Biomarker robustness reveals the P38 network as driving disease outcome in GBM patients in multiple studies and miR-9 as a possible control agent

Rotem Ben-Hamo[1] and Sol Efroni[1,*]

[1]The Mina and Everard Goodman Faculty of Life Science, Bar Ilan University, Ramat-Gan, 52900, Isreal.
[*] sol.efroni@biu.ac.il

Glioblastoma Multiforme (GBM) is the most common, aggressive and malignant primary tumor of the brain and associated with one of the worst 5-year survival rates among all human cancers [1]. This tumor diffusely infiltrates the brain early in its course, making complete resection impossible. Advances in treatment for newly diagnosed GBM have led to the current 5-year survival rates of 9.8%. Despite therapy, once GBM progresses, the outcome is uniformly fatal, with median overall survival historically less than 30 weeks[2].

Cancer is a genomic alterations disease: changes in DNA sequence, epigenetic aberrations in DNA methylation and genomic variations in copy number together scaffold the development and progression of human malignancies, GBM is no different. However, the clinical value of most Glioblastoma-associated molecular aberrations in term of their significant in diagnostic, prognostic, or predictive molecular markers has remained unclear [3]. A better understanding of the molecular characteristics and biology of GBM may help improve treatment and identification of cellular factors that drive prognosis may provide a key for novel treatment.

One of the most comprehensive efforts in molecular characterization of cancer in general and Glioblastoma Multiforme in particular is The Cancer Genome Atlas (TCGA) [4]. The types of data provided through TCGA, for over 370 patients are: expression abundance through microarrays, DNA methylation, copy number variation, and microRNA expression data.

DNA methylation plays an important role in the development of cancer and other diseases due to its ability to control gene-expression. Research has shown that the silencing effect of methylation achieved through the interaction of methylcytosine binding proteins with other structural components of chromatin, which make DNA inaccessible to transcription factors through histone deacetylation and chromatin structure changes [5-7]. Hypermethylation of CpG islands located in the promoter regions of tumor suppressor genes is now firmly established as an important mechanism for gene inactivation in cancers [8]. Somatic copy number variations are extremely common in cancer. Deletions and amplifications contribute to alteration in the expression of tumor suppressor genes and oncogenes. Detection and mapping of copy number abnormalities provides an approach for associating aberrations with disease phenotype and for localizing critical genes [9]. MicroRNAs (miRNAs) are small, endogenous non-coding RNA molecules that contribute to modulating the expression level of specific proteins based on sequence

complementarities with their target mRNA molecules. Their role in many human diseases in general and cancer in particular is well established, and their ability to act both as therapeutic agent and disease prognostic biomarker is what makes them so important to understand [10]. By studying these changes and their versatility, we can find targets for sophisticated therapeutics approaches.

In this work, we analyzed methylation, copy number, microRNA and gene-expression data in more than 370 GBM patients from The Cancer Genome Atlas database, and validation data from two additional datasets obtained from the Gene Expression Omnibus (GEO) database [11] (accession number: GSE4412 [12] and GSE13041 [13]).

Here, we took an approach that utilizes network graph structure and combine it with clinical outcome and identified significant curated pathways that can serve as biomarker for survival. To make use of network graph structure, we applied methods to merge expression data with network knowledge [14]. These methods quantify expression behavior in specific sub-networks (such sub-networks can be specific pathways or any other defined subnetwork) and produce two metrics: network activity and network consistency. In brief, a pathway's activity is a measure of how likely the interactions within a pathway are to be active in the specific sample at hand. A sample's pathway consistency is a measure of the compatibility between gene-expression abundance in that sample and molecular description as it detailed in the pathway's graph. Further details are in the Methods section and in [14].
To apply this network-based methodology, we used gene-expression data from TCGA and made use of expression levels to deduce pathway metrics. Each sample was thus re-represented using its pathway metrics. This representation assigns 579 pathway metric scores (a score for each pathway in the database) to each sample. Interaction and pathway information has been obtained from The National Cancer Institute's Pathway Interaction Database (PID) [15]. We then iterated across the set of samples, using the pathway scores, to assign KM p-values for each of the pathways. This procedure allows us to rank each of the pathways according to their ability to stratify patients into prognosis groups. We then validated this set of pathways within the two additional data sets [12, 13]. ***We found one pathway, p38 signaling mediated by mapkap kinases*** (NCI/Nature), that ***significantly and robustly stratify prognosis in all three datasets*** (figure 1). Importantly, none of the gene members in that pathway, by themselves, show statistical power in survival analysis. In addition, the groups emerged from the pathway had no correlation with any clinical feature which strengthens the hypothesis that this pathway is a core mechanism of the disease.

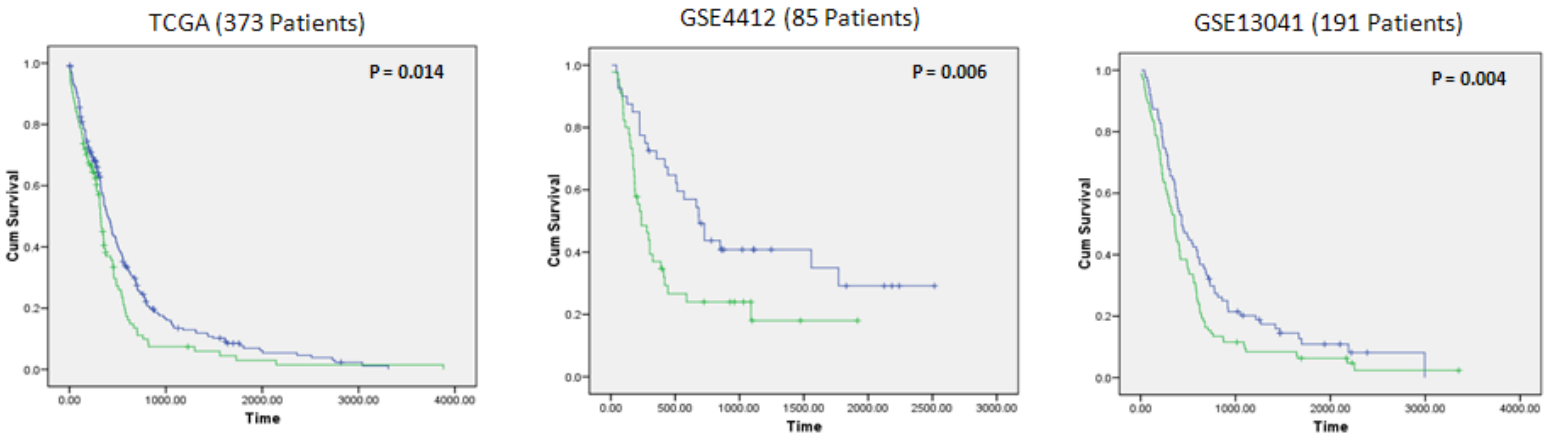

Figure1. Kaplan-Meier curves generated according to values of the P38 pathway across all three datasets. Across three panels, Group1 (blue line), which is affiliated with better prognosis, shows lower pathway activity values and Group2 (green line) shows higher pathway activity values. The affiliation of pathway metric levels with prognosis is highly robust in this case, as it shows low p-values and consistent behavior across datasets.

This pathway, when highly activated, has low survival rates. In addition, previous works found that when this pathway is highly activated it induce migration of glioblastoma cells [16].

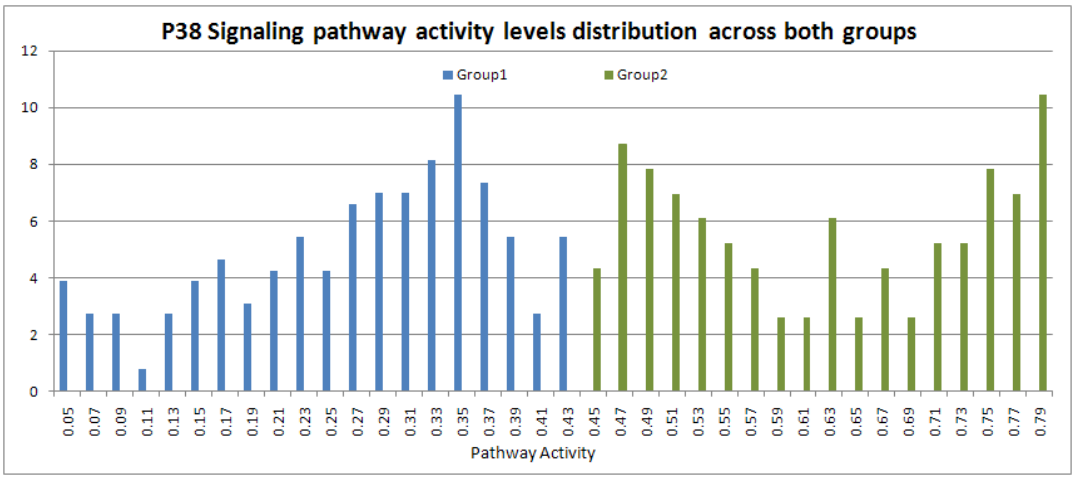

Figure2. P38 Signaling pathway activity levels distribution across groups. Group1 (blue, higher survival rates) has low pathway activity. Group2 (green, lower survival rates) has higher activity levels. This figure demonstrates the large range in the activity levels between groups, and the distinct difference between them.

To further study the molecular characteristics of this pathway, we made use of the intensive molecular features available through TCGA. As mentioned earlier, TCGA avails genetic and epigenetic information for each tumor sample. We analyzed copy number and methylation profiles of the pathway genes. Using Mann-Whitney U test we examine the copy number aberrations in both tumor and matched normal samples in order to see if the copy number

expression levels in tumor and normal for a specific gene are independent samples from identical continuous distributions with equal medians, against the alternative that they do not have equal medians. This analysis revealed that 11 out of the 13 genes in this pathways are highly targeted to copy number changes (p value<0.05) (Table1).

Amplified genes

Gene Symbol	Tumor	Normal
HSP27	21%	2%
CREB1	27%	16%
TCF3	14%	2%
ER81	45%	6%
CDC25B	36%	20%

Deleted genes

Gene Symbol	Tumor	Normal
MAPKAPK3	20%	11%
LSP1	31%	25%
TH	37%	14%
YWHAZ	63%	27%
ALOX5	68%	7%
RAF1	13%	9%

Table1. Eleven out of the 13 genes in the p38 pathway had significant change (according to Mann-Whitney test), in amplification or deletion in copy number between the tumor and its matched normal sample across all patients.

These results reveal that the pathway is highly targeted, and can explain the significant robust connection with patient's outcome. When analyzing the methylation status of the genes in the p38 pathway we found that 4 of the 13 genes in the pathway are consistently methylated across all samples and the remaining 9 genes have no change in their methylation profile.

Next, we combined the activity levels of the pathway with all 1510 microRNA in order to find microRNAs that shows significant correlation with pathway activity and thus can be considered as pathway regulators. Interestingly, we were able to find significant negative correlation (p-value < 0.0001) between the p38 pathway activity levels and hsa-Mir-9*. Deeper examination of the gene sequences revealed that 4 out of the 13 genes in the pathway have a possible binding site to hsa-Mir-9*, this analysis was performed by doing a blast search of the microRNA mature sequence and the 3' UTR region of all 13 genes in the pathway. The fact that this pathway can be targeted by single microRNA in different positions strengthens the hypothesis that this pathway is indeed can act as a robust single biomarker for GBM.

Our results demonstrate that pathway interactions are either affiliated with improved prognosis by "helping" the pathway counter the tumor or poor prognosis by "breaking down" the pathway's normal activity. Through better understanding of the pathway mechanisms and the interactions that undergo changes, we may find targets for new treatments. The fact that the pathway we identified did not correlate with age or tumor diameter and was found in all three datasets strengthens the hypothesis that this pathway is a core mechanism of the disease. The fact that 4 of the 13 genes in the pathway have a possible target sites for a single microRNA that also has a significant correlation with the pathway activity levels can imply on a possible therapeutic agent for maintaining "normal" pathway activity.

References

1. Krex D, Klink B, Hartmann C, von Deimling A, Pietsch T, Simon M, Sabel M, Steinbach JP, Heese O, Reifenberger G *et al*: **Long-term survival with glioblastoma multiforme**. *Brain* 2007, **130**(Pt 10):2596-2606.
2. Prados M, Cloughesy T, Samant M, Fang L, Wen PY, Mikkelsen T, Schiff D, Abrey LE, Yung WK, Paleologos N *et al*: **Response as a predictor of survival in patients with recurrent glioblastoma treated with bevacizumab**. *Neuro Oncol*, **13**(1):143-151.
3. Weller M, Felsberg J, Hartmann C, Berger H, Steinbach JP, Schramm J, Westphal M, Schackert G, Simon M, Tonn JC *et al*: **Molecular predictors of progression-free and overall survival in patients with newly diagnosed glioblastoma: a prospective translational study of the German Glioma Network**. *J Clin Oncol* 2009, **27**(34):5743-5750.
4. **Comprehensive genomic characterization defines human glioblastoma genes and core pathways**. *Nature* 2008, **455**(7216):1061-1068.
5. Razin A: **CpG methylation, chromatin structure and gene silencing - a three-way connection**. *Embo Journal* 1998, **17**(17):4905-4908.
6. Bibikova M, Lin ZW, Zhou LX, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang YH, Vollmer E *et al*: **High-throughput DNA methylation profiling using universal bead arrays**. *Genome Research* 2006, **16**(3):383-393.
7. Bestor TH: **Methylation meets acetylation**.
8. Dehan P, Kustermans G, Guenin S, Horion J, Boniver J, Delvenne P: **DNA methylation and cancer diagnosis: new methods and applications**. *Expert Rev Mol Diagn* 2009, **9**(7):651-657.
9. Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y *et al*: **High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays**. *Nature Genetics* 1998, **20**(2):207-211.
10. Li M, Li J, Ding X, He M, Cheng SY: **microRNA and cancer**. *AAPS J*, **12**(3):309-317.
11. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository**. *Nucleic Acids Res* 2002, **30**(1):207-210.
12. Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, Liau LM, Mischel PS, Nelson SF: **Gene expression profiling of gliomas strongly predicts survival**. *Cancer Res* 2004, **64**(18):6503-6510.
13. Lee Y, Scheck AC, Cloughesy TF, Lai A, Dong J, Farooqi HK, Liau LM, Horvath S, Mischel PS, Nelson SF: **Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age**. *BMC Med Genomics* 2008, **1**:52.
14. Efroni S, Schaefer CF, Buetow KH: **Identification of key processes underlying cancer phenotypes using biologic pathway analysis**. *PLoS ONE* 2007, **2**:e425.
15. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH: **PID: the Pathway Interaction Database**. *Nucleic Acids Res* 2009, **37**(Database issue):D674-679.
16. Nomura N, Nomura M, Sugiyama K, Hamada J: **Phorbol 12-myristate 13-acetate (PMA)-induced migration of glioblastoma cells is mediated via p38MAPK/Hsp27 pathway**. *Biochem Pharmacol* 2007, **74**(5):690-701.