A Radiogenomic Analysis of the TCGA Glioblastoma Data Set

David A Gutman[1], William Dunn Jr[1], Lee AD Cooper[1], Lisa Scarpace[2], Erich Huang[3], Tom Mikkelsen[2], Rajan Jain[2] ,Scott N Hwang[1], Chad A Holder[1], John Freymann[4], Justin Kirby[4], Carl Jaffe[5] , Adam Flanders[3], Rivka Colen[6], Dan Brat[1]

[1]Departments of Biomedical Informatics, Emory University Hospital, Atlanta, GA. [2]Henry Ford University Hospital, Detroit, Michigan. [3]National Institute of Health, Bethesda, MD. [4]SAIC-Frederick Inc., Frederick, MD. [5]Boston University School of Medicine, Boston, MA. [6]MD Anderson Cancer Center Anderson, Houston TX

EMORY WINSHIP CANCER INSTITUTE NCI–Designated

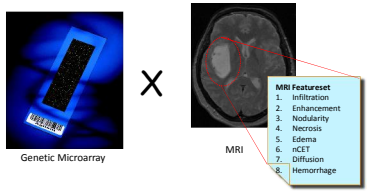
BACKGROUND

Genomic based approaches are increasingly being used to define subtypes of glioblastoma (GBM), in part due to the availability of The Cancer Genome Atlas (TCGA) data as well as to advances in bioinformatics methods. While the TCGA was primarily focused on molecular/genomic analysis, corresponding pathology and radiology data is also available from several related sources, The Cancer Imaging Archive(TCIA) and the Cancer Digital Slide Archive (CDSA).

While imaging plays a critical role in the initial evaluation of gliomas, as well as in monitoring progression, currently its value in subtyping tumors is quite limited. Marked heterogeneity is often apparent when visually inspecting MRI images, however capturing this information in robust and quantifiable ways can be challenging. Importantly, it allows researchers to identify the microenvironment of the tumor which very likely impacts genetic expression patterns and other genomic aberrations.

Evaluation of genomic data within the macroscopic context of the tumor, which can be extracted from the MRI images, allows more thorough characterization of a tumor. This will ultimately improve our ability to identify patient subtypes, but ideally advances in patient care.

In this presentation, we will describe a work flow pipeline and initial findings of a radio-genomic analysis of a subset of TCGA GBM tumors.



MATERIALS & METHODS

T2-FLAIR and post gadolinium (GD) contrast T1-weighted pre-surgical image sequences from each of 89 TCGA patients were obtained through The Cancer Imaging Archive. Next, binary masks identifying tumor location were generated on axial slices, with the flair envelope consisting of all regions of abnormal signal including non-contrast enhancing tumor and edema. The contrast enhancing region from the T1-Post GD image were subsequently parcellated into a contrast enhancing region (CE) and a necrotic region (N) (Fig 2). These indices, along with their derivatives (e.g. ratios such as N/CE), were then correlated against corresponding mRNA expression TCGA data using a simple regression analysis corrected for repeated tests (SAM).

MATERIALS & METHODS

Using this methodology, we generated a list of genes that were strongly correlated with a specific macroscopic property of the tumor; these gene lists then underwent enrichment analysis using Ingenuity Pathway Analysis to identify underlying biological phenomena associated with these imaging features.

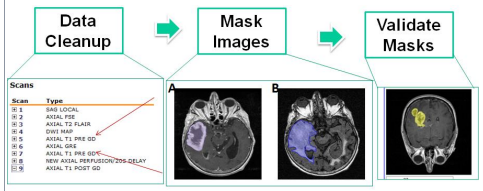


Figure 1: Overview of radio-genomic analysis workflow

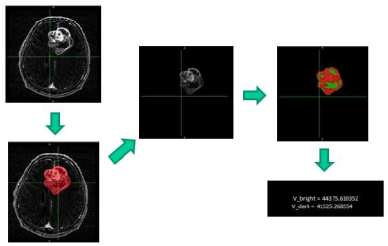


Figure 2: Semi-automated *in silico* segmentation and measurement method for T1 weighted post Gd images. Tumor volume is identified by a manually drawn binary mask (Red). K means clustering divides pixels covered by the mask into bright (red, contrast enhancement) and dark (green, necrosis) clusters based on relative pixel intensity.

RESULTS

I. Mask Generation – Mask Validation

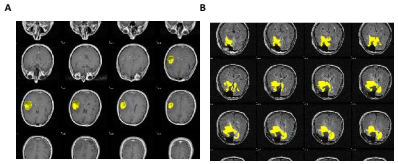


Figure 3: 'TumorView' Mask Validator. Image sets of TCGA patients with accompanying binary masks were uploaded to a web server for validation. Patients with acceptable images and mask overlay (A) were cleared for further analysis whereas patients with poor images or post-surgical images (B) were excluded.

RESULTS

II. Imaging feature volumetric analyses

Variable	Hazard Ratio (HR)	95% Confidence Interval	P Value
Gender	1.115	0.854 - 1.376	0.684
KPS	**1.473**	**0.941 - 2.005**	**0.046**
Age	1.015	0.699 - 1.331	0.124
Necrosis (mm3)	1.365	0.986 - 1.89	0.061
Contrast Enhancement (mm3)	1.276	0.929 - 1.752	0.132
Edema (mm3)	0.766	0.569 - 1.031	0.078
Total Abnormal (mm3)	1.028	0.698 - 1.516	0.887
Total Tumor (mm3)	1.339	0.963 - 1.863	0.083
Necrosis / Edema	**1.578**	**1.207 - 2.063**	**0.001**
Necrosis / Total Tumor	2.932	0.515 - 16.685	0.225
Percent Necrosis	**1.862**	**1.163 - 2.982**	**0.01**
Contrast Enhancement / Edema	**0.658**	**0.497 - 0.871**	**0.003**
Contrast Enhancement / Tumor	0.481	0.119 - 1.947	0.305
Percent Edema	**0.353**	**0.213 - 0.585**	**<0.001**
Total Tumor / Edema	**1.837**	**1.116 - 3.026**	**0.017**
Contrast Enhancement / Necrosis	0.64	0.294 - 1.394	0.262
Total Tumor / Total Abnormality	**1.817**	**1.108 - 2.978**	**0.018**

Table 1: Cox proportional hazards model univariate analyses of individual parameters for correlations with overall survival rate. KPS: Karnofsky Performance Scale score. Significant values (P<0.05) are indicated in bold.

III. Genetic correlation to volumetric features

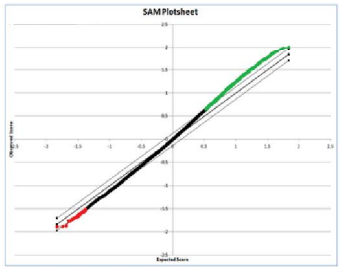


Figure 4: Graphical representation of the results of a Statistical Analysis of Microarray (SAM) analysis. Genes above upper FDR (false-discovery rate) threshold are considered significantly up regulated in necrosis patients and are indicated in green. Red genes are inversely correlated to increasing levels of tumor necrosis.

IV. Pathway correlation to volumetric features

Pathway	P-value	Gene Symbol	Gene Name
Hypoxia-Inducible Factor Signaling	0.0000643	EIF2B3	Eukaryotic translation initiation factor 2B, subunit 3
		EIF2B4	Eukaryotic translation initiation factor 2B, subunit 4
		EIF2B5	Eukaryotic translation initiation factor 2B, subunit 5
		SUMO1	SMT3 suppressor of mif two 3 homolog 1
		UBE2I	Ubiquitin-conjugating enzyme E2I
Anti-Apoptosis	0.0159	BIRC7	Baculoviral IAP repeat containing 7
		BNIP1	BCL2/adenovirus E1B 19kDa interacting protein 1
Cell Cycle: G2/M DNA Damage Checkpoint Regulation	0.0341	CKS2	CDC28 protein kinase regulatory subunit 2
		MYT1	Myelin transcription factor
Mitochondrial Dysfunction	0.0501	NDUFA2	NADH dehydrogenase (ubiquinone) 1 alpha
		NDUFS5	NADH dehydrogenase (ubiquinone) Fe-S protein 5
		RHOT2	Ras homolog gene family, member T2

Table 2: Results of Ingenuity Pathway Analysis (IPA) functional analysis for 118 genes differentially expressed between patients with low and high levels of percent necrosis and the respective levels of significance.

RESULTS

IV. Pathway correlation to volumetric features

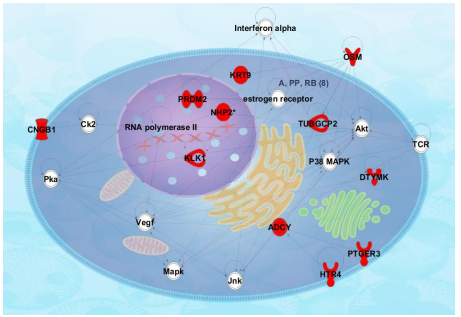


Figure 5: Representation of a subset of genes differentially expressed between patients with low and high levels of percent necrosis. Genes up-regulated in high necrosis patients (red) are involved in cell-to-cell signaling and interaction, cellular growth and proliferation, and connective tissue development and function pathways (white). Dashed lines indicate experimentally determined indirect interactions. Complete lines indicate direct interactions.

CONCLUSIONS

- We have developed a pipeline which efficiently identifies and accurately measures relevant MR imaging features.
- Our analyses show that several of these imaging features can be statistically correlated to genetic expression events as well as to underlying biological pathways.
- As genomic analyses of tumor become routine, the additional incorporation of macroscopic data will become important to more comprehensively characterize the tumor as well as to provide physicians with more accurate diagnostic and prognostic tools.

REFERENCES

- Gutman DA, Cooper LAD, Hwang S, Flanders A, Brat D. MR Imaging Predictors of Molecular Profile and Survival: A Multi-Institutional Study of the TCGA Glioblastoma Data Set. Radiology. 2013 May; 267(2):560-569.
- Zinn, P. O., Majadan, B., Sathyan, P., Singh, S. K., Majumder, S., Jolesz, F. A., & Colen, R. R. (2011). Radiogenomic Mapping of Edema/Cellular Invasion MRI-Phenotypes in Glioblastoma Mu
- Tusher, Virginia G., Robert Tibshirani, and Gilbert Chu. "Significance Analysis of Microarrays Applied to the Ionizing Radiation Response." *PNAS* 98.9 (2001): 5116-121.

ACKNOWLEDGEMENTS

This work was supported in part with funds from the Georgia Cancer Coalition, the NCI funded In Silico Brain Tumor Research Center (SAIC-F/NCI 29XS193) and NCI R01