://imaging.cci.emory.edu/wiki/display/BDS/Downloads>. Accessed January 31, 2014.
- [32] Fielding RT (2000). *Architectural Styles and the Design of Network-Based Software Architectures*. University of California, Irvine.
- [33] Gutman DA, Cooper LA, Hwang SN, Holder CA, Gao J, Aurora TD, Dunn WD, Scarpace L, Mikkelsen T, and Jain R (2013). MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. *Radiology* **267**, 560–569.
- [34] McLennan M and Kennell R (2010). HUBzero: a platform for dissemination and collaboration in computational science and engineering. *Comput Sci Eng* **12**, 48–53.
- [35] Luo XZ, Kennedy DN, and Cohen Z (2009). Neuroimaging informatics tools and resources clearinghouse (NITRC) resource announcement. *Neuroinformatics* **7**, 55–56.
- [36] Kennedy DN and Haselgrove C. The three NITRC's: software, data and cloud computing for brain science and cancer imaging research. *Front Neuroinform.* Conference Abstract: Neuroinformatics 2013, Stockholm, Sweden, 27 Aug–29 Aug, 2013. DOI:10.3389/conf.fninf.2013.09.00024
- [37] Bell G, Gray J, and Szalay A (2006). Petascale computational systems. *Computer* **39**, 110–112.