High-Throughput Quantitative MRI Phenotyping of TCGA/TCIA Breast Cancers for Prognostic Prediction

Collaboration between The University of Chicago and the TCGA/TCIA Breast Phenotype Group

THE UNIVERSITY OF CHICAGO

NorthShore University HealthSystem

THE CANCER IMAGING ARCHIVE

TCGA/TCIA Breast Phenotype Group

Quantitative Image Analysis and Computational Researchers:

University of Chicago :
- Maryellen Giger
- Hui Li
- Karen Drukker
- Li Lan

NorthShore:
- Yuan Ji
- Yitan Zhu
- Wentian Guo

Radiologists:
- Elizabeth Morris – Memorial Sloan Kettering
- Ermelinda Bonaccio – Roswell
- Kathleen Brandt – Mayo
- Elizabeth Burnside – U Wisconsin Madison
- Basak Dogan – MD Anderson
- Marie Ganott – Magee
- Jose Net – U Miami
- Elizabeth Sutton – Memorial Sloan Kettering
- Gary Whitman – MD Anderson
- Margarita Zuley – U Pittsburgh

National Cancer Institute:
- Carl Jaffe
- John Freymann
- Erich Huang
- Justin Kirby
- Brenda Fevrier-Sullivan

Purpose

We aim to demonstrate, using the TCGA/TCIA Breast image dataset, the performance of quantitative magnetic resonance imaging (MRI) analysis (i.e., MRI-based phenotyping) in characterizing the molecular subtypes of breast cancer. Understanding of the relationships between quantitative image analysis, molecular subtypes, and genomic data of breast cancer may ultimately improve prognostic assessment and contribute to more effective cancer treatment plans.

Quantitative Image Analysis and Image-based Phenotypes

Quantitative image analysis methods are being developed with the goal to serve as an aid in cancer diagnosis and patient management. In some instances – such as computer-aided detection in screening mammography – such methods have already become an integral part of today's clinical practice (1-2). Quantitative imaging aims to provide objective and effective tools for clinical decision making – serving as "measuring devices" rather than "imaging devices" (3). Quantitative image analyses address various biomedical questions using computer vision and artificial intelligence techniques. The role of quantitative medical image analysis in the detection and diagnosis of disease continues to increase, with methods being developed and evaluated for use in
- assessment of risk
- computer-aided detection (CADe)
- computer-aided diagnosis (CADx)
- assessment of prognosis
- assessment of response to therapy
- imaging genomics

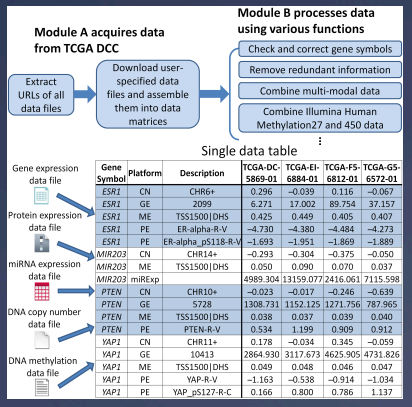
Computer Vision: Quantitatively extract lesion characteristics

Patient-specific: Image-based tumor signature for precision medicine

Population-based: Data-mining of computer-extracted features on large datasets for cancer discovery

TCGA/TCIA Dataset

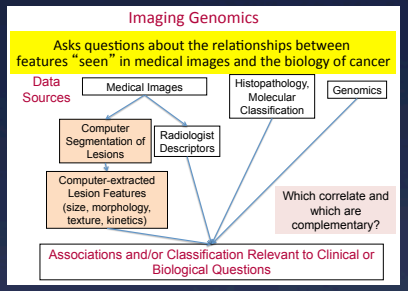
This study used the 98 breast cancer cases for which magnetic resonance images (MRIs) were available in The Cancer Imaging Archive (TCIA) and which corresponded to breast cancer cases available in the The Cancer Genome Atlas (TCGA). Clinical, genomics, epigenomics and proteomics data of these cancer cases were prepared using TCGA-Assembler (4), an open-source software freely available at *www.compgenome.org* for retrieving and processing TCGA Data.



Gene Symbol	Platform	Description	TCGA-DC-5869-01	TCGA-EI-6884-01	TCGA-F5-6812-01	TCGA-G5-6572-01
ESR1	CN	CHR6+	0.296	–0.039	0.116	–0.067
ESR1	GE	2099	6.271	17.002	89.754	37.157
ESR1	ME	TSS1500|DHS	0.425	0.449	0.405	0.407
ESR1	PE	ER-alpha-R-V	–4.730	–4.380	–4.484	–4.273
ESR1	PE	ER-alpha_pS118-R-V	–1.693	–1.951	–1.869	–1.889
MIR203	CN	CHR14+	–0.293	–0.304	–0.375	–0.050
MIR203	ME	TSS1500|DHS	0.050	0.090	0.070	0.037
MIR203	miRExp		4989.304	13159.077	2416.061	7115.598
PTEN	CN	CHR10+	–0.023	–0.017	–0.246	–0.639
PTEN	GE	5728	1308.731	1152.125	1271.756	787.965
PTEN	ME	TSS1500|DHS	0.038	0.037	0.039	0.040
PTEN	PE	PTEN-R-V	0.534	1.199	0.909	0.912
YAP1	CN	CHR11+	0.178	–0.034	0.345	–0.059
YAP1	GE	10413	2864.930	3117.673	4625.905	4731.826
YAP1	ME	TSS1500|DHS	0.049	0.048	0.046	0.047
YAP1	PE	YAP-R-V	–1.163	–0.538	–0.914	–1.034
YAP1	PE	YAP_pS127-R-C	0.166	0.800	0.786	1.137

Imaging Genomics

Imaging genomics involves the correlative investigation of image data, clinical data, histopathology data, and genomic data in order to understand the molecular biology behind the imaging characteristics of tumors and normal structures. [e.g., (7-13)]

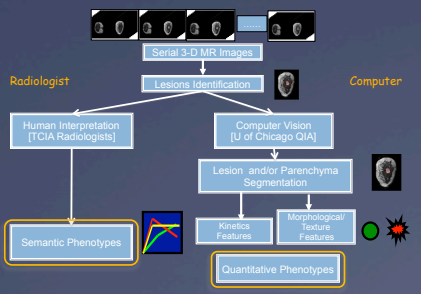


From medical images, either semantic descriptors, given subjectively by radiologists, or quantitative mathematical descriptors calculated using computer vision techniques, i.e., quantitative image-based phenotypes, can be used for input to data mining techniques and association analyses (2,5-12). By learning which phenotypes are correlated and which are complimentary to each other, investigators aim to merge the various multi-scale, multi-modality image-based phenotypes into advanced tumor signatures for predicting cancer risk and/or for patient management.

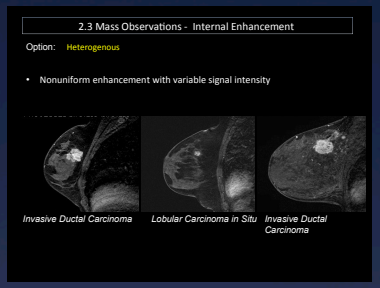
Computer output from quantitative image analysis is expected to aid in understanding disease by yielding methods of "high-throughput image-based phenotyping" (11,12), or "radiogenomics" (13). Thus, computer vision workstations need to effectively and efficiently analyze large populations in a user-friendly manner. The workstation below (demonstrated at RSNA 2010), automatically performs lesion segmentation, feature extraction (i.e., image-based phenotyping), and feature merging (i.e., classification) to yield image-based tumor signatures.

High-Throughput Diagnostic Phenotyping

Image Analysis Workflow

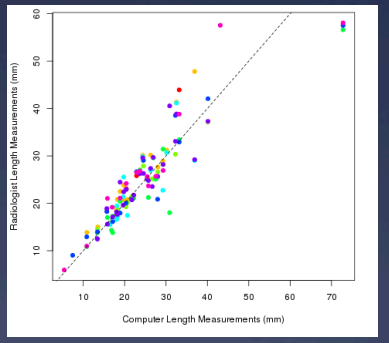


Radiologist Semantic Phenotypes



Above: Examples illustrating radiologist-extracted descriptors of internal enhancement.

Radiologist vs. Computer



Above: Size measurement visually determined by the radiologists and the size measurement automatically calculated by the computer workstation

References

1. Giger ML, Chan H-P, Boone J. Anniversary Paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM. Medical Physics 35: 5799-5820, 2008.
2. Giger ML, Karssemeijer N, Schnabel J. Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. Annual Review of Biomedical Engineering 15:327-357, 2013.
3. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC, et al. . A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging . Radiology 258: 906-914, 2011.
4. Zhu Y, Qiu P, Ji Y. TCGA-Assembler: Open-Source Software for Retrieving and Processing TCGA Data. Nature Method, 11(6), 2014.
5. Huo Z, Giger ML, Olopade OI, et al. Computerized analysis of digitized mammograms of BRCA1/BRCA2 gene mutation carriers. Radiology 225:519-526, 2002.
6. Bhooshan N, Giger ML, Jansen S, et al. Image-based Prognostic Markers from Computerized Characterization of a Clinical Breast MRI Database. Radiology 254: 680-690, 2010.
7. Aleza G, Monaco J, Doyle S, et al. Towards Improved Cancer Diagnosis and Prognosis Using Analysis of Gene Expression Data and Computer Aided Imaging. Experimental Biology and Medicine 234, 860-879, 2009.
8. Gevaert O, Xu J, Hoang CD, et al. Non-Small Cell Lung Cancer: Identifying Prognostic Imaging Biomarkers by Leveraging Public Gene Expression Microarray Data – Methods and Preliminary Results . Radiology 264 (2): 387-396, 2012.
9. Segal E, Sirlin CB, Ooi Clara, et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. Nature Biotechnology 25(6): 675-680, 2007.
10. Li H, Giger ML, Sun C, Ponsukcharoen U, Huo D, Lan L, Olopade OI, Jalmeson AR, Bancroft Brown J, Di Renzo A. Association study of image-based phenotypes and genomic biomarkers, potential in breast cancer risk assessment. Medical Physics 41, 2014.
11. Jamieson A, Giger ML, Drukker K, et al. Exploring non-linear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and t-SNE. Medical Physics 37: 339-351, 2010.
12. Giger ML, Huo Z, Lan L, Vyborny CJ. Intelligent search workstation for computer-aided diagnosis. Proc. of Computer Assisted Radiology and Surgery (CARS'2000), pp. 822-827, 2000.
13. Kuo MD, Rubman AM. Radiogenomics: Creating a link between molecular diagnostics and diagnostic imaging. European Journal of Radiology 70: 232-241, 2009.

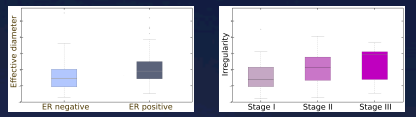
Acknowledgements & COI

Supported over the years by various research, shared instrument, and training grants from the NIH (NCI, NIBIB, NIAMS), the DOD, the DOE, and the Segal Family Foundation, as well as internal grants including those from the UC Comprehensive Cancer Center, the SPORE, and the ITMCTSA.

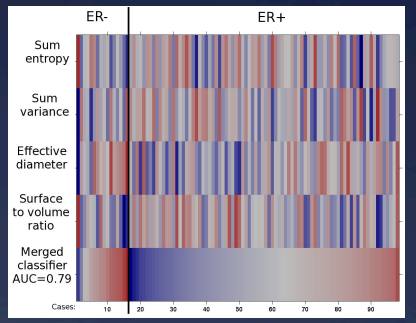
This project has been funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

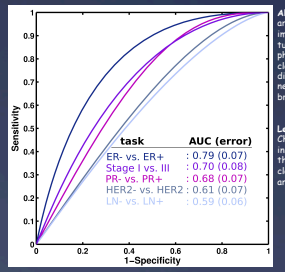
COI: M L Giger is a stockholder in R2/Hologic, co-founder and equity holder in Quantitative Insights, and receives royalties from Hologic, GE Medical Systems, MEDIAN Technologies, Riverain Medical, Mitsubishi, and Toshiba. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interest that would reasonably appear to be directly and significantly affected by the research activities.

Image-based Quantitative Phenotypes



Above: Box plots of normalized image-based quantitative phenotypes for the 98 breast cancer cases stratified by ER status and pathological stage, respectively





Above: MRI-based phenotype array of computer-extracted image-based phenotypes and a tumor signature from phenotypes merged by a classifier in the task of distinguishing between ER-negative and ER -positive breast cancers

Left: Receiver Operating Characteristic (ROC) curves indicating the performance of the computer in several classification tasks. *AUC:* area under the curve.

Summary and Conclusion

Quantitative image analysis was performed on 98 de-identified MRI studies depicting biopsy-proven breast cancers from the NCI's multi-institutional The Cancer Imaging Archive (TCIA) and The Cancer Genome Atlas (TCGA) project. Computerized image-based phenotyping was completely automated apart from the indication of the lesion center and included: 1) 3D lesion segmentation, 2) feature extraction (i.e., extraction of image-based phenotypes) , 3) leave-one-case-out linear stepwise feature selection, and 4) leave-one-case-out cross-validation merging image-based phenotypes to form a prognostic predictive classifier. The performance of the classifier model for molecular subtyping was evaluated using ROC analysis with the area under the ROC curve (AUC) as the figure of merit.

The results from this study indicate that quantitative MRI analysis shows promise as a means for high-throughput image-based phenotyping in the discrimination of breast cancer subtypes. In the future, the merging of image-based phenotypes with genomic data may lead to improved prognostic predictors.