Article

IGFBP2 is a potential master-regulator driving dysregulated gene network responsible for short survival in Glioblastoma multiforme

Manasa KP [1,2], Darius Wlochowitz[1], Edgar Wingender[2], Tim Beißbarth[1], Alexander Kel [2,3] *

1. Department of Medical Bioinformatics, University Medical Center Göttingen, 37099 Göttingen, Germany
2. geneXplain GmbH, 38302 Wolfenbüttel, Germany
3. Institute of Chemical Biology and Fundamental Medicine SBRAS, 630090, Novosibirsk, Russia

* Correspondence: Alexander E. Kel: alexander.kel@genexplain.com

Simple Summary: The present work thrives in identifying the drivers of short survival in glioblastoma multiforme. We identified 5 important regulators – IGFBP2, VEGFA, PDGFA, OSMR and AEBP1 regulating the dysregulated gene networks of short survival. Out of them, IGFBP2 is found to be highly upregulated in short survivors and has established relevance in glioblastoma pathology. Further investigation on gene regulatory network revealed that FRA-1 transcription factor is regulated by IGFBP2 via MEK2/RAF/ERK5 pathway. FRA-1 is reported to be upregulated and to have significant impact on survival in GBM. It is said to dysregulate at-least 50 genes involved in tumor invasiveness in tumor xenografts making it a therapeutic target for GBM intervention. We propose that IGFBP2 drives dysregulated gene network responsible for short survival in GBM via FRA-1 transcription factor.

Abstract

Only 2% of Glioblastoma multiforme (GBM) patients respond to standard care and survive beyond 36 months (long-term survivors, LTS) while the majority survives less than 12 months (short-term survivors, STS). To understand the mechanism leading to poor survival, we analyzed publicly available datasets of 113 STS and 58 LTS. This analysis revealed 198 differentially expressed genes (DEGs) that co-occur with aggressive tumor growth and may be responsible for the poor prognosis. These genes belong largely to the GO-categories “epithelial to mesenchymal transition” and “response to hypoxia”. Promoter and network analysis of the DEGs identified 5 potential master regulators that may explain dysregulation of the DEGs in the STS. The following 5 important master-regulators were identified: IGFBP2, VEGFA, PDGFA, OSMR and AEBP1. It is known that IGFBP2 confers increasing malignancy leading to poor prognosis. However, the molecular mechanism by which IGFBP2 affects disease progression and patient prognosis is unclear. Here we found that IGFBP2 is highly upregulated in short survivors and significantly impact survival. Further investigation of the gene regulatory network revealed that IGFBP2 expression can be regulated by FRA-1 transcription factor via MEK2/RAF/ERK5 pathway. FRA-1 is found to be upregulated and to have significant impact on survival in GBM. It is previously reported that FRA-1 can dysregulate at-least 50 genes involved in tumor invasiveness in tumor xenografts making it a therapeutic target for GBM intervention. We propose that IGFBP2 drives dysregulated gene network responsible for short survival in GBM via FRA-1 transcription factor.

Keywords: Glioblastoma; master regulators; IGFBP2; survival; tumorigenicity; transcription factors

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

1. Introduction

Glioblastoma multiforme (GBM) is the most common, highly malignant primary brain tumor [1]. Despite huge developments in treatment strategies, GBM poses unique treatment challenges due to tumor recurrence (34%) and drug resistance leading to poor survival rates of less than 15 months even after advanced chemoradiotherapy [2]. There are as little as 2% of patients who actually respond to standard care and survive beyond 36 months (3years) [2,3] clinically called as long-term survivors (LTS). Another group termed as short-term survivors (STS) are those who survive less than 12 months [4]. The factors that predict the long survival are not completely known.

Though several factors like age, gender, Karnofsky performance score, extent of tumor resection, radiotherapy, chemotherapy and many more are associated with survival and treatment responses[5–8] , it is evident from recent research that there are certain molecular signatures which might be driving treatment responses and thereby survival. Therefore, understanding these extreme survivor groups at molecular level may shed important light towards biological aspects driving their malignancy and survival. Promoter methylation of the gene MGMT, mutations in the genes IDH1/2 and loss of heterozygosity in chromosome 1p/19q are confirmed to be highly informative about survival and treatment responses [2,3,7,9–12] . Furthermore, "a decreased expression of the genes CHI3L1, FBLN4, EMP3, IGFBP2, IGFBP3, LGALS3, MAOB, PDPN, SERPING1 and TIMP1 have been reported to be associated with prolonged survival" [10,11,13,14]. Understanding these extreme survivor groups at molecular level may shed important light towards biological aspects driving their malignancy and survival.

With the advent of gene expression profiling and remarkable developments in high- throughput technologies, it is possible for us to achieve higher level molecular insights into disease biology. Databases like GEO (www.ncbi.nlm.nih.gov/geo/), Array Express (www.ebi.ac.uk/arrayexpress/) and The Cancer Genome Atlas – TCGA (www.cancer.gov/) serve as open platforms for retrieving high quality multi-omics data to find markers in cancer research. Identification of differentially expressed genes (DEGs) already serves as an important in-silico strategy towards finding potential drivers of cellular state transitions. For a more refined analysis, functional annotation of genes of interest, using a priori known biological categories from the Gene Ontology – GO (www.geneontology.org/) and pathway databases, e.g. TRANSPATH® (http://genexplain.com/transpath/), KEGG (www.genome.jp/kegg/), PANTHER (www.pantherdb.org/) and Reactome (reactome.org) has proven to be an effective hypothesis-driven approach in cancer research. Moreover, with the advent of state of art promoter-analysis it is now possible to establish gene regulatory networks that have been used to understand the causes for gene dysregulation and identifying potential master regulators driving them. In this regard, the Genome-enhancer (my-genome-enhancer.com) incorporates an automated pipeline for such an analysis called "upstream analysis". "This strategy comprises two major steps: (1) analysis of promoters of DEGs to identify transcription factors (TFs) involved in the process under study (done with the help of the TRANSFAC® database" [15] and the binding site identification algorithms, Match [16] and CMA [17]; (2) reconstruction of signaling "pathways that activate these TFs and identification of master-regulators on the top of such pathways (done with the help of the TRANSPATH® signaling pathway database" [18] and special graph search algorithms).

We used two datasets of Affymetrix U133 plus 2.0 platform of NCI Repository for Molecular Brain Neoplasia Data (REMBRANDT) cohort and GSE53733 making up to 113 STS and 58 LTS (explained in detail in the Methods section). Comparative analysis between survival groups was performed to determine DEGs. Gene set enrichment is performed to gain insights into the impact of gene dysregulation. In this paper we applied an upstream analysis algorithm to identify master regulators potentially responsible for short survival in GBM. The gene regulatory network thus developed gives us important insights about dysregulated genes, transcription factors and their master regulators. The then developed regulatory networks are critically analyzed and we propose a key role of insulin-like

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

growth factor-binding protein 2 (IGFBP2) as an important driver of short survival in GBM by driving the dysregulated gene network via FRA-1 transcription factor.

2. Results

2.1. Identification of differentially expressed genes

Identifying the DEGs gives us insight on the biological semantics of a cellular state and helps to identify promising biomarkers of various disease states. The differential gene expression analysis between STS and LTS groups of GBM, from the batch corrected GSE dataset was performed using LIMMA suing FDR cutoff of 5%. The analysis revealed 115 significantly (adj_pvalue <0.05) upregulated (log2FC>0.5) and 83 significantly downregulated (log2FC <–0.5) genes. Top 5 upregulated and down regulated genes and their corresponding log2FC are shown in Table1 and the full list is given in supplementary 1A

Table 1. The list of top 5 upregulated and top 5 downregulated genes identified using LIMMA for differential gene expression analysis between STS and LTS in GSE dataset.

Gene Symbol	Description	log2FC	p-value	adj. p-value
Upregulated genes				
CHI3L1	Chitinase-3-like 1	1.370857	9.73E-05	0.0127
POSTN	Periostin	1.33145	7.88E-07	0.001627
PDPN	Podoplanin	1.241589	6.45E-04	0.028191
MEOX2	Mesenchyme homeobox 2	1.159601	4.87E-05	0.010796
IGFBP2	Insulin-like growth factor-binding protein 2	1.149656	5.79E-05	0.01101
Downregulated genes				
KLRC1	Killer cell lectin-like receptor C1	-1.2187	3.63E-04	0.022188
KLRC2	Killer cell lectin-like receptor C2	-1.2187	3.63E-04	0.022188
FUT9	Fucosyl-transferase 9	-1.0709	1.15E-04	0.013921
DPP10	Dipeptidyl peptidase-like 10	-1.02781	2.97E-05	0.008732
GABRB3	Gamma-aminobutyric acid type A receptor beta3 subunit	-0.96352	6.73E-05	0.011057

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

2.2. Functional annotation of differentially expressed genes

Functional annotation was performed to investigate biological roles of these DEGs. As shown in Figure 1A, the top biological process is extracellular structure organization with counts of 35 DEGs. Figure 1B shows results of cellular component enrichment revealing dysregulation of genes that belong to extracellular matrix and synaptic membrane. The important molecular functions enriched are channel activity and transmembrane transporter activity (Figure 1C). The disruption in extracellular matrix organization is one of the important signatures in glioblastoma treatment response dealing with invasiveness and malignancy [14]. Deeper biological insights are required in this aspect. It is interesting to see enrichment of genes known to be involved in glioblastoma multiforme (Figure 1D). Gene signature enrichment based on hallmark gene sets of MSigDB clearly signifies the enrichment of epithelial to mesenchymal transition depicted as dotplot on Figure 1E. The process of epithelial to mesenchymal transition plays a very important role in GBM survival by driving tumor invasiveness and drug resistance [19]. Important pathways like Aurora signaling, G2/M phase, TGFbeta are found to be enriched Table 2 according to enrichment of TRANSPATH® pathways. Full list in Supplementary 1B

Table 2. Transpath® Pathway (2019.3) enrichment among DEGs in STS vs. LTS

ID (TRANSPATH)	Title	Group size	Expected hits	Nominal P-value	ES	Rank at max	NES	FDR	Number of hits
CH000001004	Aurora-A cell cycle regulation	68	67.26247	0	0.42261	8347	4.13803	0	68
CH000000919	Cyclosome regulatory network	77	76.16486	0	0.348952	7336	3.728426	0	77
CH000000694	G2/M phase (cyclin B: Cdk1)	66	65.28416	0	0.374627	6641	3.587532	0	66
CH000000879	Caspase network	83	82.09978	0	0.333518	8414	3.523174	0	83
CH000000711	TGFbeta pathway	153	151.3406	0	0.2327	8431	3.346102	0	151

2.3. Identifying the master regulators of dysregulated molecular networks

Upstream analysis approach to analyze dysregulated genes, aims to identify the clue about why certain genes are dysregulated (up or down regulated) in the study.

First, we analyzed enrichment of transcription factor binding sites in promoters of differentially expressed genes using DNA-binding motifs listed in the TRANSFAC® library. Supplementary 1C 274

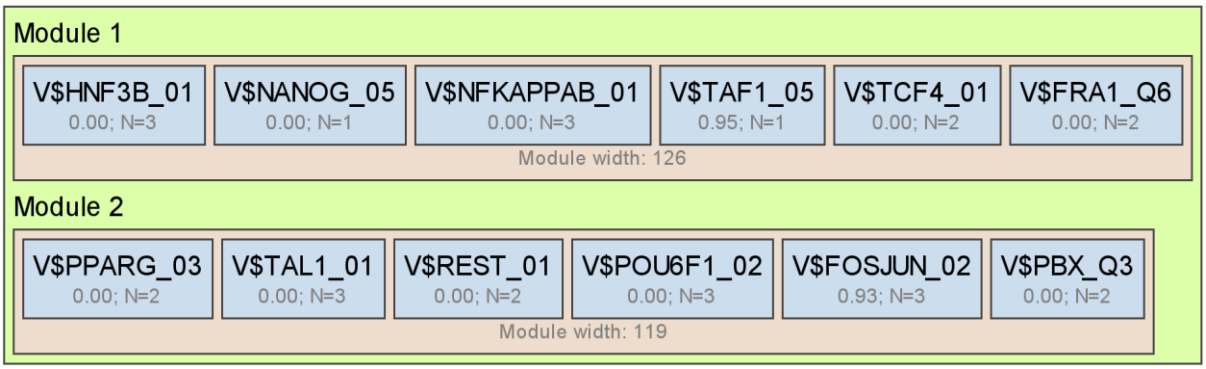
 doi:10.20944/preprints202010.0046.v1

transcription factors enriched for CCKR signaling, interleukin signaling, PDGF signaling, WNT signaling were found to have their binding sites enriched, full enrichment results in supplementary 1D

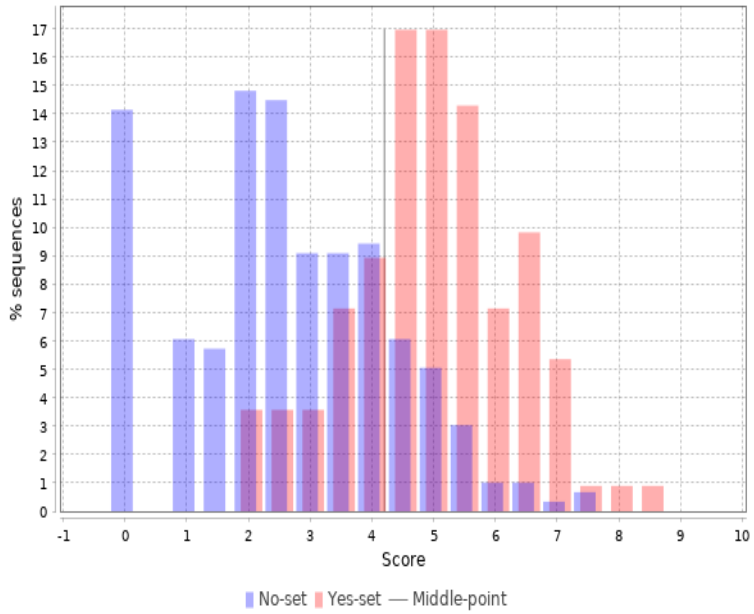
Next, we applied Composite Module Analyst and identified 2 modules involving 12 transcription factor binding site combinations that regulate the expression of genes of interest. CMA revealed the following modules comprising clustering binding sites for the following TFs: Module1: HNF3B, NANOG, NFKAPPAB, TAF1, TCF4 & FRA1; Module2: PPARG, TAL1, REST, POU6F1, FOSJUN & PBX. The modules, transcription factors and their significance are depicted in Figure 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

(A)



(B)



Model score (-p*log10(pval)): 14.01
Wilcoxon p-value (pval): 1.10e-28
Penalty (p): 0.501
Average yes-set score: 4.92
Average no-set score: 2.67
AUC: 0.85
Middle-point: 4.20
False-positive: 17.51%
False-negative: 24.11%

expected for a random set of **regulatory regions**

(C)

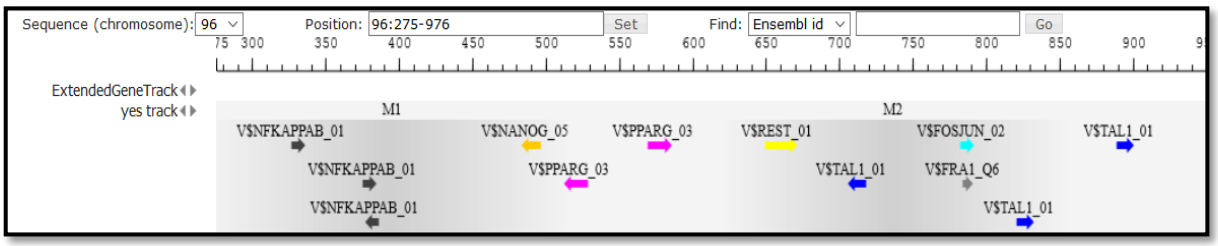


Figure 2. Results of CMA analysis of upregulated genes in short-term survivors. A Combination of 12 PWMs (Position weigh matrices) with their optimized cut-offs identified by genetic algorithm. (cut-off 0.00 means that the algorithm chose the default cut-off; parameter N represent maximal number of top scoring TF binding sites that are used in the module) B The discriminative parameters of the composition of the Composite Score (p-value of the Wilcoxon test, AUC, rates of false positives and false negatives) and two histograms of the distributions of the Composite Score values in Yes and No promoters. C An example of the site location in the promoter of CHI3L1 gene which is usually found upregulated in GBM. The promoter of the gene contains predicted sites for NFkappaB1, FRA1, NANOG and several other transcription factor

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

Finally, we reconstructed the signaling network that activates the TFs revealed by CMA analysis and thereby identifying the top regulators in these networks using TRANSPATH database. The process identified 5 important master regulators that are plausible drivers of short survival in GBM: IGFBP2, VEGFA/VEGF165, PDGFA, AEBP1 and OSMR. All the master regulators were found to be significantly upregulated in short-term survivors. The regulatory network reconstructed along with 6 master regulators is shown in Figure 3, the master regulators and their LogFC in short-term survivors are listed in Table 4. Since VEGF165 is a splice variant of VEGF-A, only the latter will be considered further on.

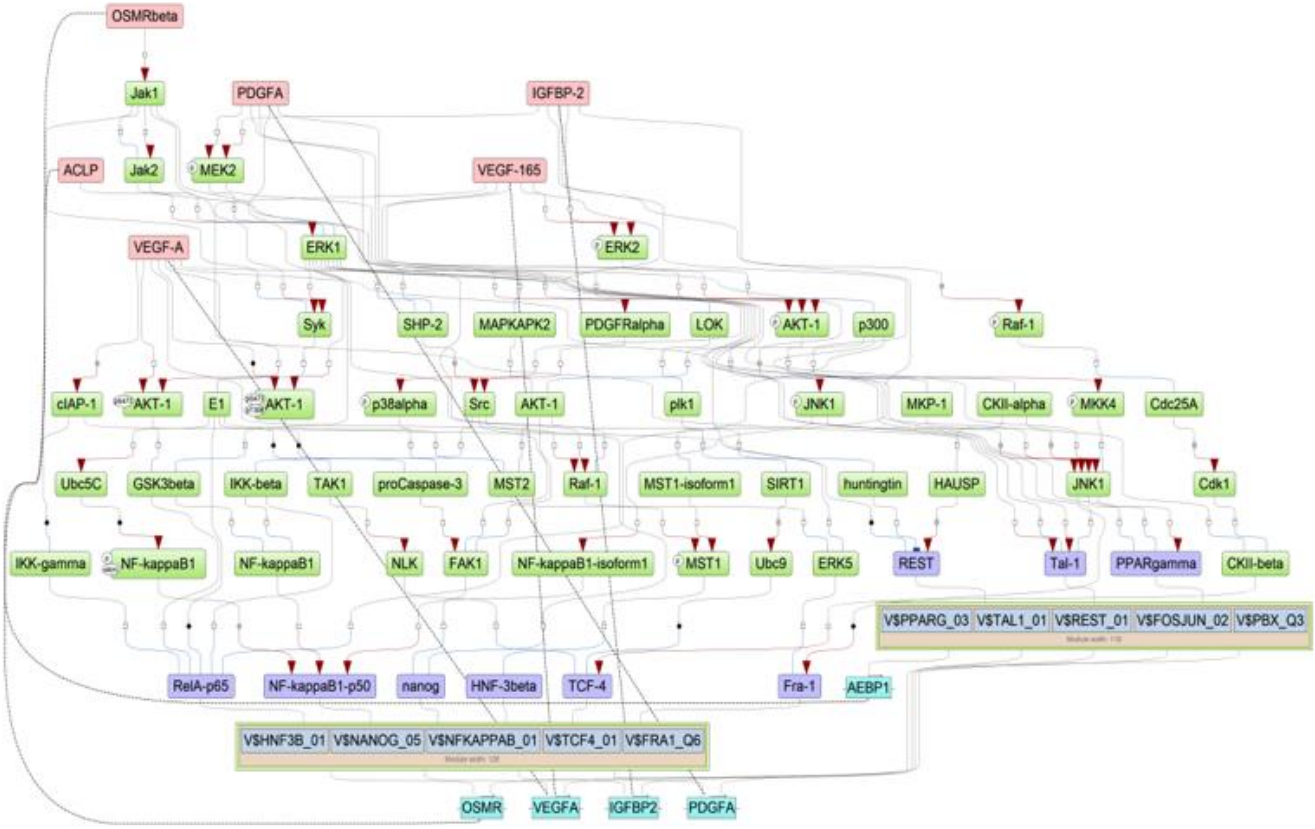


Figure 3. Gene regulatory network of 6 master regulators (red nodes) regulating the 2 transcription factor modules (purple nodes) enriched in promoters in highly upregulated genes of short-term survivors. Shown is the visualization generated by Genome Enhancer. The master regulators identified are IGFBP2, VEGFA, VEGF165 (an VEGF-A splice variant), PDGFA, OSMR and AEBP1.

Table 4. Table of the master regulators identified, their description, LogFC in STS and number of TFs regulated

Molecule Name	Gene Description	HGNC Gene symbol	LogFC in STS	No. of TFs regulated
IGFBP-2	Insulin like growth factor binding protein 2	IGFBP2	1.12	9
ACLP	AE binding protein 1	AEBP1	0.7829	9
VEGF-A	vascular endothelial growth factor A	VEGFA	0.77833	9
VEGF-165	vascular endothelial growth factor A	VEGFA	0.77833	9
OSMRbeta	Oncostatin M receptor	OSMR	0.6345	8
PDGF-A	platelet derived growth factor subunit A	PDGFA	0.52932	9

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

2.4. Validating the expression of master regulators in other cohorts

Expression patterns of the above identified master regulators have been validated in 2 different cohorts. A) TCGA-GBM microarray data and B) GSE16011. The expression patterns were similar and there is a significant upregulation of all master regulators except VEGFA (GSE16011: adj_pvalue=0.069, TCGA-GBM: adj_pvalue=0.075) Supplementary1E,1F. The expression patterns are depicted as boxplots Figure 4

2.5. Impact of master regulators on survival in GBM

Univariate survival analysis is used to study impact of these master regulators and the regulated transcription factors were looked for their impact upon survival in GBM. Tier 3 data belonging to 158 GBM patients of RNA seq data is used for this purpose. Out of 5 unique master regulators, all but one (VEGFA) were found to have a significant impact upon survival with GBM. Additionally, Fra-1 transcription factor, also known as Fos-related antigen 1 (FOSL1), was found to be significant in its impact upon survival Figure 5.

3. Discussion

Gene regulatory networks represent the causal regulatory relationships between transcription factors (TFs) and their gene targets [20] . This will enable us to explain as to why some genes are dysregulated in certain biological state. Comparative studies of the short-term survivors and long-term survivors of GBM showed that gene-expression programs executed across survival groups vary significantly. Using this approach, we identified a set of 12 transcription factors cooperatively binding to the promoters of genes upregulated in short-term survivors. Graph analysis of signal transduction network upstream of these transcription factors allowed us to identify 5 potential master-regulators responsible for such low survival in GBM patients, namely insulin like growth factor binding protein - IGFBP2, vascular endothelial growth factor A - VEGF-A, its isoform VEGF165, platelet Derived growth Factor A- PDGFA, Oncostatin M - OSMR and Adipocyte Enhancer binding protein- AEBP1.

VEGFA is the most important mediator of angiogenesis in glioblastoma. VEGF sequestration, vascular disruption and suppressing VEGFR have been explored as therapeutic strategies in glioblastoma [21]. VEGF/VEGFA is said to promote proliferation of GBM stem like cells [22]. PDGF (here: PDGFA) and PDGF receptors are commonly co-expressed in gliomas and play important role in glial tumorigenesis. PDGFRα, receptor of PDGFA was identified as the "third of the top 11 amplified genes in clinical GBM specimens" and serves as a therapeutic target [23]."OSMR is a critical regulator of glioblastoma tumor growth that orchestrates a feed-forward signalling mechanism with EGFRvIII and STAT3 to drive tumorigenesis" [24].AEBP1 – is a critical oncogene and "both cellular proliferation and survival were affected upon AEBP1 silencing" in glioma cells. AEBP1 overexpression leads to "uncontrollable activation of NF-κB, which may have severe pathogenic outcomes" in GBM [25]

Except VEGFA, all master regulators were found to impact survival according to TCGA-GBM RNA seq data. Master regulators were differentially expressed across the survival groups and the expression patterns are validated in 2 other independent cohorts. IGFBP2 is found to be the most upregulated master regulator driving gene dysregulation network in STS of GBM.

Insulin-like growth factor binding protein 2 (IGFBP2) has emerged as the most potential glioma oncogene and functions as hub of oncogenic signalling pathways by regulating pro-tumorigenic signals of tumor initiation and progression. IGFBP2 predicts an unfavorable prognosis serving as biomarker [26,27] and confers increasing aggressive malignant status[28] . However, the molecular mechanism by which IGFBP-2 affects disease progression and patient prognosis is unclear.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

Pericellular IGFBP2 activates IGFR either by increasing bioavailability of IGFs or by direct interaction with its functional domain. This IGF/IGFR activates RAS-RAF-MAPK signalling and mediates cell survival through PI3K-AKT pathway [29,30].Integrin functions as a receptor for collecting IGFBP2 extracellular signals[31,32]. This is involved in modulating NF-κB signalling to promote glioma progression via Integrin linked kinase (ILK)[28]. IGFBP2 induced activation of NF-κB is said to drive epithelial to mesenchymal transition and invasiveness in certain cancer types[28]."Exogenous IGFBP2 promotes proliferation, invasion, and chemoresistance to temozolomide in glioma cells via integrin β1" by promoting ERK phosphorylation and nuclear translocation [32,33]. PTEN, a tumor suppressor, acts as a checkpoint in IGFBP2 signalling cascade. "IGFBP2 is the most significant marker of PTEN loss and AKT activation" in glioblastoma[34].Both exogenous and cellular IGFBP2 overexpression lead to activation of STAT3 [35].IGFBP2 potentiates STAT3 by activation of nuclear EGFR signalling pathway [35]. "STAT3 and NF-κB are said to be the two major downstream transcriptional factors of IGFBP2 that direct tumorigenic" intracellular signalling [35].Finally, IGFBP2 is capable of nuclear translocation [36] and is involved transcriptional regulation of the VEGF gene and modulates angiogenesis [36]. "Silencing of IGFBP-2 in human glioblastoma cells reduces both progression and invasion" [37]. Additionally, based on our results from gene regulatory analysis, we propose that IGFBP2 can modulate gene dysregulation via FRA-1 transcription factor.

The role of AP1 transcription factor, specially c-Fos and c-Jun is said to be a critical regulator of tumor invasiveness [38].FRA-1, which is a Fos transcription factor and, thus, is an AP-1-related factor is consistently reported to be upregulated in GBM[39]. "Ectopic Fra-1 induced prominent phenotypic changes in malignant glioma cell lines - H4, U-87 MG and A-172 MG". FRA-1 has earlier been reported to be important in maintenance/progression in malignant glioma[40]. In addition, FRA-1 transgene has allowed H4 cells that do not form xenografts of the tumor to recover tumorigenic ability. The genotype of these cells also changed, as 50 of the 1,056 genes examined are dysregulated [41]. Hence, FRA-1 is suggested to be an important target in GBM therapeutic interventions. Debinski et al., (2005) hypothesized that any AP1 stimulating signals like epidermal growth factor (EGF), leukemia inhibitory factor, Oncostatin M, FGF-2 can positively regulate FRA-1. VEGF-D seems to be regulated by FRA-1 and is a known prognostic factor in other aggressive cancers [42,43]

In our data we found FRA-1 to be upregulated in the short-term survivors. The regulatory network showed that FRA-1 transcription factor is regulated by IGFBP2 via MEK2/RAF/ERK5 pathway. IGFBP2 and FRA-1 both are known to regulate transcription of VEGF and VEGFA is found to be the second most dysregulated master regulator in STS. We find that promoters of the genes of all the 5 master regulators reported in the study have binding sites for FRA-1 transcription factor. It is possible that FRA-1 mediates the positive feedback loop that leads to the activation of the master regulators of STS in a cooperative manner. Additionally, it is known that activated ERK signalling, which is triggered by the master-regulators, may lead to mitogen-induced FRA-1 transcription [44] as well as protection from proteasomal degradation [45]. Figure 6. shows how the FRA-1 interacts with five master regulators and their intermediates. Hence, it is possible that FRA-1 and the master regulators are in a cooperative positive feedback loop.

Thus, we propose that IGFBP2 orchestrates a complex pro-tumorigenic program via FRA-1 transcription factor in GBM and serves not only as a prognostic factor but also a therapeutic target Figure 6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

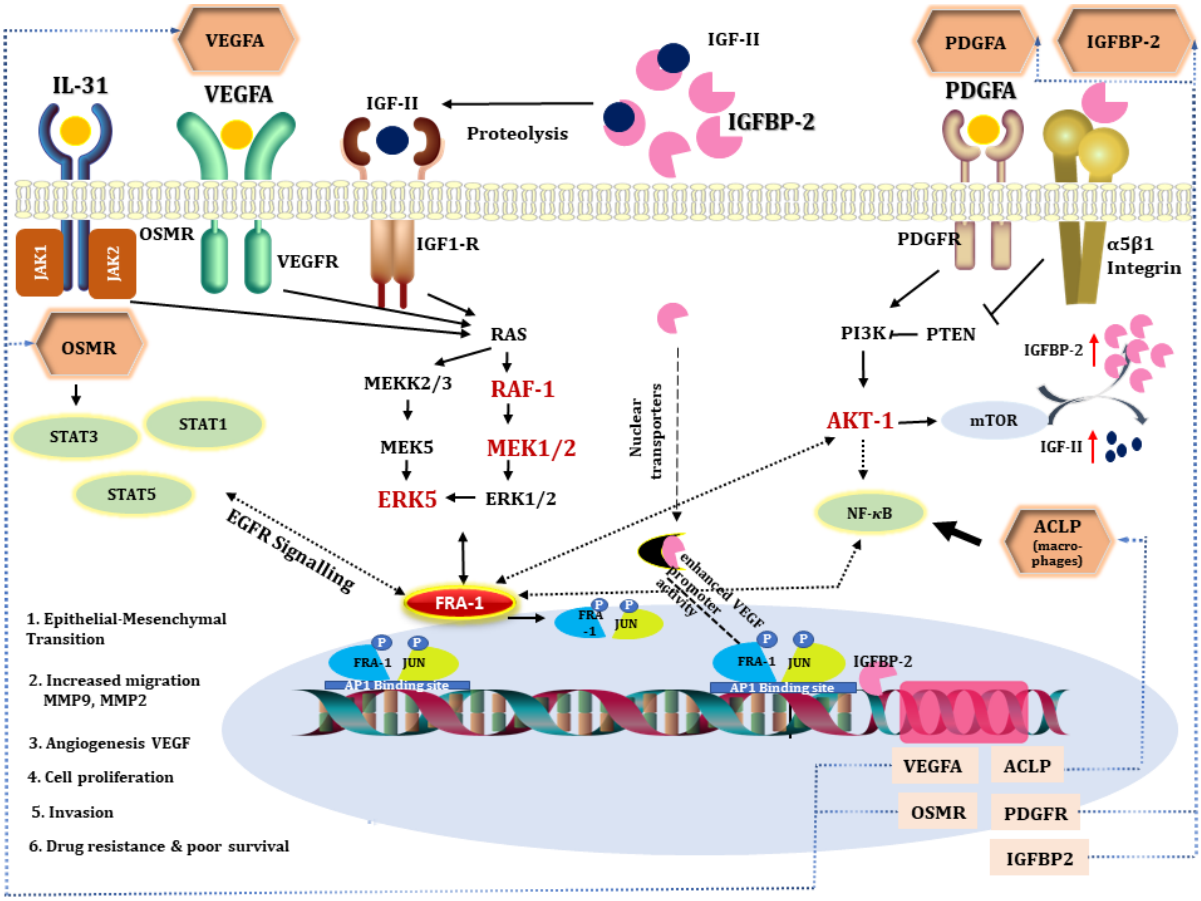


Figure 6. Master regulators, their intermediates and association with FRA-1transcription factor. VEGFA, PDGFA, IGF-II and IL-31 activate RAF/MEK/ERK signalling and mediates cell survival through PI3K-AKT pathway [30]. There is a direct relationship between human GBM patient outcome and both AKT-1 and AKT-2 mRNA levels [46]. MEK2/RAF1/ERK5 and AKT-1 are found to be upregulated in STS. Oncostatin M is a receptor for cytokine IL31 and is involved in enhancing EGFR signalling[24]. IGFBP2 is said to potentiate nuclear translocation of EGFR -STAT3 signalling[47]. Integrin acts as receptor for IGFBP2 extracellular signals [32,33] and PTEN acts as a checkpoint in IGFBP [48] signalling cascade. Enhanced ERK signalling via these master regulators can contribute to drug resistance [49,50]. STAT3 and NF-κB are said to be downstream regulators of IGFBP2 signalling[51]. Nuclear translocation of IGFBP2 via nuclear receptors enhances promoter activity of VEGF [42,43]. All the 5 master regulators have binding site for FRA-1. Hence, it is possible that the FRA-1 and the master regulators are in a positive feedback loop to orchestrate complex tumorigenic program of invasiveness, migration, drug resistance and angiogenesis.

4. Materials and Methods

4.1. Data Collection

The genome wide expression profiles based on Human Genome U133 plus 2.0 array of patients with GBM were collected from public repository of GEO database – GSE108474 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108474) and GSE53733 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53733). The 2 datasets were pooled together leading to 113 and 58 samples corresponding to short-term survivors (STS; survival < 12 months) and long-term survivors (LTS; survival > 36 months) with GBM, respectively Table 5.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

Table 5. Statistics of datasets studied in this work. The datasets with labels GSE' were collected from GEO database.

	Platform	**Short-term survivors**	**Long-term survivors**
GSE53733	HU133 plus 2.0 arrays	16	23
GSE108474	HU133 plus 2.0 arrays	97	35

4.2. Affymetrix microarray data pre-processing

The raw data files (. CEL format) for GSE108474 and GSE53733 were collected from GEO database- from here on called as GSE dataset. RMA algorithm is used in R (affy package) for background correction, quality check and normalization to obtain log2 transformed expression values [52]. Batch correction of the pooled expression data was performed using empirical Bayes framework is performed [53] . This batch corrected file is used for further analysis. Multiple Affymetrix ids were summarized to genes ids by choosing the maximum out of probe intensities of multiple probes belonging to single gene. The final expression matrix comprised 21526 probes and 171 samples.

Differential gene expression (DEG) analysis

LIMMA (Linear Models for Microarray Data) method was applied to identify differentially expressed genes [54]. It is an efficient tool which is stable even for experiments with small samples. Differential gene expression analysis of 171 samples of GSE dataset was performed with Benjamini-Hochberg adjusted P_Value. Only genes with $|\log2 \text{ fold change (FC)}| \geq 0.5$ and adjusted P values < 0.05 were selected as the DEGs for subsequent analysis.

4.3. Databases used in the study

Transcription factor binding sites in promoters and enhancers of differentially expressed genes were analyzed using known DNA-binding motifs described in the TRANSFAC® library, release 2019.3 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transfac)[15]. The master regulator search uses the TRANSPATH® database (BIOBASE), release 2019.3 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transpath)[18]. A comprehensive signal transduction network of human cells is built by the software based on reactions annotated in TRANSPATH®. The information about drugs corresponding to identified drug targets and clinical trials references were extracted from HumanPSD™ database, release 2020.2 (https://genexplain.com/humanpsd). The Ensembl database release Human99.38 (hg38) (http://www.ensembl.org) was used for gene IDs representation and Gene Ontology (GO) (http://geneontology.org) was used for functional classification of the studied gene set.

4.4. Functional Annotation

To explore the biological importance of DEGs, gene set enrichment analysis is performed. GSEA is an efficient method to determine whether the genes of interest show statistically significant enrichment between different biological states. Gene ontology enrichments for cellular component,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

biological process and molecular functions were performed. To investigate the top enriched ontology terms 1000 random permutations were done and adjusted p-value cutoff of 0.05 is used. The dysregulated gene networks enrichment also gives useful insight about known disease signatures [55]. The hallmark gene set of MsigDB defines specific biological states or processes [56]. Enrichment analysis is performed in R using DOSE package [57]. PANTHER pathway enrichment of the identified transcription factors was performed using EnrichR tool [58]. TRANSPATH pathway enrichment was performed using geneXplain platform [18].

4.5. Genome Enhancer pipeline

The approaches mentioned above helps us in understanding the impact of the differentially expressed genes in GBM biology. To understand the reason behind this dysregulation, Genome enhancer pipeline of geneXplain is used (my-genome-enhancer.com). Significantly upregulated genes in STS were used in this workflow.

It works in 2 steps.

A. Analysis of enriched transcription factor binding sites and composite modules

Binding of transcription factors to the transcription factor binding sites in promoters and like enhancers is key to mediation of transcriptional regulation of genes. Classically, enhancers are defined as regions in the genome that increase transcription of one or several genes when inserted in either orientation at various distances upstream or downstream of the gene [17]. Enhancers typically have a length of several hundreds of nucleotides and are bound by multiple transcription factors in a cooperative manner [18].

Identifying such clusters of binding sites for such cooperating transcription factors (composite-modules) that act as potential condition-specific enhancers of the target genes in their upstream regulatory regions (-1000 bp upstream of transcription start site (TSS)) and the transcription factors which regulate the genes through such enhancers is a determining step to understand regulatory mechanism. PANTHER pathway enrichment of the above identified transcription factors were identified using EnrichR tool [58].

We use Composite Module Analyst (CMA) [17] method to detect such potential enhancers, as targets of multiple TFs bound in a cooperative manner to the regulatory regions of the genes of interest. CMA applies a genetic algorithm to construct a generalized model of the enhancers by specifying combinations of TF motifs (from TRANSFAC®) whose sites are most frequently clustered together in the regulatory regions of the studied genes. CMA identifies the transcription factors that through their cooperation provide a synergistic effect and thus have a great influence on the gene regulation process.

B. Finding master regulators in networks

The second step involves the signal transduction database TRANSPATH® and special graph search algorithms to identify common regulators of the revealed transcription factors. These master regulators appear to be the key candidates for therapeutic targets as they have a master effect on regulation of intracellular pathways that activate the pathological process of our study.

The tool also generates a visualization output of selected master regulators with possibility to map the LogFC, p value on the created regulatory network.

4.6. Validation of observed gene signatures

The raw TCGA HT-Hg_U133 microarray data of 560 GBM samples were downloaded from TCGA legacy archive. The GSE16011 raw. CEL data was downloaded from GEO repository. Both raw datasets were processed and analysed independently following same steps as mentioned earlier. These 2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 October 2020 doi:10.20944/preprints202010.0046.v1

datasets are used to observe and validate the expression pattern of master regulators across the two survival groups (see Table 6).

Table 6. Statistics of two validation datasets.

Datasets	Platform	Short-term survivors	Long-term survivors
GSE16011	HU133 plus 2.0 arrays	93	16
TCGA-GBM microarray	HU133	271	49

4.7. Impact on survival

OncoLnc (http://www.oncolnc.org/)[59] is a web-based interactive tool to explore survival correlations and to retrieve clinical data matched with expression data for mRNAs, miRNAs, and long non-coding RNAs saha et.al,2020. Master regulators and their target transcription factors affect the whole regulatory network and therefore can have an independent impact on survival in GBM patients. RNA sequencing data for 152 TCGA GBM cohort in OncoLnc is been used to understand the impact of individual master regulator on survival in GBM.

5. Conclusion

In the work presented, we have established probable master regulators responsible for gene dysregulation in short-term survivors of GBM. These genes have sufficient experimental evidence towards their role in GBM. Out of the reported 5 master regulators, IGFBP2 is established as the most promising master regulator. Through the gene regulatory network analysis, we also propose that FRA-1 is a plausible downstream regulator of IGFBP2-induced signalling in short-term survivors of GBM.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 766069.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wen, P. Y. & Kesari, S. Malignant gliomas in adults. *New England Journal of Medicine* vol. 359 492–507 (2008).
2. Krex, D. *et al.* Long-term survival with glioblastoma multiforme. *Brain* **130**, 2596–2606 (2007).
3. Das, P. *et al.* A clinicopathological and molecular analysis of glioblastoma multiforme with long-term survival. *J. Clin. Neurosci.* **18**, 66–70 (2011).
4. Shinawi, T. *et al.* DNA methylation profiles of long- and short-term glioblastoma survivors. *Epigenetics* **8**, 149–156 (2013).
5. Sonoda, Y. *et al.* Long-term survivors of glioblastoma: Clinical features and molecular analysis. *Acta Neurochir. (Wien).* **151**, 1349–1358 (2009).
6. Scott, J. N. *et al.* Which glioblastoma multiforme patient will become a long-term survivor? A

population-based study. *Ann. Neurol.* **46**, 183–188 (1999).

7. Zhang, X., Zhang, W., Cao, W. D., Cheng, G. & Zhang, Y. Q. Glioblastoma multiforme: Molecular characterization and current treatment strategy (Review). *Experimental and Therapeutic Medicine* vol. 3 9–14 (2012).

8. Lee, Y. *et al.* Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age. *BMC Med. Genomics* **1**, 52 (2008).

9. Chen, J. R., Yao, Y., Xu, H. Z. & Qin, Z. Y. Isocitrate dehydrogenase (IDH)1/2 mutations as prognostic markers in patients with glioblastomas. *Med. (United States)* **95**, (2016).

10. Franceschi, S. *et al.* Investigating molecular alterations to profile short- and long-term recurrence-free survival in patients with primary glioblastoma. *Oncol. Lett.* **10**, 3599–3606 (2015).

11. Han, S., Meng, L., Han, S., Wang, Y. & Wu, A. Plasma IGFBP-2 Levels after Postoperative Combined Radiotherapy and Chemotherapy Predict Prognosis in Elderly Glioblastoma Patients. *PLoS One* **9**, e93791 (2014).

12. Reifenberger, G. *et al.* Molecular characterization of long-term survivors of glioblastoma using genome- and transcriptome-wide profiling. *Int. J. Cancer* **135**, 1822–1831 (2014).

13. Bi, W. L. & Beroukhim, R. Beating the odds: Extreme long-term survival with glioblastoma. *Neuro-Oncology* vol. 16 1159–1160 (2014).

14. De Vega, S., Iwamoto, T. & Yamada, Y. Fibulins: Multiple roles in matrix structures and tissue functions. *Cellular and Molecular Life Sciences* vol. 66 1890–1902 (2009).

15. Wingender, E., Dietze, P., Karas, H. & Knüppel, R. TRANSFAC: A database on transcription factors and their DNA binding sites. *Nucleic Acids Research* vol. 24 238–241 (1996).

16. Kel, A. E. *et al.* MATCH™: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* **31**, 3576–3579 (2003).

17. Waleev, T. *et al.* Composite Module Analyst: Identification of transcription factor binding site combinations using genetic algorithm. *Nucleic Acids Res.* **34**, W541 (2006).

18. Krull, M. *et al.* TRANSPATH®: An integrated database on signal transduction and a tool for array analysis. *Nucleic Acids Research* vol. 31 97–100 (2003).

19. Iwadate, Y. Epithelial-mesenchymal transition in glioblastoma progression. *Oncology Letters* vol. 11 1615–1620 (2016).

20. Marbach, D. & Stolovitzky, G. Robert Küffner 6,11 , Nicole M Vega 3-5. *Robert J Prill* **11**, (2012).

21. Reardon, D. A., Wen, P. Y., Desjardins, A., Batchelor, T. T. & Vredenburgh, J. J. Glioblastoma multiforme: An emerging paradigm of anti-VEGF therapy. *Expert Opinion on Biological Therapy* vol. 8 541–553 (2008).

22. Xu, C., Wu, X. & Zhu, J. VEGF promotes proliferation of human glioblastoma multiforme stem-like cells through VEGF receptor 2. *Sci. World J.* **2013**, (2013).

23. Martinho, O. & Reis, R. M. Malignant Gliomas: Role of Platelet-Derived Growth Factor Receptor A (PDGFRA). in *Tumors of the Central Nervous System, Volume 1* 109–118 (Springer Netherlands, 2011). doi:10.1007/978-94-007-0344-5_12.

24. Jahani-As, A. *et al.* Control of glioblastoma tumorigenesis by feed-forward cytokine signaling. *Nat. Neurosci.* **19**, 798–806 (2016).

25. Majdalawieh, A. F., Massri, M. & Ro, H. S. AEBP1 is a Novel Oncogene: Mechanisms of Action and Signaling Pathways. *Journal of Oncology* vol. 2020 (2020).

26. McDonald, K. L. *et al.* IQGAP1 and IGFBP2. *J. Neuropathol. Exp. Neurol.* **66**, 405–417 (2007).

27. Phillips, L. M. *et al.* Glioma progression is mediated by an addiction to aberrant IGFBP2 expression and can be blocked using anti-IGFBP2 strategies. *J. Pathol.* **239**, 355–364 (2016).

28. Holmes, K. M. *Elucidating the IGFBP2 signaling pathway in glioma development Elucidating the IGFBP2 signaling pathway in glioma development and progression and progression*. https://digitalcommons.library.tmc.edu/utgsbs_dissertations (2012).

29. Subramani, R. *et al.* Targeting insulin-like growth factor 1 receptor inhibits pancreatic cancer growth and metastasis. *PLoS One* **9**, 97016 (2014).

30. Yao, X., Sun, S., Zhou, X., Guo, W. & Zhang, L. IGF-binding protein 2 is a candidate target of therapeutic potential in cancer. *Tumor Biology* vol. 37 1451–1459 (2016).

31. Firth, S. M., McDougall, F., McLachlan, A. J. & Baxter, R. C. Impaired blockade of insulin-like growth factor I (IGF-I)-induced hypoglycemia by IGF binding protein-3 analog with reduced ternary complex-forming ability. *Endocrinology* **143**, 1669–1676 (2002).

32. Schütt, B. S., Langkamp, M., Rauschnabel, U., Ranke, M. B. & Elmlinger, M. W. Integrin-mediated action of insulin-like factor binding protein-2 in tumor cells. *J. Mol. Endocrinol.* **32**, 859–868 (2004).

33. Yau, S. W., Azar, W. J., Sabin, M. A., Werther, G. A. & Russo, V. C. IGFBP-2 - taking the lead in growth, metabolism and cancer. *J. Cell Commun. Signal.* **9**, 125–142 (2015).

34. Li, T. *et al.* IGFBP2: integrative hub of developmental and oncogenic signaling network. *Oncogene* vol. 39 2243–2257 (2020).

35. Phillips, L. M. *et al.* Glioma progression is mediated by an addiction to aberrant IGFBP2 expression and can be blocked using anti-IGFBP2 strategies. *J. Pathol.* **239**, 355–364 (2016).

36. Azar, W. J. *et al.* IGFBP-2 enhances VEGF gene promoter activity and consequent promotion of angiogenesis by neuroblastoma cells. *Endocrinology* **152**, 3332–3342 (2011).

37. Fukushima, T., Tezuka, T., Shimomura, T., Nakano, S. & Kataoka, H. Silencing of insulin-like growth factor binding protein-2 (IGFBP-2) in human glioblastoma cells reduces both invasiveness and expression of progression-associated gene CD24. (2007) doi:10.1074/jbc.M609567200.

38. Jochum, W., Passegué, E. & Wagner, E. F. AP-1 in mouse development and tumorigenesis. *Oncogene* vol. 20 2401–2412 (2001).

39. Kesari, S. & Bota, D. A. Fos-related antigen-1 (Fra-1) is a regulator of glioma cell malignant phenotype. *Cancer Biology and Therapy* vol. 11 307–310 (2011).

40. Debinski, W. & Gibo, D. M. Fos-related antigen 1 modulates malignant features of glioma cells. *Mol. Cancer Res.* **3**, 237–249 (2005).

41. Debinski, W. & Gibo, D. M. Fos-related antigen 1 modulates malignant features of glioma cells. *Mol. Cancer Res.* **3**, 237–249 (2005).

42. Debinski, W. *et al.* VEGF-D is an X-linked/AP-1 regulated putative onco-angiogen in human glioblastoma multiforme. *Mol. Med.* **7**, 598–608 (2001).

43. Azar, W. J., Zivkovic, S., Werther, G. A. & Russo, V. C. IGFBP-2 nuclear translocation is

mediated by a functional NLS sequence and is essential for its pro-tumorigenic actions in cancer cells. *Oncogene* **33**, 578–588 (2014).

44. Adiseshaiah, P., Peddakama, S., Zhang, Q., Kalvakolanu, D. V. & Reddy, S. P. Mitogen regulated induction of FRA-1 proto-oncogene is controlled by the transcription factors binding to both serum and TPA response elements. *Oncogene* **24**, 4193–4205 (2005).

45. Vial, E. & Marshall, C. J. Elevated ERK-MAP kinase activity protects the FOS family member FRA-1 against proteasomal degradation in colon carcinoma cells. *J. Cell Sci.* **116**, 4957–4963 (2003).

46. Joy, A. *et al.* The role of AKT isoforms in glioblastoma: AKT3 delays tumor progression. *J. Neurooncol.* **130**, 43–52 (2016).

47. Chua, C. Y. *et al.* IGFBP2 potentiates nuclear EGFR-STAT3 signaling. *Oncogene* **35**, 738–747 (2016).

48. Li, T. *et al.* IGFBP2: integrative hub of developmental and oncogenic signaling network. *Oncogene* vol. 39 2243–2257 (2020).

49. Salaroglio, I. C., Mungo, E., Gazzano, E., Kopecka, J. & Riganti, C. ERK is a pivotal player of chemo-immune-resistance in cancer. *International Journal of Molecular Sciences* vol. 20 (2019).

50. Abrams, S. L. *et al.* The Raf/MEK/ERK pathway can govern drug resistance, apoptosis and sensitivity to targeted therapy. *Cell Cycle* **9**, 1781–1791 (2010).

51. Chua, C. Y. *et al.* IGFBP2 potentiates nuclear EGFR-STAT3 signaling. *Oncogene* **35**, 738–747 (2016).

52. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).

53. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

54. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).

55. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

56. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

57. Yu, G., Wang, L.-G., Yan, G.-R. & He, Q.-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608–609 (2015).

58. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, (2013).

59. Anaya, J. OncoLnc: Linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *PeerJ Comput. Sci.* **2016**, (2016).