19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011
http://mssanz.org.au/modsim2011

Identifying gene alterations required for the development of astrocytoma

Brian Kunkle[a], Changwon Yoo[b], Quentin Felty[a] and Deodutta Roy[a]

[a]Departments of Environmental and Occupational Health and [b]Biostatistics, Robert Stempel College of Public Health and Social Work, Florida International University, Miami, FL 33199, USA
Email: Droy@fiu.edu

ABSTRACT: A growing body of evidence suggests that there are critical periods of time extending from conception to puberty when the central nervous system in children may be more affected by environmental toxin exposures. These exposures may likely interact with the genome/epigenome of the fetus or young child to produce alterations in their genetic makeup which can predispose to development of disease, including glial tumors. However, very few studies to date have investigated gene-environment interaction in relation to the development of glial tumors. Interactions of environmental factors with genetic and epigenetic changes are expected to contribute in the development of the particular type of glial tumor in an individual. Glial tumors are usually broken down into more specific subtypes based on their predicted cell type of origin. The most common glial tumors include astrocytoma (originated from astrocytes), oligodendroglioma (originated from oligodendrocytes), brain stem glioma (originated from brain stem cells), and ependymoma (originated from ependymal cells). The assessment of gene–environment interaction in glial tumors has been more complex because of the lack of sound molecular epidemiological studies with a more complete picture of individual cancer risk associated with environmental exposure and genetic analysis.

Our goal in this study was to identify gene-environment interactions that are critical in the development of glioblastoma multiforme (GBM), the most common and aggressive type of human brain tumor. A GBM is a grade IV astrocytoma. We have used environmental bioinformatic resources for investigation of gene-environment interactions in the development of astrocytoma. We also combined these analytic approaches through first combining available microarray data on astrocytoma using a meta-analysis approach, and then conducting gene pathway networking analysis on results of this meta-analysis. Genes responsive to environmental exposures were identified using the Environmental Genome Project, Comparative Toxicology, and Seattle SNPs databases. These genes were then compared to a curated list of genes altered in GBM. The list of genes responsive to the environment and important to GBM was then further investigated using gene networking tools such as RSpider and Cytoscape.

Overlapping of final list of GBM alterations with the environmental genes found 173 genes that had an environmental exposure link and were altered in glioblastoma. Of these 173 genes, a Pubmatrix search found that 65 overlapping genes had not been previously assessed in glioblastoma research. A specific search for chemical-gene interactions producing mutagenesis in our genes found 226 results. The main biological functions of these genes included Signaling by Nerve Growth Factor (NGF), DNA Repair, Integrin Cell Surface Interactions, Biological Oxidations, Apoptosis, Synaptic Transmission, Cell Cycle Checkpoints, and Arachidonic Acid Metabolism. Four separate analyses were run in Banjo in order to search for genes critical for Grade I Astrocytoma development. Top Bayesian network and Markov blanket genes identified for Grade I Pilocytic Astrocytoma were IGFB5, TIMP4, SSR2, LPL, DUSP7, GABRA5, SH3GL3, C1S, ANK3, HLAA, EIF4A1, PTGER3, CCND2. Many of the genes identified in this study have in fact been implicated in the development of astrocytoma, including EGFR, HIF-1α, c-Myc, WNT5A, and IDH3A. In summary, this study was able to identify a set of key genes significantly dysregulated during the development of GBM. Findings of this study have a major implication for the role of gene-environment interactions in the development of GBM, suggesting some of the key genes with potential to contribute to GBM.

Keywords: *Bioinformatics, Gene-Environment Interactions, Glioblastoma*

1. INTRODUCTION

Astrocytomas are neoplasms of the brain that originate in a type of glial cell called an astrocyte. Very few epidemiological studies to date have investigated gene-environment interactions (GEI) in relation to brain tumor development. Recent advances in high-throughput microarrays have produced a wealth of information concerning molecular biology of astrocytoma. Bioinformatics tools that help for the assessment of pathway and gene relationships, text mining of published literature, and integration of large amounts of diverse biological and environmental data allow for hypothesis driven investigations of gene-gene and gene-environment interactions. In particular, microarrays have been used to obtain genetic and epigenetic changes between normal non-tumor tissue and brain tumor tissue. Due to the relative rarity of gliomas, microarray data for these tumors is often the product of small studies, and thus pooling this data becomes desirable. Additionally, analysis of microarray data has been an evolving field as techniques such as cluster analysis, networking analysis and principal components analysis have been used in order to tease biologically relevant information from the large amount of data produced from microarrays. Methods that exploit these tools have been applied to modeling of gene-environment interactions in depression and alcohol use [1], and bipolar disorder and its interaction with both tobacco use [2] and lithium treatment [3]. Integration of toxicological and pharmacological databases such as the Comparative Toxicological Database (CTD) [4] and Environmental Genome Projects (EGP) [5] with data on genetic alterations has also proven useful in developing hypothesis for research into GEI related diseases [6,7].

Evidence of environmental exposures causing copy number variations (CNV) is still developing, while the environments ability to cause single nucleotide polymorphism (SNP) mutations is quite established. Although evidence suggests that most common copy number variants are inherited and therefore caused by ancestral structural mutations, there is growing evidence that many or most normal and sporadic, non-recurrent CNVs, which account for the majority of disease-associated CNVs in humans and those in cancers, arise via mechanisms coupled to aberrant DNA replication and/or non-homologous repair of DNA damage. This suggests an unexpected mitotic, rather than meiotic, cell origin for many CNVs and has a number of important implications for the role of environmental exposures in their formation. This evidence has led some to hypothesize that the two types of environmental agents most likely to be associated with CNV formation are: 1) agents that lead to replication stress, which might lead to CNVs through secondary breakage or replicative template switching, and 2) agents that directly induce DNA double-strand breaks (DNA DSBs), which might lead to CNVs through inappropriate joining of broken ends. Moreover, the ability of environmental agents to cause CNVs and induce epigenetic transgenerational effects in the sperm epigenome separate from methylation effects has recently been established.

In an attempt to identify genes potentially important in environmentally related alterations in GBM, we have applied bioinformatic methods for identifying potentiall environmentally-related GBM genes. We have searched for specific gene-chemical observations important for 'environmentally responsive genes', and develop a gene network using these genes and alterations in GBM. We used environmental bioinformatic resources for investigation of gene-environment interactions in the development of astrocytoma. We also combined these analytic approaches through first combining available microarray data on astrocytoma using a meta-analysis approach, and then conducting gene pathway networking analysis on results of this meta-analysis.

2. METHODS

The copy number alterations and SNP mutations in GBM were curated from published literature and the COSMIC database [8]. Gene lists from six studies on glioblastoma multiforme were used for this curation [8-14]. Design of study, analysis platform, sample size and region of genome analyzed were criteria used to select studies to include in our alteration results. Gene lists from three environmental databases were used for the compilation of possible environmentally important genes in glioblastoma. These databases were: (a) Environmental Genome Project (EGP): http://egp.gs.washington.edu/finished_genes.html. (b) Seattle SNPs: http://pga.gs.washington.edu/finished_genes.html. (c) Comparative Toxicogenomics Database (CTD): http://www.mdibl.org/research/ctd.shtml. The gene lists from the glioblastoma alterations search and our environmental genes database search were then inputted into the GeneVenn program [15] to assess their overlap. Gene overlaps between the 3 environmental gene databases and our glioblastoma alterations list were determined. Overlapping genes were used for further analysis including, Pubmatrix (http://pubmatrix.grc.nia.nih.gov). The USCS golden pathway database (http://pubmatrix.grc.nia.nih.gov/), and SNPper (http://chip.org/bio/snpper-entergene), were used to identify functional SNPs and genes that have functional SNPs. SNPper searches both the USCS Genome Browser and the NCBI dbSNP database for relevant SNPs and genes. The overlapping genes list was searched in the Comparative Toxicology Database for relevant gene-chemical interactions, chemical associations, pathway associations, and GO processes associations. Both the overlapping genes list and the entire glioblastoma alterations list was subjected to gene networking analysis using the PathCluster [16] and Moksiskaan [17] gene networking tools.

3. RESULTS AND DISCUSSION

Glioblastoma associated copy number alterations and SNP mutations were found in the 6 total studies and 1 database we searched (Table 1). A total of 217 amplified genes, 214 copy number gain genes, 350 deleted genes, 161 copy number loss genes, and 2410 SNP mutated genes were found. According to the COSMIC database, 2,129 genes have been found to be mutated in glioblastoma, while 15,733 genes have been sequenced in glioblastoma where no mutation has been found. The top 15 mutated GBM genes in our COSMIC with the percentage of mutated genes per tumors analyzed are shown in Figure 1. Sixty seven amplified/gain genes and 45 deleted/loss genes due to structural mutational changes are shown in Figure 2. Six amplified/gain genes were shown to have copy number deletion or loss.

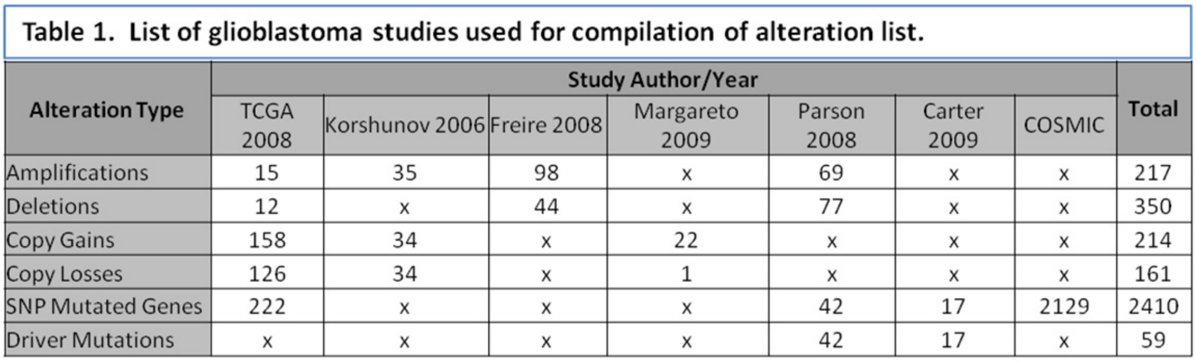

Table 1. List of glioblastoma studies used for compilation of alteration list.

Alteration Type	Study Author/Year							Total
	TCGA 2008	Korshunov 2006	Freire 2008	Margareto 2009	Parson 2008	Carter 2009	COSMIC	
Amplifications	15	35	98	x	69	x	x	217
Deletions	12	x	44	x	77	x	x	350
Copy Gains	158	34	x	22	x	x	x	214
Copy Losses	126	34	x	1	x	x	x	161
SNP Mutated Genes	222	x	x	x	42	17	2129	2410
Driver Mutations	x	x	x	x	42	17	x	59

The search of the environmental databases returned 648 Environmental Genome Project (EGP) genes (environmentally responsive genes), 319 Seattle SNP (SSNP) genes (inflammatory genes), and 15 Comparative Toxicology Database (CTD) genes (toxicogenomic genes). Very little overlap existed between these gene sets (4 between EGP and CTD, 3 between SSNP and CTD, 8 between EGP and SSNP, and 1 between all three databases). Overlapping of final list of GBM alterations with the environmental genes found 173 genes that had an environmental exposure link and were altered in glioblastoma. Of these 173 genes, a Pubmatrix search found that 65 overlapping genes had not been previously assessed in glioblastoma research.

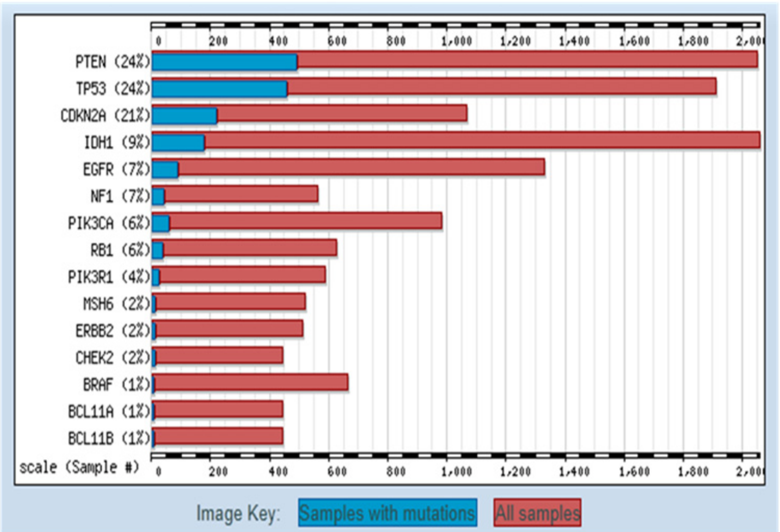

Figure 1. Top Glioblastoma genes with mutations

The list of 173 genes potentially important in GEI in glioblastoma formation was then subjected to analysis in the Comparative Toxicology Database. Results indicated substantial potential for chemical interaction with these set of genes including interactions with chemicals such as arsenic and pesticides such as chlorpyrifos for gene ABCB1, benzene and bisphenol A for EGFR, and estradiol for MDM2 and NCOA1. In total, the list of 173 genes produced 30,983 gene-chemical interactions and showed 13,779 chemical associations. A specific search for chemical-gene interactions producing mutagenesis in our genes found 226 results. Four separate analyses were run in Banjo in order to search for genes critical for Grade I Astrocytoma development. Top Bayesian network and Markov blanket genes identified for Grade I Pilocytic Astrocytoma were IGFB5, TIMP4, SSR2, LPL, DUSP7, GABRA5, SH3GL3, C1S, ANK3, HLAA, EIF4A1, PTGER3, CCND2. This study produced several major findings including identification of a list of top over- and under-expressed genes among 12 sub-studies on astrocytoma, identification of several genes important to development of astrocytomas, identification of important signaling pathways in astrocytic tumors, and identification of possible mechanisms which explain the genes and pathways identified as important to the development of astrocytoma.

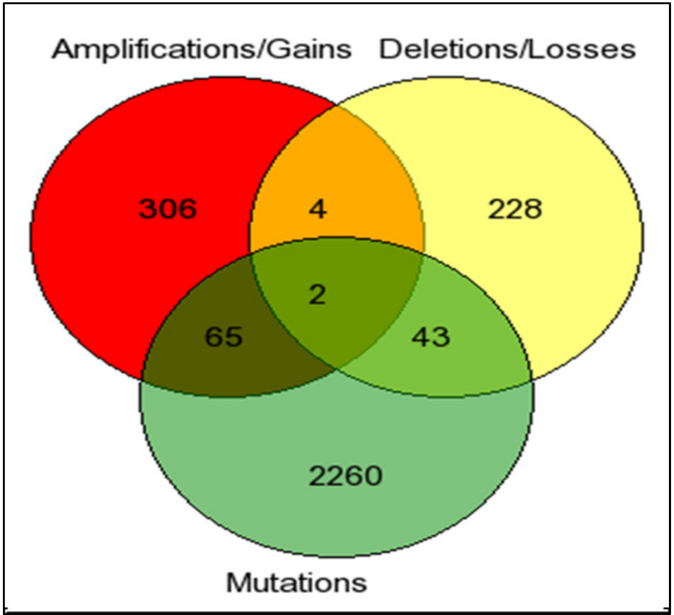

Figure 2. Overlap between glioblastoma amplified/gain genes, deleted/loss genes and SNP mutated geneswith mutations.

4. REFERENCES

1. McEachin,R.C., Keller,B.J., Saunders,E.F., and McInnis,M.G. (2008). Modeling gene-by-environment interaction in comorbid depression with alcohol use disorders via an integrated bioinformatics approach. *BioData Mining*, 1**,** 2.
2. McEachin,R.C., Saccone,N.L., Saccone,S.F., Kleyman-Smith,Y.D., Kar,T., Kare,R.K., Ade,A.S., Sartor,M.A., Cavalcoli,J.D., and McInnis,M.G. (2010). Modeling complex genetic and environmental influences on comorbid bipolar disorder with tobacco use disorder. *BMCMedical Genetics*, 11, 14.
3. McEachin,R.C., Chen,H., Sartor,M.A., Saccone,S.F., Keller,B.J., Prossin,A.R., Cavalcoli,J.D., and McInnis,M.G. (2010). A genetic network model of cellular responses to lithium treatment and cocaine abuse in bipolar disorder. *BMC Systemic Biology*, 4, 158.
4. Davis,A.P., Murphy,C.G., Saraceni-Richards,C.A., Rosenstein,M.C., Wiegers,T.C., and Mattingly,C.J. (2009). Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical-gene-disease networks. *Nucleic Acids Research*, 37, D786-D792.
5. Rieder,M.J., Livingston,R.J., Stanaway,I.B., and Nickerson,D.A. (2008). The environmental genome project: reference polymorphisms for drug metabolism genes and genome-wide association studies. *Drug Metabolism Review*, 40, 241-261.
6. Bauer-Mehren,A., Bundschus,M., Rautschka,M., Mayer,M.A., Sanz,F., and Furlong,L.I. (2011). Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. *PLoS One.*, 6, e20284.
7. Herbert,M.R., Russo,J.P., Yang,S., Roohi,J., Blaxill,M., Kahler,S.G., Cremer,L., and Hatchwell,E. (2006). Autism and environmental genomics. *Neurotoxicology*, 27, 671-684.
8. Forbes,S.A., Bindal,N., Bamford,S., Cole,C., Kok,C.Y., Beare,D., Jia,M., Shepherd,R., Leung,K., Menzies,A., Teague,J.W., Campbell,P.J., Stratton,M.R., and Futreal,P.A. (2011). COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Research*, 39, D945-D950.

9. The Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455, 1061-1068.
10. Carter,H., Chen,S., Isik,L., Tyekucheva,S., Velculescu,V.E., Kinzler,K.W., Vogelstein,B., and Karchin,R. (2009). Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Research*, 69, 6660-6667.
11. Freire,P., Vilela,M., Deus,H., Kim,Y.W., Koul,D., Colman,H., Aldape,K.D., Bogler,O., Yung,W.K., Coombes,K., Mills,G.B., Vasconcelos,A.T., and Almeida,J.S. (2008). Exploratory analysis of the copy number alterations in glioblastoma multiforme. *PLoS One.*, 3, e4076.
12. Korshunov,A., Sycheva,R., and Golanov,A. (2006). Genetically distinct and clinically relevant subtypes of glioblastoma defined by array-based comparative genomic hybridization (array-CGH). *Acta Neuropathology*, 111, 465-474.
13. Margareto,J., Leis,O., Larrarte,E., Pomposo,I.C., Garibi,J.M., and Lafuente,J.V. (2009). DNA copy number variation and gene expression analyses reveal the implication of specific oncogenes and genes in GBM. *Cancer Investigation*, 27, 541-548.
14. Parsons,D.W., Jones,S., Zhang,X., Lin,J.C., Leary,R.J., Angenendt,P., Mankoo,P., Carter,H., Siu,I.M., Gallia,G.L., Olivi,A., McLendon,R., Rasheed,B.A., Keir,S., Nikolskaya,T., Nikolsky,Y., Busam,D.A., Tekleab,H., Diaz,L.A., Jr., Hartigan,J., Smith,D.R., Strausberg,R.L., Marie,S.K., Shinjo,S.M., Yan,H., Riggins,G.J., Bigner,D.D., Karchin,R., Papadopoulos,N., Parmigiani,G., Vogelstein,B., Velculescu,V.E., and Kinzler,K.W. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science*, 321, 1807-1812.
15. Pirooznia,M., Nagarajan,V., and Deng,Y. (2007). GeneVenn - A web application for comparing gene lists using Venn diagrams. *Bioinformation.*, 1, 420-422.
16. Kim,T.M., Yim,S.H., Jeong,Y.B., Jung,Y.C., and Chung,Y.J. (2008). PathCluster: a framework for gene set-based hierarchical clustering. *Bioinformatics*, 24, 1957-1958.
17. Laakso,M. and Hautaniemi,S. (2010). Integrative platform to translate gene sets to networks. *Bioinformatics*, 26, 1802-1803.