TCGA Breast Phenotype Research Group Data sets (TCGA-Breast-Radiogenomics)

Description

At the time of our study, 108 cases with breast MRI data were available in the The Cancer Genome Atlas Breast Invasive Carcinoma Collection (TCGA-BRCA) collection. In order to minimize variations in image quality across the multi-institutional cases we included only breast MRI studies acquired on GE 1.5 Tesla magnet strength scanners (GE Medical Systems, Milwaukee, Wisconsin, USA) scanners, yielding a total of 93 cases. We then excluded cases that had missing images in the dynamic sequence (1 patient), or at the time did not have gene expression analysis available in the TCGA Data Portal (8 patients). After these criteria, a dataset of 84 breast cancer patients resulted, with MRIs from four institutions: Memorial Sloan Kettering Cancer Center, the Mayo Clinic, the University of Pittsburgh Medical Center, and the Roswell Park Cancer Institute. The resulting cases contributed by each institution were 9 (date range 1999-2002), 5 (1999-2003), 46 (1999-2004), and 24 (1999-2002), respectively. The dataset of biopsy proven invasive breast cancers included 74 (88%) ductal, 8 (10%) lobular, and 2 (2%) mixed. Of these, 73 (87%) were ER+, 67 (80%) were PR+, and 19 (23%) were HER2+. Various types of analyses were conducted using the combined imaging, genomic, and clinical data. Those analyses are described within several manuscripts created by the group (cited below).

Data Access

Data Access

Data Type	Download all or Query/Filter	License
Radiologist Annotations (XLS)	*Download*	CC BY 3.0
Segmentations (ZIP, XLS)	*Download*	CC BY 3.0
Quantitative Radiomic Features	*Download*	CC BY 3.0
MammaPrint, Oncotype DX, and PAM50 Multi-gene Assays (XLS)	*Download*	CC BY 3.0
Clinical Data (XLS)	*Download*	CC BY 3.0

Click the Versions tab for more info about data releases.

Collections Used in this Third Party Analysis

Below is a list of the Collections used in these analyses:

Source Data Type	Download all or Query/Filter	License
Corresponding Original Images from TCGA-BRCA (DICOM, 52GB)	*Download* (Open with the NBIA Data Retriever)	CC BY 3.0

- TCGA-BRCA

Please contact help@cancerimagingarchive.net with any questions regarding usage.

Detailed Description

Detailed Description

How to use the Segmentations

With regards to the naming structure, *S2-1.les: S2 means DCE-MRI sequence 2, lesion #1. Sometimes, there are multiple DCE-MRI sequences on TCIA data, and so the team used the sequence that corresponded to the one on which the radiologists annotated the truth. Each of our tumor segmentation files is a binary file, consisting of the following format:

1. six uint16 values for the inclusive coordinates of the lesion's cuboid , relative to the image:
y_start y_end
x_start x_end
z_start z_end

2. the N int8 on/off voxels (0 or 1) for the above specified cube, where N = (y_end y_start +1) * (x_end - x_start + 1) * (z_end - z_start + 1).

A voxel value of 1 denotes that it is part of the lesion, while a value of zero denotes it is not.

Please reference these data extracted using version **V2010** of the UChicago MRI Quantitative Radiomics workstation.
Citations & Data Usage Policy

Citations & Data Usage Policy

Users must abide by the TCIA Data Usage Policy and Restrictions. Attribution should include references to the following citations:

Data Citation

Morris, E., Burnside, E., Whitman, G., Zuley, M., Bonaccio, E., Ganott, M., Sutton, E., Net, J., Brandt, K., Li, H., Drukker, K., Perou, C., & Giger, M. L. (2014). **Using Computer-extracted Image Phenotypes from Tumors on Breast MRI to Predict Stage [Data set]**. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2014.8SIPIY6G

TCIA Citation

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). **The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository**. Journal of Digital Imaging, 26(6), 1045–1057. https://doi.org/10.1007/s10278-013-9622-7

In addition to the dataset citation above, please be sure to cite the following if you utilize these data in your research:

Publication Citation

Guo, W., Li, H., Zhu, Y., Lan, L., Yang, S., Drukker, K., Morris, E., Burnside, E., Whitman, G., Giger, M. L., Ji, Y., & TCGA Breast Phenotype Research Group. (2015). **Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data**. Journal of Medical Imaging, 2(4), 041007. https://doi.org/10.1117/1.jmi.2.4.041007

Publication Citation

Burnside E, Drukker K, Li H, Bonaccio E, Zuley M, Ganott M, Net JM, Sutton E, Brandt K, Whitman G, Conzen S, Lan L, Ji Y, Zhu Y, Jaffe C, Huang E, Freymann J, Kirby J, Morris EA, Giger ML. (2016) **Using computer-extracted image phenotypes from tumors on breast MRI to predict breast cancer pathologic stage.** Cancer 122(5): 748-757 . DOI: 10.1002/cncr.29791

Publication Citation

Zhu Y, Li H, Guo W, Drukker K, Lan L, Giger ML*, Ji Y*: **Deciphering genomic underpinnings of quantitative MRI-based radiomic phenotypes of invasive breast carcinoma**. Nature – Scientific Reports 5:17787. doi: 10.1038/srep17787, 2015.

Publication Citation

Li H, Zhu Y, Burnside ES, Drukker K, Hoadley KA, Fan C, Conzen SD, Whitman GJ, Sutton EJ, Net JM, Ganott M, Huang E, Morris EA, Perou CM, Ji Y, Giger ML. (2016) **MR Imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of gene assays of MammaPrint, Oncotype DX, and PAM50.** Radiology 281(2):382-391. doi: 10.1148/radiol.2016152110

Publication Citation

Li H, Zhu Y, Burnside ES, …. Perou CM, Ji Y, Giger ML: **Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA Dataset**. npj Breast Cancer (2016) 2, 16012; doi:10.1038/npjbcancer.2016.12; published online 11 May 2016.

Acknowledgment

Please also include the following acknowledgement:

"The authors would like to thank the TCGA Breast Phenotype Research Group for providing the computer-extracted tumor segmentation data used in this study. The tumor segmentation data comes from the University of Chicago lab of Maryellen Giger, whose lab members participated in the TCGA Breast Phenotype Research Group. In any presentation, poster, paper, etc, the segmentations should be identified as "Chicago Dynamic MRI Explorer 2005 Version". We would also like to acknowledge The Cancer Imaging Archive and The Cancer Genome Atlas initiatives for making the imaging and the clinical data used in this study publicly available."

Other Publications Using This Data

TCIA maintains a list of publications that leverage TCIA data. If you have a manuscript you'd like to add please contact the TCIA Helpdesk.

Versions

Version 1 (Current): 2018/09/04

Data Type	Download all or Query/Filter
Images (DICOM)	Download
Annotations (XLS)	Download
Segmentations (ZIP, XLS)	Download Download
Multi-gene Assays (XLS)	Download
Clinical Data (XLS)	Download