# An Assessment of Imaging Informatics for Precision Medicine in Cancer

C. Chennubhotla[1], L. P. Clarke[2], A. Fedorov[3], D. Foran[4], G. Harris[5], E. Helton[6], R. Nordstrom[2], F. Prior[7], D. Rubin[8], J. H. Saltz[9], E. Shalley[2], A. Sharma[10]

[1]  Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, USA
[2]  Cancer Imaging Program, NCI, NIH, Bethesda, MD, USA
[3]  Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[4]  Department of Pathology and Laboratory Medicine, Rutgers Robert Wood Johnson Medical School, Rutgers, New Brunswick, NJ, USA
[5]  Harvard Medical School, Boston, MA, USA
[6]  Center for Biomedical Informatics and Information Technology, NCI, NIH, Bethesda, MD, USA
[7]  Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA
[8]  Department of Radiology, Stanford University, Palo Alto, CA, USA
[9]  Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY, USA
[10] Department of Biomedical Informatics, Emory University, Atlanta, GA, USA

## Summary

**Objectives:** Precision medicine requires the measurement, quantification, and cataloging of medical characteristics to identify the most effective medical intervention. However, the amount of available data exceeds our current capacity to extract meaningful information. We examine the informatics needs to achieve precision medicine from the perspective of quantitative imaging and oncology.

**Methods:** The National Cancer Institute (NCI) organized several workshops on the topic of medical imaging and precision medicine. The observations and recommendations are summarized herein.

**Results:** Recommendations include: use of standards in data collection and clinical correlates to promote interoperability; data sharing and validation of imaging tools; clinician's feedback in all phases of research and development; use of open-source architecture to encourage reproducibility and reusability; use of challenges which simulate real-world situations to incentivize innovation; partnership with industry to facilitate commercialization; and education in academic communities regarding the challenges involved with translation of technology from the research domain to clinical utility and the benefits of doing so.

**Conclusions:** This article provides a survey of the role and priorities for imaging informatics to help advance quantitative imaging in the era of precision medicine. While these recommendations were drawn from oncology, they are relevant and applicable to other clinical domains where imaging aids precision medicine.

## Introduction

Precision medicine [1] requires the measurement, quantification, and cataloging of medical characteristics to identify the most effective medical intervention. The National Academy of Sciences defines precision medicine as "*the tailoring of medical treatment to the individual characteristics of each patient. It does not literally mean the creation of drugs or medical devices that are unique to a patient, but rather the ability to classify individuals into subpopulations that differ in their susceptibility to a particular disease, in the biology and/or prognosis of those diseases they may develop, or in their response to a specific treatment*" [2]. In other words, through precision medicine we can classify patients into cohorts that share characteristics such as diagnosis, prognosis, response to a certain therapy etc. This requires ready access to networks of data that can be queried using many different types of search criteria across many different types of data. And to create such classifiers, large quantities of diverse data must be accessed, analyzed, and reduced to actionable knowledge for patients and encounters.

Imaging, which includes radiology, radiation oncology, and pathology, complements clinical and molecular data and offers crucial insights that help stratify patients into cohorts and guide care using the principles of precision medicine [3-7]. In addition to diagnosis and treatment planning, imaging also has the potential to provide deep and novel insights by evaluating a patient's response to therapy during treatment, as well as predicting outcome at an earlier time point [6-8]. Treatment response and early outcome prediction thus create opportunities for adaptive medicine. For example, in breast cancer patients with ER+, PR+, and HER2– invasive ductal carcinoma MRI-based features (texture and morphological) could predict the likelihood of recurrence and magnitude of chemotherapy benefit [9]. Clustering morphological signatures extracted from digitized whole-slide pathology images of

glioblastoma patients helped identify significant prognostic sub-classifications, in which clusters are correlated with transcriptional, genetic, and epigenetic events [10]. Imaging can also be used for surveillance in certain low-grade cancers and help avoid unnecessary biopsies. In such patients, imaging features can also be used to identify sub-populations that are likely to advance to higher-grades, and would thus be candidates for specific treatments [11, 12]. These four examples, from breast cancer, glioblastoma, low-grade glioma, and prostate cancer, are illustrative of the role imaging can play in precision medicine for cancer. A more detailed survey of imaging and its role in precision medicine was recently reviewed by the Association of University Radiologists Radiology Research Alliance [13]. This survey, in many ways, complements the review by providing a survey of challenges, and an assessment of the needs of imaging informatics, for the advancement of precision medicine in cancer.

Today, imaging is predominantly digital; however, image interpretation and its use in diagnosis and treatment assessment have remained largely qualitative. This has been changing steadily through initiatives such as the Quantitative Imaging Biomarker Alliance (QIBA) [14-16], as well as research programs such as the Quantitative Imaging Network (QIN) [16, 17]. This article surveys the landscape of quantitative imaging, its role in advancing precision medicine, some of the informatics priorities, and challenges, and presents some recommendations. It should be noted, that while this article focuses on cancer imaging, the underlying needs and challenges are by no means unique to cancer. Domains such as neurology or cardiology have similar characteristics and requirements. For example, in neuroimaging, imaging data is closely linked with observational data or connectome data. These data types are not typically seen in oncology. However, both oncology and non-oncological conditions share the pattern where the integration of imaging and associated/derived non-imaging data with clinical and genomic data can be used to classify patient populations based on diagnosis and response to treatment [6, 7]. For the sake of readability, this article does do not make a distinction between imaging and cancer imaging. The

challenges, approaches, and recommendations surveyed here are applicable across the broad landscape of quantitative imaging and its application to precision medicine.

It is also worth emphasizing that while much of this work was predominantly geared in recent years towards radiology, there has been a steady increase in the research and development of similar techniques in pathology. Cancer diagnosis is primarily based on pathology; outcome prediction and treatment recommendations are highly dependent on pathologist observations. While digital pathology imaging has lagged behind radiology imaging due to the continuing use of glass slides in clinical diagnostic pathology, the advent of high quality, high throughput, digital scanners has led to the widespread adoption of digital pathology in cancer research studies. DICOM (Digital Imaging and Communications in Medicine), the de facto standard for medical imaging now includes specifications for digital pathology [18, 19]. Digital pathology data management, visualization, and analysis tools have been developed by both research groups and private companies. Work is rapidly progressing on the development of standards for pathology data management, annotation, and markup. Such advances as well as this increased adoption has led to digital pathology sharing with radiology many of the same imaging-based, informatics design patterns for disease classification, patient stratification, response assessment, and outcome prediction.

The inclusion of information obtained from digital pathology is crucial to the success of efforts to improve precision of quantitative imaging-based predictions. Diagnostic and treatment guidelines call for quantitative measurements that are challenging for human observers (*e.g.,* tumor infiltrating lymphocytes, mitoses and immunohistochemistry (IHC) staining). There is also an increasing tendency to mandate detailed assessments of tumor heterogeneity across tumor type. A highly pertinent example is non-small cell lung cancer (NSCLC) adenocarcinoma WHO guidelines that specify that for each patient, pathologists break down sub-type composition in 5% increments. Digital pathology machine-learning methods also promise to reduce inter-observer variability arising from the sole reliance on human-generated pathology classifications.

## Quantitative Imaging, and Informatics Methodologies for Precision Medicine

Quantitative imaging is the process of extracting measurable (numerical) information from images to determine the amount, extent, or severity of disease, where imaging devices behave as standard measurement instruments providing reliable and reproducible numerical results. It has benefitted from advances in image acquisition that have led to improvements in quality and resolution of imaging and diversity imaging modalities. Quantitative imaging, through advances in high performance computing and machine learning, has enabled the process of radiomics (extraction and mining of quantitative imaging features) [18-20] and radiogenomics (integrating radiomic features with clinical and molecular data). It is enabling optimized treatments, surveillance, and better prediction of response to treatment, and it offers great promise for precision medicine. Numerous groups have developed methodologies to extract rich collections of imaging features, linked them with clinical outcome and molecular characterizations, and studied their relevance in clinical research [10, 18-87].

Informatics is the practice of information processing and the engineering of information systems, focusing on the collection, classification, storage, retrieval, and dissemination of recorded knowledge. Quantitative imaging, therefore, benefits from the methodologies, tools, and capabilities that are offered by informatics to help convert the information contained in images into actionable knowledge. The combined information that is gathered can be used to enhance individual and population health outcomes, simplify patient care, and improve the quality of clinical workflow. One notable effort in this area is the National Cancer Institute (NCI) Quantitative Imaging Network (QIN). QIN has encouraged the development and validation of quantitative imaging methods for the measurement of tumor response to therapies in clinical trials and routine care.

It is imperative that by incorporating the science of informatics into medical imaging we create a powerful driver for precision

Chennubhotla et al.

medicine activities. Open and regular communication among scientists, clinicians, informatics specialists, and regulatory experts is needed. We discuss some of the highest priority informatics activities that will help bring quantitative imaging into the realm of clinical decision support. These priority activities were identified through a series of workshops, the first of which was convened in October 2015, to discuss the joint roles of quantitative imaging and informatics within the context of present and future precision medicine needs. The primary needs and challenges, as well as some representative efforts in areas, are presented here.

## N1  Curated Data Repositories:

*Access to well-curated image repositories with support for semantically integrated datasets and the ability to integrate information across type and scale are critical.*

Archiving vast amounts of existing data is a major informatics effort, not only because of the rapid growth in the volume of data to be stored, but also because of the challenges in accessing, retrieving, analyzing, and displaying the results. Imaging datasets require hundreds of terabytes to petabytes of storage. Imaging features, produced by a pipeline, can depend on a variety of parameters, leading to an explosion of post-processing feature data. Meaningful comparison and subsequent downstream use of the imaging features necessitate a standardized representation. Additionally, clinical information about each patient should be linked to the image data. The linked clinical data must be available when searching for sub-populations. It is therefore imperative that the research community stops reinventing the wheel in the context of imaging biomarker development and comes up with common ways to share tools and data to help improve interoperability. There are three types of data that researchers and practitioners of quantitative imaging share, namely:

- **Clinical Data**: Clinical data includes demographic information, diagnosis, exposure, family history, treatment, and outcomes data. Clinical data must be harmonized against a common vocabulary. This is an active area of research. One possible direction is the use of DICOM

to represent clinical data [88]. This is an attractive proposition given the near-universal acceptance of DICOM, especially in the clinical domain. However, the DICOM specification merely provides a data representation format. Work is needed to create an ontology that can be used to encode the data. An example of such an ontology that helps encode clinical data in DICOM has been developed for Head and Neck cancers [88]. Another option that is being explored is the use of the clinical data model used in The Cancer Genome Atlas (TCGA) [89]. The TCGA clinical data model includes site-specific terms, with mappings to the NCI Thesaurus, and would provide the ability to create image cohorts that span different imaging studies. The TCGA clinical data model has also been adopted by The Genomic Data Commons (GDC) [90, 91].

- **Images and Image Metadata**: This includes the raw pixels as well as metadata that describe the image such as patient level information, acquisition data, etc. Image metadata is frequently stored in DICOM formats and follows the DICOM information hierarchy. Other formats such as NIfTI (Neuroimaging Informatics Technology Initiative) are also widely used [92]. DICOM is not widely used in digital pathology, since most digital pathology scanner vendors prefer their own formats. There are however open source libraries, such as OpenSlide [93], that allow researchers to interact with these images using a shared library and application programming interfaces (APIs).

- **Image Annotations and Features**: These include human and machine generated annotations and features. QIN agreed to adopt DICOM as a standard for images and segmentation maps. The term "features" refers to the quantitative characteristics extracted from images; these are represented in various open formats. Since imaging features are at the core of quantitative imaging, a detailed description about their representation and storage is presented separately (see N3 & N4).

**Data Curation:** The data used in the development of imaging-based methodologies for precision medicine must be well curated

to reduce any uncertainty in its history or content. The use of standards such as DICOM is therefore essential. In addition to standards, the data management system, as well as the processes of data curation are very important. In recent years, several imaging repositories have come online line [92, 94-99], with The Cancer Imaging Archive (TCIA) being an exemplar of a well curated, diverse, imaging repository. Since its inception in 2011, TCIA has evolved into NCI's primary resource for curating, managing, and distributing images. A significant component of TCIA operations and tools involves the curation and de-identification processes. With the adoption of data standards, as well as the deployment of easy to use tools and shared best practices, the process of data preparation and submission is greatly simplified. This results in reducing the burden of data sharing, and in faster submission and quicker dissemination of data. This has the added advantage of encouraging data sharing, since the burden of data sharing is frequently cited as one of the common roadblocks to data sharing.

**De-identification:** A key component of data curation is having well-documented processes and tools that facilitate the de-identification of data and the removal of any patient identifiers. It is often assumed that de-identification involves the scrubbing of protected health information (PHI) from DICOM headers. However, in practice, it has been observed that scanners frequently encode identifiable information in private DICOM tags. There are additional challenges with de-identification when dealing with time-series data. In such situations, the chosen heuristic for date de-identification must be cognizant of time elapsed between successive studies. This information allows users to run queries such as: "*find all lung screening studies where 3 or more studies were performed, and each study was within 6 months of the prior study.*" In addition to maintaining the elapsed time, users may want the ability to integrate these imaging studies with other non-imaging data. Therefore, enough metadata must be preserved to ensure compliance with appropriate rules and regulations, while ensuring that researchers can unambiguously locate imaging and associated non-imaging data [100].

## N2 Data Exploration, Access, and Integration

*Exploring and accessing images along with associated data is critical to research in the era of Precision Medicine.*

**Data Exploration and Access**: How data is managed is critical when it comes to versatility and ease of use of the data. While methods of storage are important for creating useful and minable image troves, efficient content-based methods of data retrieval may be even more essential to making these data accessible and usable. Search engines capable of returning image data along with all appropriate metadata are necessary assets in the era of large electronic datasets. It is necessary for data associated with the images (see N1) to be accessible and integrated with image data. It is not practical, nor are we advocating for a centralized meta repository that manages all data types. Rather a mechanism that allows one to use some of the data types to create a cohort and then access images and relevant data directly from the individual repositories that manage that data.

**Data Integration**: An example of this would be an integration of TCIA with the Genomic Data Commons (GDC). It would give researchers the ability to create a cohort using genomic, clinical, and imaging attributes, and then access the images and genomic data for the identified cases. One popular and decentralized approach to achieving such integration is through the adoption of REST APIs. Our goal should be adoption of APIs and when appropriate, convergence on shared API specifications. The underlying API implementations are best left to the repositories, and will have minimal impact on facilitating an integrated exploration and retrieval of data.

The development and adoption of an API economy has the added advantage of encouraging developers to directly and programmatically integrate with the various data repositories. Doing so allows, for example, an image analysis algorithm to directly retrieve images from TCIA without requiring that a user first download a dataset, then upload it to a local cluster, and finally launch the algorithm on this dataset. Integration via APIs allows one to run large, cloud-based, pipelines that can exploit the cost and scale benefits of clouds (See N6). Similarly, research worksta-

tions like 3DSlicer [101] can directly integrate with, and utilize the search and retrieval capabilities of image repositories, giving their users an optimized experience.

## N3 Algorithm Validation and Reproducibility

In addition to the reduction of hardware errors in data collection, quantitative imaging deals with the development and optimization of robust algorithms capable of extracting useful information from the collected images. The potential of quantitative imaging can only be realized if the algorithms are reproducible and validated. These algorithms are individually designed to serve specific functions in a chain of analysis, that begins with the collected images, and ends with the extracted quantitative information. Functions such as segmenting suspicious regions in the image and then processing the information within those regions for information correlative with disease are included. Quantitative methods differ in how much information is extracted and used, and in how the information is assembled for dissemination. In addition, as there are many different imaging systems in use, the information extracted must be available in a standardized form that can be read and interpreted across multiple devices. Using this information, researchers can then generate new diagnostic and prognostic techniques. One notable mechanism for advancing these goals are Grand Challenges that have proven to be a successful means for promoting innovation in algorithm development [102].

**Feature Generation**: Imaging features cover the gamut of tumor segmentations, observations, and features captured by humans, as well as features that are computed by algorithms. They include qualitative, quantitative, and mixed features. The underlying objective is to capture a set of features that can act as numeric surrogates for an image and can then be used to explore correlations with clinical or genomic data. They could be used to train classifiers that can guide diagnosis, prognosis, or response to therapy. For example, Aerts et. al., extract ~400 features from CT and MRI images of lung cancer, and head and neck cancer pa-

tients, and identify feature signatures that are strong predictors of outcome [18]. Features here included morphological features, tumor intensity, texture, and other higher-level features. In digital pathology, a similar process is followed, leading to the coinage of the term pathomics. An illustrative example of pathomics is the work done by Cooper et. al., where they processed glioblastoma images and extracted 74 different features [103], that included morphological characterizations of nuclei, nuclear intensity, texture, and gradient statistics. These features were extracted from 200M nuclei and revealed three prognostically significant clusters with associations to genetic mutations and outcomes [10]. A similar study was done by Huang et. al. on breast cancer images [104].

**Deep Learning and Medical Imaging**: In recent years, there has been a strong interest in the application of neural networks and deep learning for quantitative imaging. These methodologies have been around for a long time, and as far back as the early 90s, during the early days of digital imaging, they were used in a variety of applications, such as the detection of lung nodules [105, 106], and classification of regions of interest (ROIs) from mammograms as benign or malignant [107]. However, it was not until very recently that deep learning gained popularity and has emerged as one of the most promising tools for image classification. This popularity is driven, not only by advances in algorithms, but also, in large part, by advances in high performance computing (HPC), incl. graphics processing units (GPUs), and the fact that the cost of these HPC systems have come down significantly. Since these algorithms are easily parallelizable, they can take advantage of the inherent parallelization of GPUs. A comprehensive survey, as well as a series of articles, covering the use of deep learning and medical imaging can be found in an IEEE special issue on medical imaging and deep learning edited by Greenspan et. al. [108]. There are however a set of challenges that have slowed down the success of deep learning, the biggest one being access to large quantities of annotated data sets (see N1). Data needs to be well annotated (N4), and researchers should have the ability to integrate the extracted 'hidden' features with clinical and/or genomic data (N2).

**Algorithm Validation**: The performance characteristics of the algorithms must eventually be tested and validated under a variety of clinically relevant conditions before they can be useful in clinical workflow. This often requires the use of large datasets of clinical images as an environment in which to test the performance and robustness of the algorithms. Ideally, metadata such as annotations, clinical information from patient history, and patient outcomes will be a part of the information included in algorithm validation. The need for informatics in this process is critical and is integral to the process not only through the function of the final quantitative algorithm as it performs in clinical workflow, but also during the degree of required testing and validation needed to ensure algorithm performance before it reaches the clinical setting.

**Role of Grand Challenges for Algorithm Validation and Reproducibility**: Grand challenges have proven to be very successful in helping with the development and validation of novel and innovative algorithms such as brain tumor segmentation [109]. They are a good way to crowdsource the annotation of data. This results in an annotated data set which is critical to the advancement of quantitative imaging through deep learning. Grand challenges explicitly encourage open science and open source, best-of-breed algorithms. They do so, by presenting informaticians with specific problems, constraints, and incentives for innovation [110]. They also address reproducibility and integration by providing access to clinical and -omics data that are not always readily available and can help move a tool to readiness for the commercialization pipeline. Grand challenges are conducted in specially-designed environments that import specific data sets from selected archives, for training and testing on the algorithms, and provide comparisons of results and performance. However, they need to be organized in a manner that simulates clinical workflows and other real-world constraints (*e.g.,* computational pathology challenges must use whole slide images (WSIs) at 40x magnification or higher depending on the task, or test datasets that are noisy and need cleaning). They also need to be organized in a manner where the submissions are run on the test data by organizers, thereby fostering a culture of tool reproducibility (more details in N5).

## N4 Representing and Managing Quantitative Image Features:

*Effective curation of, and integration with, imaging features is critical to the realization of the potential of quantitative imaging. Non-proprietary annotation and markup will allow for cross-hardware compatibility, interoperability, and sharing of data from many sources and between institutions.*

**Feature Representation**: Integrating radiology, pathology, clinical, and -omics data requires that image annotations be stored in a standardized and interoperable manner. One example of image annotations is the segmentation of the image regions corresponding to the tissues or objects of interest. Such annotations can be displayed during image viewing, can be used to extract quantitative measures from the image (*e.g.,* tumor volume, vessel permeability), and can capture aspects of key regions within images that are meaningful to the radiologist and oncologist. For example, image annotations can record the location and measurements of target lesions or point out non-target lesions. Frequently, the annotations, created on commercial image viewing workstations, are collected and stored in either proprietary formats or as DICOM presentation state objects, which are like graphical overlay objects. This enables rendering the information visually, but does not support search of, and access to the annotations, nor any computation on them. One is therefore forced to rely on vendor-specific implementations and software. Even if one were to use vendor-specific software, these software tools are often closed, and do not adopt standards for annotation, thus hindering interoperability. Consequently, all annotations currently must be created and maintained within siloed commercial applications, and there is no interoperability of image annotations across platforms and applications. To realize the potential value of integrative radiology-pathology-omics, it is vital that image annotations be stored in standardized interoperable formats such as the Annotation and Image Markup (AIM) standard or DICOM; a harmonization effort is underway to unify these two standards.

The goal of the AIM project [111] is to provide a standardized, interoperable mechanism for modeling, capturing, and serializing image annotation and markup data that would be adopted widely within the medical imaging community. Both human- and machine-read-able artifacts are possible. The variability in methods of storing annotations with the image data is a concern that can be addressed by developing standard DICOM objects to store this information. DICOM Working Group 8 is working to harmonize and unify the AIM and DICOM standards and create a DICOM Structured Reporting object to store AIM image annotations. When adopted by commercial platforms, this will provide a standardized interoperable format for image annotations. Adopting this as the standard format to store image annotations will streamline software development and enable the work to focus on providing rich annotation features and functionality and on amassing a large collection of minable image data. Designing the tools to be compatible with other standards will enable a high degree of interoperability and the incorporation of the annotation standard into commercial, clinical, information systems.

**Data Visualization**: Image viewing platforms that support AIM/DICOM-SR will permit consuming annotations from a variety of sources and linking them to other types of image data as well as non-image data. Moreover, large collections of image data will become "minable" to enable discovery from historical collections of Radiology/Pathology image data. Such activities will be particularly important in cooperative groups, who routinely collect and store large amounts of such image data and annotations during clinical trials.

## N5 Scaling Quantitative Imaging via Container and Cloud Deployments:

*Novel container technologies will allow for portability and interoperability, critical to sharing algorithms in a distributed research environment. Increasing adoption of cloud environments will allow researchers to compute and process at significantly larger scales.*

Advances in systems software such as containers provide the ability to encapsulate algorithms, and their implementations, thus enhancing reuse and portability [89, 112]. Containers, popularized by Docker, make it possible for researchers to share their algorithms and pipelines in a robust and self-contained fashion. These systems integrate nicely with modern distributed version control system, thereby greatly simplifying

the deployment of data processing codes. Additionally, in instances where investigators are unable to share source code, containers give them the ability to create images that are equivalent to platform-agnostic, binary executables of their data processing codes.

In recent years, cloud computing has become much more popular within the research community. This increased interest has been spurred, in part by the launch of the NCI supported, genomics cloud pilots [113]. These cloud pilots are now serving as exemplars that allow researchers to perform genomic studies on the cloud, without having to first download large quantities of genomic data and then upload them to institutional clusters for processing and analysis. The adoption of containers ease this migration by greatly simplifying the complexity of deploying diverse code-bases on a single cloud [91, 114].

The imaging community should consider these technologies as a means of sharing methods and tools. Some key issues in this area are still to be addressed. The cost of processing on the cloud is still high, though this is being addressed through the recent launch of the NCI Commons Credit Pilot [115].

## Open Standards and Open Source Architecture:

These enable flexible and more rapid technology developments, which are reproducible and are more likely to see an accelerated adoption in the marketplace.

Open source refers to software that is accompanied by its source code and is made available through a license which allows users to change and re-distribute the software under the conditions stipulated by the license. Different flavors of open source licenses exist [112]. Examples include GNU Public License (GPL) that limits commercial use of the source code, and MIT or FreeBSD licenses that do not limit modification and reuse of the source code by anyone and for any purpose, including commercialization. There are numerous examples of software tools developed within NCI-supported programs that are being made available as open source. One example is ePAD [113, 115], a quantitative imaging informatics platform that provides web-based access to AIM-compliant metadata and semantic image annotation on any platform and any image

workstation. Another open source solution is LIBRA [95, 116], a software package developed at the University of Pennsylvania that is a fully-automated breast density estimation solution based on a published algorithm that works on both raw and vendor post-processed digital mammography images. The DICOM Toolkit (DCMTK) is another example of openly available software [117]. DCMTK is a collection of libraries and applications that implement large components of the DICOM standard, including software for examining, constructing, and converting DICOM image files, handling offline media, and sending and receiving images over a network connection.

For the developer of quantitative imaging algorithms, whether for data collection or image analysis, the use of open source software as modules or components in the total algorithm package can be a shortcut to success. Open source development has seen a significant growth and transformation with the release of git [118] (a distributed version control systems) and github.com (a publically accessible, centralized, hosted git service). A commitment on the part of the developer to use a modular, open-source architecture, encourage reuse, thereby introducing efficiencies in algorithm development is required. A significant development is the widespread use of containerization platforms such as Docker [119, 120] and related projects, which are enabling more broad dissemination of methods through facile packaging and execution of algorithms. In other words, the inherent flexibility in open source programming permits the programmer to focus on building custom interfaces, to create new capabilities, and to customize the performance of the overall algorithm. It also allows for parallel development on independent components. Importantly, open source development is critical for community building and a continuity of the development that might be more tolerant to the interruptions in funding or fluctuations in the personnel at individual academic labs.

Innovation is important to science, but we also need to balance that innovation with pragmatism, developing what researchers need today and what can facilitate progress in steps. We can learn from the success of communities such as DICOM and The Biomedical Research Integrated Domain Group (BRIDG) [121] to make sure what we develop resonates with

research communities. If we do not take this approach, reproducibility of research results and outputs, which is critical to scientific research, will never be a reality. Additionally, there is an urgency to demand and reward the sharing of both imaging data and data analysis results to enable secondary analysis, support reproducibility of findings, and to allow aggregation of standardized datasets. These datasets can include radiological images, digital pathology, immunohistochemistry, and data from other modalities that can be standardized and integrated for analysis. Efforts such as the Informatics Technology for Cancer Research (ITCR) Program [101], which is funding the development of open source tools and algorithms, have been very successful in generating interest and engagement with the imaging community. Over a dozen tools that support visualization, storage, and analysis have been developed from ITCR funding.

## Bringing Quantitative Imaging into the Clinical Workflow

For quantitative imaging to become a part of precision medicine, it is critical that images and features connect with other diagnostic approaches. Genomics, for example, is receiving a great deal of scientific focus for its ability to chart the progression of disease and to unlock the molecular basis for cancer. Radiomics and radiogenomics are creating a culture change in imaging and in the use of informatics to predict patient outcome. Showing the benefit in combining genomic information with subtle imaging results to gain greater insight into cancer progression is important to speed adoption of imaging methods and incorporate them into the clinical workflow.

Educating clinicians on the benefits of imaging methods in clinical practice is key to their adoption. For example, morphologic diagnosis is required in many cases, as genomic analysis alone is sometimes inadequate. Genomic analysis will not reveal carcinoma versus benign growth and mutations analyses alone cannot provide a specific diagnosis. For example, in the case of Leiomyoma (benign disease) vs. Leiomyosarcoma (cancer), the genetic mutation is the same, but human cognition and the use of microscopes are

required to accurately diagnose cancer versus benign growth, although novel deep learning techniques may be helpful in aiding differential diagnosis in the future. Spatial phenotypic heterogeneity is not captured by genomic data. There is no way of understanding interactions between the various cell types in a tumor microenvironment (TME). If the cell composition is the same, but the interactions are different, in two different TMEs, genomics cannot tell them apart. Hence, the study of images, and their spatial data, is crucial.

**A Case Study - The Open Health Imaging Foundation at Dana Farber / Harvard Cancer Center:** One of the fundamental drivers for an integrated cancer imaging informatics infrastructure is the ability to easily view and share images across sites and modalities, and provide a standards-based platform and plug-in architecture for developers. There are many proprietary commercial web-viewers on the market, which are not easily customized or open to collaborative development, and those systems that are open are not typically of a professional grade that would allow translation and collaboration between academics and industry.

The Tumor Imaging Metrics Core (TIMC) at Harvard created the Open Health Imaging Foundation (OHIF; http://ohif.org). OHIF supports open-source, web-based, imaging technologies, and is building a vendor-neutral, open source, extensible, zero-footprint web-viewer and supporting server for display and analysis of DICOM images. The platform is designed with a plug-in architecture to allow the group to integrate this web-viewer with oncology applications across the cancer research community.

One use case of this zero-footprint web-viewer is the replacement of the group's existing thick-client with an open source image workstation from the Precision Imaging Metrics [116] clinical trials management system. The system was developed by the Dana Farber / Harvard Cancer Center (DF/HCC) TIMC and is presently in use across six NCI-designated Cancer Centers. To make the system broadly available to the oncology research community, the team is developing an interface to the TIMC's Precision Imaging Metrics web-based application, and implementing an annotation and overlay standards-compatible interface. The group has been actively working with investigators from several other NCI-funded projects to integrate their viewer with other oncology research platforms. The viewer will meet all the basic requirements for radiology tumor measurements specific to the needs of oncology clinical trials, yet also be flexible enough to be configured for user preferences and extended via plug-ins to support varied research workflows as a shared research resource. To achieve these design goals, the viewer and all its functionality will be delivered to client machines exclusively through the web browser requiring nothing to install on client computers or mobile devices, which greatly simplifies and reduces the cost and support requirements of software deployments, and increases accessibility. The proposed viewer will enable researchers, imaging software developers, clinicians, and patients to access oncology clinical trials images in a freely available and openly extensible environment. This will facilitate remote image viewing and collaborative image consultations among a wide-range of imaging professionals.

# On-Going and Future Initiatives

There are numerous ongoing initiatives that help advance the integration of imaging into the clinical and research workflow. We summarize a few of these that cover a cross-section of research and clinical use cases, ranging from those that enable imaging-based epidemiologic studies, to others that advance the quality and reproducibility of imaging algorithms, to a few that enable the management and processing of imaging data at large scales. This is by no means a comprehensive list, rather a sampling of informatics projects with a shared theme; one that includes a focus on quantitative imaging, informatics methodologies, and a specific facet of precision medicine.

**Development of a Cancer Imaging Commons:** The Blue Ribbon Panel working as part of the Vice President's Cancer Initiative (the Moonshot) made several recommendations [122], including the creation of a National Cancer Data Ecosystem. This ecosystem will comprise several commons, like the Genomic Data Commons [123], and include an Imaging Commons. The concept of a commons includes the data, compute, and analytical tools residing in one place, presumably in the cloud, for easy access and computation by researchers. The task of getting TCIA data into the CGC Pilots is a first step towards the development of a Cancer Imaging Data Commons.

**Virtual Tissue Repository (VTR):** The NCI Surveillance, Epidemiology, and End Results (SEER) program is working with participating registries to create the VTR, which will allow researchers to select cases and request that the tumor registries gather tissue, slides, and images, and generate Pathology imaging features and/or additional information. A pilot VTR is using caMicroscope [124] for online viewing of digital pathology images and will employ ITCR tools to carry out analyses on Pathology imaging features along with an integrative query system.

**Prototype Data Harmonization and Integration Project:** Using an NCI Early Detection Research Network (EDRN) breast cancer study containing clinical trial data, in-vivo images, pathology images, and biomarker images, the aim is to build the informatics connections between imaging and clinical data based on ISO standards.

**Imaging and Cloud Computing**: The Cancer Genomics Cloud (CGC) Pilots, funded by NCI, has been launched in 2016. The three platforms provide access to genomic data in combination with clinical data from The Cancer Genome Atlas. In 2017, these cloud pilots expanded their scope by incorporating proteomic data and imaging data to allow for cross-domain analysis. Such collocation of data simplifies data access (N1) and facilitates an integrated exploration of data (N3). CGCs all rely on containerized applications (N5 & N6). Thus, they meet many of the informatics needs that are outlined here. Recent NCI initiatives, encouraging the use of these resources in research activities, could provide a real-world assessment of the informatics needs that were identified in this paper, and help develop a road map for the advancement of quantitative imaging in clinical and research settings.

## Conclusion

This article provides a survey of the role and priorities for imaging informatics to help advance quantitative imaging in the era of precision medicine. It came about from a series of workshops, and dialogues between NCI staff and the academic and industrial scientists involved with imaging informatics. The community continues its work through various initiatives, and ongoing dialogues on the subject, and working to translate informatics developments to clinical utility as rapidly as possible. In addition to the six needs and challenges outlined above, there are some other broad recommendations, listed below:

- **Ensure buy-in from clinicians** and make sure the tools developed will work in the clinical workflow. **Educate clinicians** on the value of imaging and its potential contribution to diagnosis, guiding treatment plans, and scientific research.
- **Incentivize and reward sharing** of both the imaging data and the data analysis results to enable secondary analysis, support reproducibility of findings, and to allow aggregation of standardized datasets.
- Create solutions that **ensure data quality and veracity**, for ease of retrieval and clinical utility.
- Work towards a **flexible, extensible, integrated framework** but not a single, monolithic platform; encourage APIs, agile data management systems, and use of standards and semantics for interoperability.
- Work within organizations to **educate tech transfer and legal departments** on the importance of industry partners and to set reasonable expectations for such partnerships.

### Acknowledgements

## References

1. Ashley EA. The precision medicine initiative: a new national effort. JAMA 2015;313(21):2119-20.
2. National Research Council. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington (DC); 2011
3. Hsu W, Markey MK, Wang MD. Biomedical imaging informatics in the era of precision medicine: progress, challenges, and opportunities. J Am Med Inform Assoc 2013;20(6):1010-3.
4. Yankeelov TE, Mankoff DA, Schwartz LH, Lieberman FS, Buatti JM, Mountz JM, et al. (2016) Quantitative Imaging in Cancer Clinical Trials. Clin Cancer Res 2016;22(2):284-90.
5. NRC. Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease. National Academies Press (US); 2011.
6. Thrall JH. Moreton Lecture: Imaging in the Age of Precision Medicine. J Am Coll Radiol 2015;12(10):1106-11.
7. Herold CJ, Lewin JS, Wibmer AG, Thrall JH, Krestin GP, Dixon AK, et al. Imaging in the Age of Precision Medicine: Summary of the Proceedings of the 10th Biannual Symposium of the International Society for Strategic Studies in Radiology. Radiology 2016;279(1):226-38.
8. Thrall JH. Personalized medicine. Radiology 2004;231(3):613-6.
9. Sutton EJ, Oh JH, Dashevsky BZ, Veeraraghavan H, Apte AP, Thakur SB, et al. Breast cancer subtype intertumor heterogeneity: MRI-based features predict results of a genomic assay. J Magn Reson Imaging 2015;42(5):1398-406.
10. Cooper LA, Kong J, Gutman DA, Wang F, Gao J, Appin C,et al. Integrated morphologic analysis for the identification and characterization of disease subtypes. J Am Med Inform Assoc 2012;19(2):317-23.
11. de la Fuente MI, Young RJ, Rubel J, Rosenblum M, Tisnado J, Briggs S,et al. Integration of 2-hydroxyglutarate-proton magnetic resonance spectroscopy into clinical practice for disease monitoring in isocitrate dehydrogenase-mutant glioma. Neuro Oncol 2016;18(2):283-90.
12. Siddiqui MM, Truong H, Rais-Bahrami S, Stamatakis L, Logan J, Walton-Diaz A, et al. Clinical implications of a multiparametric magnetic resonance imaging based nomogram applied to prostate cancer active surveillance. J Urol 2015;193(6):1943-9.
13. Giardino A, Gupta S, Olson E, Sepulveda K, Lenchik L, Ivanidze J, et al. Role of Imaging in the Era of Precision Medicine. Acad Radiol 2017;24(5):639-49.
14. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, et al; Group R-QMW. Metrology Standards for Quantitative Imaging Biomarkers. Radiology 2015;277(3):813-25.
15. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, et al; Group QTPW. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 2015;24(1):27-67.
16. Nordstrom RJ. The Quantitative Imaging Network in Precision Medicine. Tomogra-phy2016;2(4):239-41.
17. Mangla R, Singh G, Ziegelitz D, Milano MT, Korones DN, Zhong J, et al. Changes in Relative Cerebral Blood Volume 1 Month after Radiation-Temozolomide Therapy Can Help Predict Overall Survival in Patients with Glioblastoma. Radiology 2010;256(2):575-84.
18. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
19. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. Magn Reson Imaging 2012;30(9):1234-48.
20. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48(4):441-6.
21. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC; Group Ft. Quantitative Imaging Test Approval and Biomarker Qualification: Interrelated but Distinct Activities. Radiology 2011;259(3):875-84.
22. Clarke LP, Croft BS, Nordstrom R, Zhang H, Kelloff G, Tatum J. Quantitative imaging for evaluation of response to cancer therapy. Transl Oncol 2009;2(4):195.
23. Clarke LP, Sriram RD, Schilling LB. Imaging as a biomarker: standards for change measurements in therapy workshop summary. Acad Radiol 2008;15(4):501-30.
24. Eliceiri KW, Berthold MR, Goldberg IG, Ibanez L, Manjunath BS, Martone ME, et al. Biological imaging software tools. Nat Meth 2012;9(7):697-710.
25. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging 2012;30(9):1323-41.
26. Gatenby RA, Grove O, Gillies RJ. Quantitative Imaging in Cancer Evolution and Ecology. Radiology 2013;269(1):8-14.
27. Gillies R. Radiomics: informing cancer heterogeneity. J Nucl Med 2013; 54:31.
28. Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 2009;48(1):63.
29. Hunter L. Radiomics of NSCLC: Quantitative CT Image Feature Characterization and Tumor Shrinkage Prediction. MS. University of Texas Graduate School of Biomedical Sciences at Houston; 2013.
30. Kalpathy-Cramer J, Freymann JB, Kirby JS, Kinahan PE, Prior FW. Quantitative Imaging Network: Data Sharing and Competitive AlgorithmValidation Leveraging The Cancer Imaging Archive. Transl Oncol 2014;7(1):147-52.
31. Kuo MD, Gollub J, Sirlin CB, Ooi C, Chen X. Radiogenomic Analysis to Identify Imaging Phenotypes Associated with Drug Response Gene Expression Programs in Hepatocellular Carcinoma. J Vasc Interv Radiol 2007;18(7):821-30.
32. Rutman AM, Kuo MD. Radiogenomics: Creating a link between molecular diagnostics and diagnostic imaging. Eur J Radiol 2009;70(2):232-41.
33. Sivakumar S, Chandrasekar C. Lung Nodule Detection Using Fuzzy Clustering and Support Vector

Machines. International Journal of Engineering and Technology; 2013;5:179-85.

34. Veiga C, McClelland J, Moinuddin S, Laurenco A, Ricketts K, Modat M, et al. Adaptive radiotherapy for head and neck patients: evaluation of a deformable registration-based "dose of the day" calculation. International Journal of Radiation: Oncology-Biology-Physics; 2013.

35. Waterton JC, Pylkkanen L. Qualification of imaging biomarkers for oncology drug development. E J Cancer 2012;48(4):409-15.

36. Gurcan MN, Pan T, Shimada H, Saltz J. Image analysis for neuroblastoma classification: segmentation of cell nuclei. Conf Proc IEEE Eng Med Biol Soc 2006;1:4844-7.

37. Ross JC, Estepar RS, Diaz A, Westin CF, Kikinis R, Silverman EK, et al. Lung extraction, lobe segmentation and hierarchical region assessment for quantitative analysis on high resolution computed tomography images. Med Image Comput Comput Assist Interv 2009;12(Pt 2):690-8.

38. Sertel O, Kong J, Shimada H, Catalyurek UV, Saltz JH, Gurcan MN. Computer-aided Prognosis of Neuroblastoma on Whole-slide Images: Classification of Stromal Development. Pattern Recognit2009;2(6):1093-103.

39. Gutman DA, Cooper LA, Hwang SN, Holder CA, Gao J, Aurora TD, et al. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. Radiology 2013;267(2):560-9.

40. Velazquez ER, Parmar C, Jermoumi M, Mak RH, van Baardwijk A, Fennessy FM, et al. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. Scientific Reports 2013;3:3529.

41. Colen R, Foster I, Gatenby R, Giger ME, Gillies R, Gutman D, et al. NCI Workshop Report: Clinical and Computational Requirements for Correlating Imaging Phenotypes with Genomics Signatures. Transl Oncol 2014;7(5):556-69.

42. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics feature quantification using semi-automatic volumetric segmentation. PloS One 2014;9(7):e102107.

43. Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RT, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 2015;114(3):345-50.

44. Parmar C, Leijenaar RT, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, et al. Radiomic feature clusters and prognostic signatures specific for Lung and Head & Neck cancer. Sci Rep 2015;5:11044.

45. Basavanhally AN, Ganesan S, Agner S, Monaco JP, Feldman MD, Tomaszewski JE, et al. Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology. IEEE Trans Biomed Eng 2010;57(3):642-53.

46. Chen W, Chu V, Hu J, Yang L, Wang F, Kurc T, et al. ImageMiner: a medical image analysis and image management UML data model. APIII: Advancing Practice, Instruction & Innovation Through Informatics, Pittsburgh, PA; 2009.

47. Cholleti S, Cooper L, Kong J, Chisolm C, Brat D, Gutman D, et al. Classification of brain tumor regions. Boston, MA: Pathology Informatics; 2010.

48. Cooper L, Kong J, Gutman D, Wang F, Cholleti S, Pan T, et al. An Integrative Approach for In Silico Glioma Research. IEEE Trans Biomed Eng 2010;57(10):2617-21.

49. Cooper L, Kong J, Gutman D, Wang F, Cholleti S, Pan T, et al. Integrative Analysis of Image and Molecular Data for Study of Brain Tumors. The NCI-NCRI Informatics Initiative Joint Conference: Biomedical Informatics without Borders: Implementing Interoperability, Bethesda, MD; 2010.

50. Cooper LAD, Carter AB, Farris AB, Wang F, Kong J, Gutman DA, et al. Digital Pathology: Data Intensive Frontier in Medical Imaging. Proc IEEE Inst Electr Electron Eng 2012;100(4):991-1003.

51. Foran DJ, Yang L, Chen W, Hu J, Goodell LA, Reiss M, et al. ImageMiner: a software system for comparative analysis of tissue microarrays using content-based image retrieval, high-performance computing, and grid technology. J Am Med Inform Assoc 2011;18(4):403-15.

52. Fuchs TJ, Buhmann JM. Computational pathology: Challenges and promises for tissue analysis. Comput Med Imaging Grap 2011;35(7):515-30.

53. Gao Y, Tannenbaum A. Combining Atlas and Active Contour for Automatic 3d Medical Image Segmentation. Proc IEEE Int Symp Biomed Imaging 2011;1401-4.

54. Gao Y, Tannenbaum A, Chen H, Torres M, Yoshida E, Yang XF, et al. Automated Skin Segmentation in Ultrasonic Evaluation of Skin Toxicity in Breast Cancer Radiotherapy. Ultrasound Med Biol 2013;39(11):2166-75.

55. Gurcan MN, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B. Histopathological Image Analysis: A Review. IEEE Rev Biomed Eng 2009;2:147-71.

56. Huang K, Mosaliganti K, Cooper LAD, Machiraju R. Quantitative phenotyping using microscopic images. In: Rittscher J, Machiraju R, Wong STC (eds) Microscopic Image Analysis for Life Science Applications. Boston, MA, USA: Artech House; 2008.

57. Huang PW, Lee CH. Automatic classification for pathological prostate images based on fractal analysis. IEEE Trans Med Imaging 2009;28(7):1037-50.

58. Kong J, Sertel O, Shimada H, Boyer K, Saltz J, Gurcan M. Computer-aided grading of neuroblastic differentiation: Multi-resolution and multi-classifier approach. IEEE International Conference on Image Processing 2007;5:V-525.

59. Lu C, Mahmood M, Jha N, Mandal M. Automated Segmentation of the Melanocytes in Skin Histopathological Images. IEEE J Biomed Health Inform 2013;17(2):284-96.

60. Lu C, Mandal M. Automated Segmentation and Analysis of the Epidermis Area in Skin Histopathological Images. 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2012:5355–9.

61. Martone ME, Zhang S, Gupta A, Qian X, He H, Price DL, et al. The cell-centered database: a database for multiscale structural and protein localization data from light and electron microscopy. Neuroinformatics 2003;1(4):379-95.

62. McKeen-Polizzotti L, Henderson KM, Oztan B, Bilgin CC, Yener B, Plopper GE. Quantitative metric profiles capture three-dimensional temporospatial architecture to discriminate cellular functional states. BMC Medical Imaging 2011;11(1):11.

63. Michailovitch O, Rathi Y, Tannenbaum A. Image segmentation using active contours driven by the Bhattacharyya gradient flow. IEEE Trans Image Processing 2007;16:2787-801.

64. Monaco J, Tomaszewski J, Feldman M, Moradi M, Mousavi P, Boag A, et al. Detection of prostate cancer from whole-mount histology images using Markov random fields. Workshop on Microscopic Image Analysis with Applications in Biology (in conjunction with MICCAI); 2008.

65. Oztan B, Kong H, Gürcan MN, Yener B. Follicular lymphoma grading using cell-graphs and multi-scale feature analysis. Proc SPIE 2012;8315:831516-831511.

66. Phan J, Quo C, Cheng C, Wang M. Multi-Scale Integration of-omic, Imaging, and Clinical Data in Biomedical Informatics. IEEE Rev Biomed Eng 2012;5:74-87.

67. Pieper S, Lorensen B, Schroeder W, Kikinis R. The NA-MIC Kit: ITK, VTK, pipelines, grids and 3D slicer as an open platform for the medical image computing community. 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro; 2016. p. 698-701.

68. Qi X, Kim H, Xing F, Parashar M, Foran DJ, Yang L. The analysis of image feature robustness using CometCloud. J Pathol Inform 2012:3-33.

69. Song Y, Treanor D, Bulpitt AJ, Magee DR. 3D reconstruction of multiple stained histology images. J Pathol Inform 2013;4(Suppl):S7.

70. Tay CH. Algorithms for Tissue Image Analysis using Multifractal Techniques. Master of Science Thesis. University of Canterbury. Computer Science and Software Engineering; 2012.

71. Viros A, Fridlyand J, Bauer J, Lasithiotakis K, Garbe C, Pinkel D, et al. Improving melanoma classification by integrating genetic and morphologic features. PLoS Medicine 2008;5(6):e120.

72. Yang L, Tuzel O, Chen W, Meer P, Salaru G, Goodell LA, et al. PathMiner: a web-based tool for computer-assisted diagonstics in pathology. IEEE Trans Inf Technol Biomed 2009;13(3):291-99.

73. Zhu LJ, Gao Y, Appia V, Yezzi A, Arepalli C, Faber T, et al. Automatic Delineation of the Myocardial Wall From CT Images Via Shape Segmentation and Variational Region Growing. IEEE Trans Biomed Eng 2013;60(10):2887-95.

74. Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, et al. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. PloS One 2015;10(3):e0118261.

75. Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. Annu Rev Pathol 2013;8:331-59.

76. Gurcan MN, Madabhushi A. Digital Pathology. Proc SPIE 2013;8676:867601-867601.

77. Wang C, Pécot T, Zynger DL, Machiraju R, Shapiro CL, Huang K. Identifying survival associated morphological features of triple negative breast cancer using multiple datasets. J Am Med Inform Assoc 2013;20(4):680-7.

78. Hamilton PW, Bankhead P, Wang Y, Hutchinson R, Kieran D, McArt DG, et al. Digital pathology and image analysis in tissue biomarker research. Methods 2014;70(1):59-73.

79. Gurcan MN, Madabhushi A. Medical Imaging 2015: Digital Pathology. Proc SPIE 2015;9420:942001-942001.

80. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Sci Transl Med 2011;3(108):108ra113.

81. Dong F, Irshad H, Oh EY, Lerwill MF, Brachtel EF, Jones NC, et al. Computational pathology to discriminate benign from malignant intraductal proliferations of the breast. PloS One 2014;9(12):e114885.

82. Qi X, Wang D, Rodero I, Diaz-Montes J, Gensure RH, Xing F, et al. Content-based histopathology image retrieval using CometCloud. BMC Bioinformatics 2014;15:287.

83. Hai S, Yong S, Fuyong X, Xin Q, Hirshfield KM, Lin Y, et al. Robust automatic breast cancer staging using a combination of functional genomics and image-omics. Conf Proc IEEE Eng Med Biol Soc 2015:7226-9.

84. Cheng J, Hipp J, Monaco J, Lucas DR, Madabhushi A, Balis UJ. Automated vector selection of SIVQ and parallel computing integration MATLAB: Innovations supporting large-scale and high-throughput image analysis studies. J Pathol Inform 2011;2:37.

85. Hipp J, Smith SC, Cheng J, Tomlins SA, Monaco J, Madabhushi A, et al. Optimization of complex cancer morphology detection using the SIVQ pattern recognition algorithm. Anal Cell Pathol (Amsterdam) 2012;35(1):41-50.

86. Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data--methods and preliminary results. Radiology 2012;264(2):387-96.

87. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep 2015;5:13087.

88. Fedorov A, Clunie D, Ulrich E, Bauer C, Wahle A, Brown B, et al. DICOM for quantitative imaging biomarker development: a standards based approach to sharing clinical data and structured PET/CT analysis results in head and neck cancer research. PeerJ 2016;4:e2057.

89. Boettiger C. An introduction to Docker for reproducible research. ACM SIGOPS Operating Systems Review 2015;49(1):71-9.

90. NCI Genomic Cloud Pilots.

91. Jacobsen DM, Canon RS. Contain this, unleashing docker for hpc. Proceedings of the Cray User Group; 2015.

92. Roelofs E, Dekker A, Meldolesi E, van Stiphout RGPM, Valentini V, Lambin P. International data-sharing for radiotherapy research: An open-source based infrastructure for multicentric clinical data mining. Radiother Oncol 2014;110(2):370-4.

93. Goode A, Gilbert B, Harkes J, Jukic D, Satyanarayanan M. OpenSlide: A vendor-neutral software foundation for digital pathology. J Pathol Inform 2013;4:27.

94. Toga AW (2012) The clinical value of large neuroimaging data sets in Alzheimer's disease. Neuroimaging clinics of North America, 22(1):107.

95. Grethe JS, Baru C, Gupta A, James M, Ludaescher B, Martone ME, et al. Biomedical informatics research network: building a national collaboratory to hasten the derivation of new understanding and treatment of disease. Stud Health Technol Inform 2005;112:100-10.

96. Marcus D, Harwell J, Olsen T, Hodge M, Glasser M, Prior F, et al. Informatics and Data Mining Tools and Strategies for the Human Connectome Project. Front Neuroinform 2011;5(4):1-12.

97. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. J Cogn Neurosci 2007;19(9):1498-507.

98. Hall D, Huerta MF, McAuliffe MJ, Farber GK. Sharing heterogeneous data: the national database for autism research. Neuroinformatics 2012;10(4):331-9.

99. Korfiatis PD, Kline TL, Blezek DJ, Langer SG, Ryan WJ, Erickson BJ. MIRMAID: A Content Management System for Medical Image Analysis Research. Radiographics 2015;35(5):1461-8.

100. Moore SM, Maffitt DR, Smith KE, Kirby JS, Clark KW, Freymann JB, et al. De-identification of Medical Images with Retention of Scientific Research Value. Radiographics 2015;35(3):727-35.

101. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging 2012;30(9):1323-41.

102. Grand Challenges in Biomedical Image Analysis. https://grand-challenge.org/. Accessed 2017-07-01.

103. Cooper LA, Kong J, Gutman DA, Wang F, Cholleti SR, Pan TC, et al. An integrative approach for in silico glioma research. IEEE Trans Biomed Eng 2010;57(10):2617-21.

104. Wang C, Pecot T, Zynger DL, Machiraju R, Shapiro CL, Huang K. Identifying survival associated morphological features of triple negative breast cancer using multiple datasets. J Am Med Inform Assoc 2013;20(4):680-7.

105. Doi K, Giger ML, Nishikawa RM, Hoffmann KR, Macmahon H, Schmidt RA, et al. Digital Radiography. A useful clinical tool for computer-aided diagnosis by quantitative analysis of radiographic images. Acta Radiol 1993;34(5):426-39.

106. Lo S-CB, Lin J-S, Freedman MT, Mun SK. Computer-assisted diagnosis of lung nodule detection using artificial convoultion neural network. Medical Imaging 1993. International Society for Optics and Photonics; 1993. p. 859-69.

107. Sahiner B, Chan HP, Petrick N, Wei D, Helvie MA, Adler DD, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. IEEE Trans Med Imaging 1996;15(5):598-610.

108. Greenspan H, Ginneken Bv, Summers RM. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. IEEE Trans Med Imaging 2016;35(5):1153-9.

109. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging 2015;34(10):1993-2024.

110. Bhargava R, Madabhushi A. Emerging Themes in Image Informatics and Molecular Analysis for Digital Pathology. Annu Rev Biomed Eng 2016;18:387-412.

111. Mongkolwat P, Kleper V, Talbot S, Rubin D. The National Cancer Informatics Program (NCIP) Annotation and Image Markup (AIM) Foundation model. J Digit Imaging 2014;27(6):692-701.

112. Müller H, Kalpathy-Cramer J, Hanbury A, Farahani K, Sergeev R, Paik JH, et al. Report on the Cloud-Based Evaluation Approaches Workshop 2015. ACM SIGIR Forum 2016;50:38-41.

113. Moreira DA, Hage C, Luque EF, Willrett D, Rubin DL. 3D markup of radiological images in ePAD, a web-based image annotation tool. Computer-Based Medical Systems (CBMS), 2015 IEEE 28th International Symposium on. IEEE; 2015. p. 97-102.

114. Belmann P, Dröge J, Bremges A, McHardy AC, Sczyrba A, Barton MD. Bioboxes: standardised containers for interchangeable bioinformatics software. Gigascience 2015;4(1):47.

115. Rubin DL. The electronic Imaging Physician Annotation Device (ePAD); 2017.

116. Keller BM, Chen J, Daye D, Conant EF, Kontos D. Preliminary evaluation of the publicly available Laboratory for Breast Radiodensity Assessment (LIBRA) software tool: comparison of fully automated area and volumetric density measures in a case-control study with digital mammography. Breast Cancer Res 2015;17:117.

117. Toolkit DD. DCMTK; 2010.

118. Torvalds L, Hamano J. Git: Fast version control system; 2010. URL http://git-scm.com.

119. Merkel D. Docker: lightweight linux containers for consistent development and deployment. Linux Journal 2014;(239):2.

120. Di Tommaso P, Palumbo E, Chatzou M, Prieto P, Heuer ML, Notredame C. The impact of Docker containers on the performance of genomic pipelines. PeerJ 2015;3:e1273.

121. Fridsma DB, Evans J, Hastak S, Mead CN. The BRIDG project: a technical report. J Am Med Inform Assoc 2008;15(2):130-137.

122. Singer DS, Jacks T, Jaffee E (2016) A US "Cancer Moonshot" to accelerate cancer research. Science, 353(6304):1105-6.

123. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. New Engl J Med 2016;375(12):1109-12.

124. caMicroscope - A Web Based Annotation and Visualization Platform for Digitized Whole Slide Images. https://github.com/camicroscope/caMicroscope. Accessed 2017-07-01.

Correspondence to:
Ashish Sharma
Woodruff Memorial Research Building
101 Woodruff Circle, #4105
Atlanta, GA 30322, USA
PhTel: +1 (404) 654 0124
E-mail: ashish.sharma@emory.edu