ORGANIZATION: THE CANCER IMAGING ARCHIVE
WEBSITE: http://www.cancerimagingarchive.net/
CONTACT INFORMATION: Justin Kirby, justin.kirby@fnlcr.nih.gov

PROBLEM CONTEXT

The Cancer Imaging Archive is a large organized archive of medical images of cancer accessible online for public download. This data are available to be utilized by cancer researchers, engineers, developers, professors, hackers and even the general public. This data has been used to find new ways to detect cancer and measure the change of the disease in response to treatments. A lot of this data comes from NCI-sponsored clinical trials or grants and thus was collected to address whatever the aim of the trial or grant happened to be. A good way to get a sense of what's been done with the data in the archive is to look over the related publications list at http://www.cancerimagingarchive.net/related-publications/.

TCIA opens the possibility for computer scientists to apply image processing methods to make novel discoveries relating to medical imaging without having to convince a hospital's IRB to give them data. Prior to TCIA it was pretty much impossible for entry-level researchers, grad students, etc to gain access to this kind of high quality medical imaging data because of PHI/HIPAA regulations and the time/effort associated with de-identifying it. Even for those who do work in medical imaging and have access to this kind of data, TCIA saves valuable time/effort by allowing re-use of existing data rather than everyone having to build up their own internal databases.

THE DATA

The data are mostly human (we have a small mouse data set) radiological images (MR, PET, CT) of patients with cancer. The images are in DICOM format. Each image contains a bunch of "tags" or metadata about the image stored in the file header (kind of like how an MP3 has the music itself, but also info about the artist, song, etc). Data are broken down into "Collections" which are just focused data sets usually relating to a particular type of cancer or research aim. Where possible we always try to augment the image data with supplemental data such as patient demographics, outcomes, genomics, or other expert-derived analysis (e.g. seed points or boundaries around the tumor). A full description of the collections themselves and the related data can be found at https://wiki.cancerimagingarchive.net/display/Public/Collections.

SOLUTION POSSIBILITIES

This problem statement is, by nature, open-ended. As such, you can be very creative with how you seek to use the data that is collected by TCIA. Here are a couple of possibilities.

Hardcore computer science/image analysis participants:

There are a huge number of data-driven questions we could ask that could lead to the hackers developing a tool which actually analyzes the images. We would just need to identify some relevant data subsets and come up with some targeted questions. Some preliminary ideas include:

"Using this set of images from lung cancer patients, can you build a tool that will find and detect the tumor using an automated algorithm and then calculate the volume?"

Related to this, we have a very well-funded competition coming up in this area over the next year with up to $1.8 million in prizes: http://www.fnih.org/press/releases/fnih-support-worldwide-competition-focused-improving-accuracy-lung-cancer-screening. I think doing something like the lung tumor detection project above could help prepare these students to participate in the Coding 4 Cancer competitions. A training data set could be provided and then a separate testing data set could be leveraged to evaluate how well the tool works.

"Using this set of images can you classify the following brain cancer patients as either high grade glioblastoma or low grade glioma?"

A training data set could be provided and then a separate testing data set could be leveraged to evaluate how well the tool works.

Web developers and usability experts:

A very common first step in cancer imaging research involves identifying the location of the tumor in the patient. Usually then the tumor is "segmented" to create a boundary that can be used to assess the size and many other characteristics of its appearance. Unfortunately identifying the location of the tumor is a tedious process and so we don't yet have this information on a large number of TCIA subjects. I think there is some promise in crowd sourcing this task but we do not have a system which would enable this. We do have an API on our database though. So perhaps we could recruit some people to design and build a web-based system which interacts with the API and pulls up some random patient's images in our system and then allow a trained user to drop a seed point on the tumor. While this kind of system would be generally targeted at getting trained radiologists to do the work, it's also feasible to train lay people how to identify certain types of tumors without much effort. We could also consider asking teams to design a system which provides some modular system in which we can easily insert sample images and instructions that can be presented as training materials before beginning use of the system to enable some citizen science.

Web developers and usability experts

The current web UI for browsing/searching the data is archaic. It would be quite valuable if we could get people to build an alternate UI via the REST API to browse, search and download the data using some modern web technologies.