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

Abstract: Only two percent of Glioblastoma multiforme (GBM) patients respond to standard care and survive beyond 36 months (long-term survivors, LTS) while the majority survive 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 characterize 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". In this paper we applied upstream analysis approach which involves state-of-art promoter analysis and network analysis of the dysregulated genes potentially responsible for short survival in GBM. Binding sites for transcription factors associated with GBM pathology like NANOG, NF-κB, REST, FRA-1, PPARG and seven others were found enriched in the promoters of the dysregulated genes. We reconstructed the gene regulatory network with several positive feedback loops controlled by five master regulators – IGFBP2, VEGFA, VEGF165, PDGFA, AEBP1 and OSMR which can be proposed as biomarkers and as therapeutic targets for enhancing GBM prognosis. Critical analysis of this gene regulatory network gives insights on mechanism of gene regulation by IGFBP2 via several transcription factors including the key molecule of GBM tumor invasiveness and progression, FRA-1. All the observations are validated in independent cohorts and their impact on overall survival is studied.

Keywords: Glioblastoma; master regulators; upstream analysis; IGFBP2; FRA-1; FOSL1;short term survivors; transcription factors

1. Introduction

Glioblastoma multiforme (GBM) is the most common, highly malignant primary brain tumor [2]. 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 [3]. There are as little as 2% of patients who actually respond to standard care and survive beyond 36 months (3years) [3,4] clinically called as long-term survivors (LTS). Another group termed as short-term survivors (STS) are those who survive less than 12 months [5]. 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, and chemotherapy are associated with survival and treatment responses[6–9] , 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

[3,4,8,10–13] . Furthermore, CHI3L1, FBLN4, EMP3, IGFBP2, IGFBP3, LGALS3, MAOB, PDPN, SERPING1 and TIMP1 gene expression is decreased in patients with long survival [11,12,14,15]. 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[16],Array Express[17] and The Cancer Genome Atlas – TCGA[18] 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[19] – GO and pathway databases e.g. TRANSPATH® [20],KEGG[21],PANTHER[22]and Reactome[23] 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, we applied Genome-Enhancer (https://genexplain.com/genome-enhancer/), multi-omics analysis TOOL which makes use of an open source BioUML[24] programming environment and incorporates an automated pipeline for such an analysis called "upstream analysis"[25,26] and newest advanced approach called "walking pathways"[27]. "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"[20,28] and the binding site identification algorithms, Match[29,30] and CMA [31]; (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" [20] and special graph search algorithms) that are characterized by positive feedback loops[27].

In this paper we applied upstream analysis on publicly available datasets of GBM - Molecular Brain Neoplasia Data (REMBRANDT) cohort and GSE53733 to understand the gene-regulatory networks driving short-survival in GBM. This regulatory network revealed set of 12 transcription factors binding at the regulatory regions of genes of interest and 5 master regulators regulating them, namely - (a) VEGFA, mediator of angiogenesis[32], promoter of stem like cells in GBM, (b) PDGF – highly amplified gene[33] and key player of tumorigenesis, (c) OSMR –that orchestrates feed-forward signalling with EGFR and STAT3[34] to regulate tumor growth, (d) AEBP1 which has key role in pathogenesis via NF-κB activation[35] and (e) IGFBP2.

IGFBP2 had higher expression in STS, and was found to have an impact on overall survival as well as an established molecule of interest in GBM. IGFBP2 expression is higher in all GBM subtypes35, but is said to drive gene programs for immunosuppression in mesenchymal subtype and is suggested as an immunotherapeutic target[36] , whereas modulates cell proliferation in non-mesenchymal (Classical, Proneural and neural) subtypes[37,38]. It is also found to be a marker of aggressive behavior and prognostic marker for survival[39]. However, the molecular mechanism by which IGFBP-2 affects disease progression and patient prognosis is not yet clear.

The current work focuses on understanding gene regulatory networks which drive short-survival in GBM and their master regulators which can potentially act as biomarkers and therapeutic targets. Later, we critically discuss the role of IGFBP2 in gene regulatory network.

2. Results

2.1. Identification of differentially expressed genes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

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[40] with FDR cutoff of 5%. The analysis revealed 957 genes which are significantly differentially expressed (DEGs) (adj.*p*-value<0.05).The analysis revealed 115 significantly (adj.*p*-value<0.05) upregulated (log2FC>0.5) and 83 significantly downregulated (log2FC<(−0.5)) genes. Top 5 upregulated and downregulated genes and their corresponding log2FC are shown in Table1 and the full list is given in **Table S1-A**.

Table 1. The list of top 5 significantly upregulated and downregulated genes in STS identified in the GSE dataset.

Gene Symbol	Description	log2FC	*p*-value	adj. *p*-value
Upregulated genes				
CHI3L1	Chitinase-3-like 1	1.371	9.73E-05	0.013
PDPN	Podoplanin	1.241	7.88E-07	0.002
MEOX2	Mesenchyme homeobox 2	1.159	6.45E-04	0.028
IGFBP2	Insulin-like growth factor-binding protein 2	1.149	4.87E-05	0.010
COL6A2	Collagen Type VI Alpha 2 Chain	1.0479	5.79E-05	0.011
Downregulate genes				
KLRC2	Killer cell lectin-like receptor C2	-1.2187	3.63E-04	0.022
KLRC1	Killer cell lectin-like receptor C1	-1.2187	3.63E-04	0.022
FUT9	Fucosyl-transferase 9	-1.0709	1.15E-04	0.014
DPP10	Dipeptidyl peptidase-like 10	-1.02781	2.97E-05	0.008
GABRB3	Gamma-aminobutyric acid type A receptor beta3 subunit	-0.96352	6.73E-05	0.011

2.2. Functional annotation of differentially expressed genes

Functional annotation was performed to investigate biological roles of these DEGs. As shown in **Figure S1A**, the top GO Biological process are extracellular structure and matrix organization with 30 DEG hits. **Figure S1B** shows the results for GO Cellular Component enrichment which revealed dysregulation of genes that belong to extracellular matrix and synaptic membrane. The important molecular function GO terms enriched are channel activity and transmembrane transporter activity **(Figure S1C).** The disruption in extracellular matrix organization is one of the important signatures in glioblastoma treatment response dealing with invasiveness and malignancy [15]. Deeper biological insights are required in this aspect. It is interesting to see enrichment of genes known to be involved in glioma (**Figure 1A**). Gene signature enrichment based on hallmark gene sets of MSigDB clearly signifies the enrichment of epithelial to mesenchymal transition depicted in **Figure 1B**. The process of epithelial to mesenchymal transition plays a very important role in GBM survival by driving tumor invasiveness and drug resistance [41]. Important pathways like Aurora signaling, G2/M phase transition, TGF-β are

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

found to be enriched in Table 2 according TRANSPATH® pathways. The full list of enrichment results can be found in **Table S1-B**.

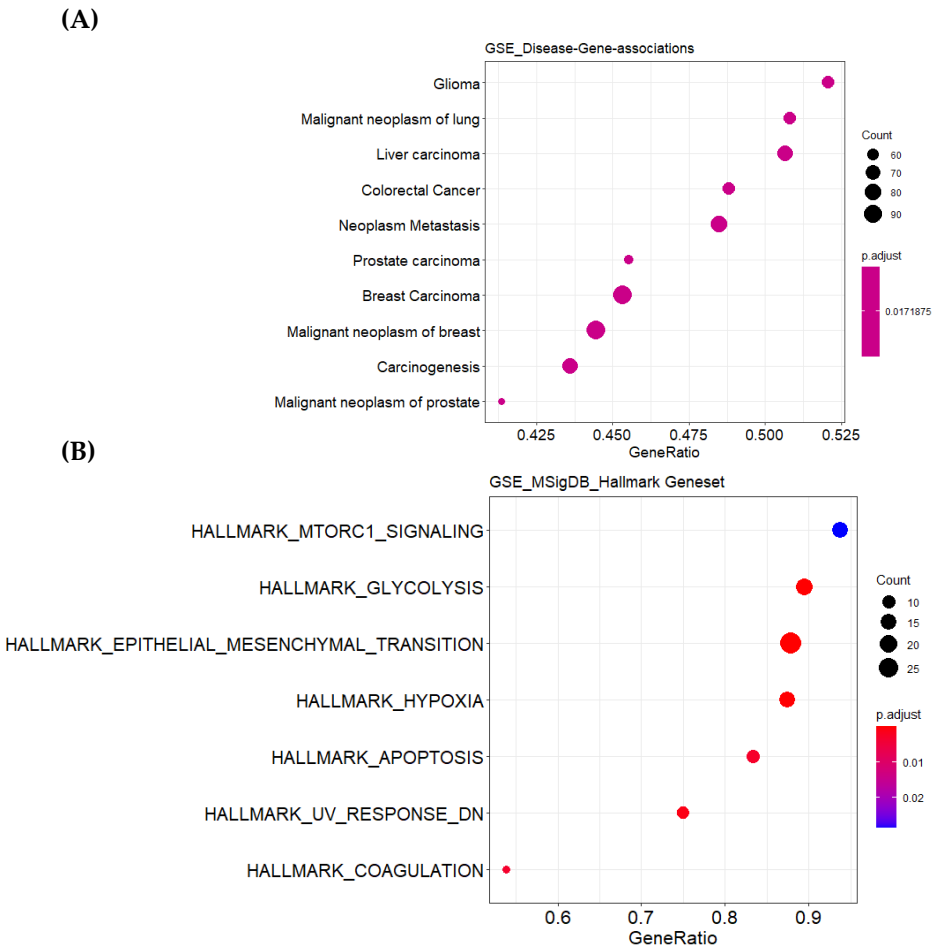

Figure 1: Functional enrichment analysis of DEGs (A) Enrichment for known Disease-Gene Networks in different diseases. Y-axis represents enriched ontology categories and X-axis represents GeneRatio. Gene ratio is count/set-size.The 'count' is the number of genes that belong to a given gene set, while 'set-size' is the total number of genes in the gene set. Y-axis is sorted based on Leading-edge. Leading edge is subset of genes which contribute most to the Enrichment Score. The dots are sized based on gene ratio and are coloured according to their adj.*p*-value. (B) Enrichment for Hallmark gene sets in the Molecular signature database similar to A.

Table 2: Pathway enrichment using TRANSPATH® Pathway (2019.3) for differentially expressed genes

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	67.262	0	0.422	8347	4.138	0	68
CH000000919	Cyclosome regulatory network	77	76.164	0	0.349	7336	3.728	0	77
CH000000694	G2/M phase (cyclin B: Cdk1)	66	65.284	0	0.375	6641	3.587	0	66
CH000000879	Caspase network	83	82.099	0	0.333	8414	3.523	0	83
CH000000711	TGFbeta pathway	153	151.340	0	0.232	8431	3.346	0	151

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

2.3. Identifying the master regulators of dysregulated gene networks

Reconstruction of the disease-specific regulatory networks can help to identify potential master regulators which may serve as mechanism based biomarkers or as therapeutic targets to block a specific pathological regulatory cascade. Using the gene regulatory analysis as a first step we analyzed enrichment of transcription factor binding sites in promoters of upregulated genes of STS using DNA-binding motifs listed in the TRANSFAC® library. 274 transcription factors **(Table S1-C)** enriched for CCKR signaling, interleukin signaling, PDGF signaling, WNT signaling were found to have their binding sites enriched, full enrichment results can be found in **Table S1-D.**

Next, we applied the Composite Module Analyst (CMA) and identified–two 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, FRA-1; Module2: PPARG, TAL1, REST, POU6F1, FOSJUN & PBX. The modules, transcription factors and their significance are depicted in **Figure S2**. Differential expression statistics for the 12 transcription factors are given in **Table S2**. Among them, FRA-1 transcription factor (also known as FOSL1) was found to be *p*-value significant and upregulated in short-term survivors of GBM (log2FC=0.023, *p*-value = 0.008, adj.*p*-value =0.093) **Table S2.**

Figure 2 validates the predicted cluster of TF binding sites from the composite modules identified in the promoter of IGFBP2 gene. We can see that four binding sites in this cluster (for TFs: c-Fos/c-Jun, Nanog, Tal-1, HNF3/FoxA1) can be confirmed by publicly available ChIP-seq data of GTRD database[1]. In addition, Fra-1 binding site can be confirmed by a cluster of mapped reads of independent publicly available ChIP-seq data (FRA1 track on the Figure 2). (Full map is shown in the **Figure S4**).

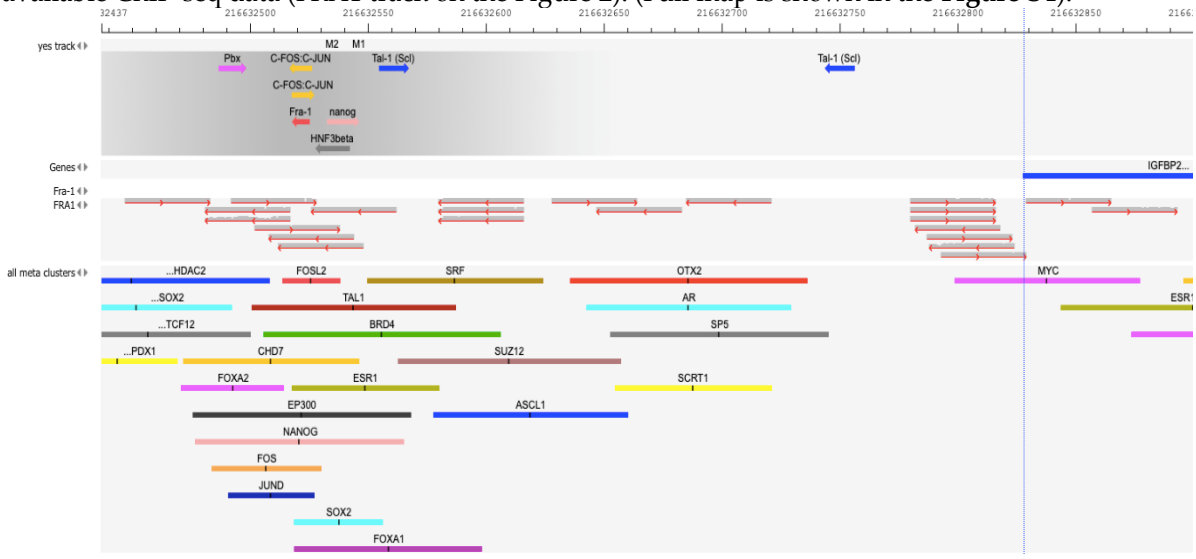

Figure 2. Map of the cluster of TF binding sites of the composite model identified within the promoter of IGFBP2 gene (-1000 to +100bp relative to transcription start site (TSS)). Position of the TSS (the beginning of the first exon on IGFBP2 gene) is shown by the vertical dotted line. Track "yes track" represents the cluster of identified TF binding sites of the composite model within the promoter. The direction of the arrows gives the orientation of the PWMs. The names of TFs binding to these sites are shown above the arrows. The track "FRA1" represents the mapped reads of the FRA1(also called FOSL1) ChIP-seq data of GEO, GSM803382. The reads were maped on the hg38 human genome using Subread-aligner[42] with default parameters.
The track "all meta clusters" shows all known meta-clusters in this region from GTRD database41 that represent the overlapping fragments of peaks for one particular TF from several ChIP-seq experiments. The name of TF is shown above each meta-cluster. Several predicted TF binding sites in the composite model are confirmed in independent ChIP-seq experiments: several overlapping reads of FRA1 ChIP-seq data in the "FRA1" track and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

FOSL2 meta- cluster in the GTRD confirm predicted site for Fra-1; FOS and JUN meta-clusters in the GTRD confirm the predicted c-Fos:c-Jun binding sites; NANOG meta-cluster confirms the predicted Nanog binding site; TAL1 meta-cluster confirms the predicted Tal-1 binding site; FOXA2 and FOXA1 meta-clusters of GTRD confirms the HNF3beta binding site.

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 five 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 genes that encode the master regulator proteins are controlled by the transcription factors revealed by CMA in their promoters, which maintains the multiple positive feedback loops in the system. It should be underlined here, that in such networks with positive feedback loops, the identified key transcription factors, such as Fra-1, are both upstream of their target genes, such as IGFBP2, as well as downstream from the master regulator proteins, such as IGFBP-2. (encoded by these genes). The regulatory network reconstructed along with six master regulators is shown in **Figure 3**, the master regulators and their LogFC in STS are listed in **Table 3**. Since VEGF165 is a splice variant of VEGF-A, only the latter will be considered further on.

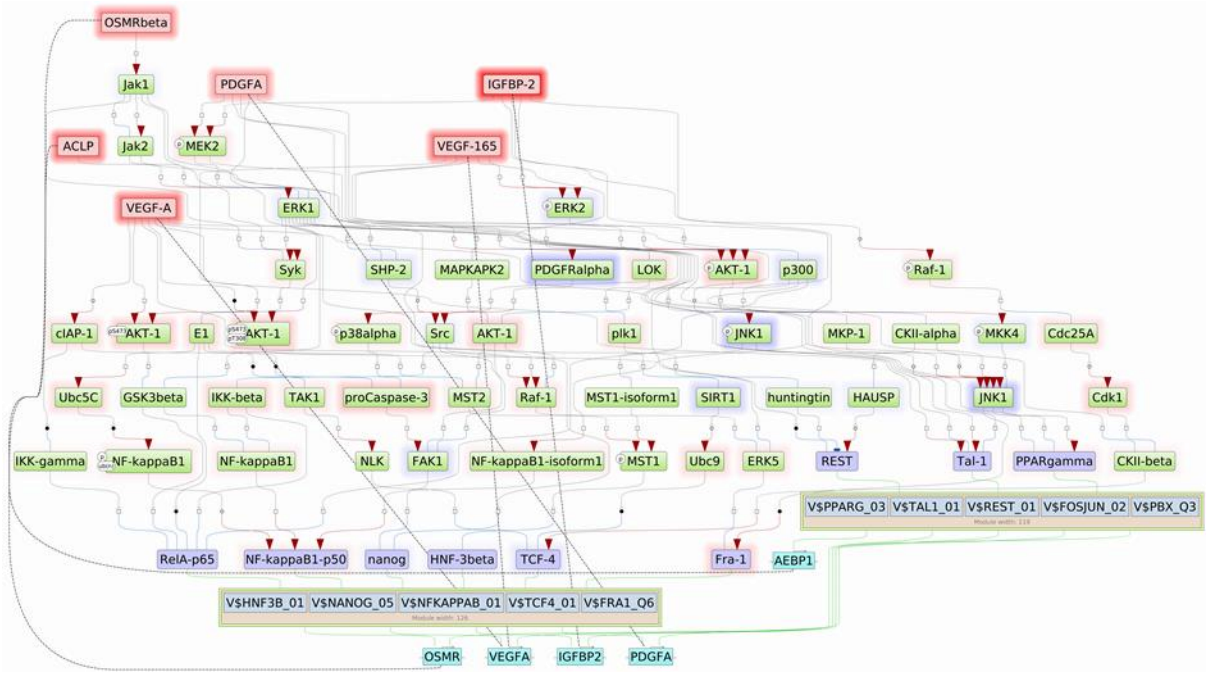

Figure 3. Signal transduction and Gene regulatory network of six master regulators (red nodes) regulating two transcription factor modules (purple nodes) enriched in promoters of highly upregulated genes of STS. The dotted lines from genes to such signalling proteins represents the transcription and translation processes (positive feedback loops). The outside box filling is based on LogFC and are filled red when upregulated (log2FC>0.2 & *p*-value<0.05) and filled blue when downregulated (log2FC<0.2 & *p*-value<0.05) in the current study[68].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

Table 3. Table of the master regulators identified, their description, log2FC in STS and number of TFs regulated.

Molecule Name	Gene Description	HGNC Gene symbol	log2FC in STS	No. of TFs regulated
IGFBP-2	Insulin like growth factor binding protein 2	IGFBP2	1.149	9
ACLP	AE binding protein 1	AEBP1	0.782	9
VEGF-A	vascular endothelial growth factor A	VEGFA	0.778	9
VEGF-165	vascular endothelial growth factor A	VEGFA	0.778	9
OSMRbeta	Oncostatin M receptor	OSMR	0.634	8
PDGF-A	platelet derived growth factor subunit A	PDGFA	0.529	9

2.4. Validating the expression of master regulators in other cohorts

Expression patterns of the above identified master regulators have been validated in two different cohorts. A) TCGA-GBM microarray[18] data and B) GSE16011[43]. The expression patterns were similar and there is a significant upregulation of all master regulators except for VEGFA (GSE16011: adj.*p*-value=0.069, TCGA-GBM: adj.*p*-value=0.075) **Table S1-E & S1-F**. The differential expression values are given in **Table 4**.

Table 4. Expression of master regulators across survival groups (STS and LTS resp.,) and across 3 datasets – GSE, TCGA-GBM microarray and GSE16011.

Master Regulator	GSE		GSE16011		TCGA	
	log2FC (STS vs LTS)	adj.*p*-value	log2FC (STS vs LTS)	adj.*p*-value	log2FC (STS vs LTS)	adj.*p*-value
IGFBP2	1.149	4.87E-05	2.030	4.598E-04	1.098	5.00E-06
AEBP1	0.782	7.75E-05	1.723	0.001	0.971	3.96E-06
PDGFA	0.529	4.55E-04	1.680	4.709E-09	0.825	2.07E-05
VEGFA	0.778	5.20E-04	0.884	0.069	0.500	0.0752
OSMR	0.634	8.65E-04	1.957	4.24E-05	0.486	0.0318

2.5 Validating the master regulators in TCGA-GBM cohort

The TCGA-GBM microarray data containing 258 STS and 49LTS is used to validate the above identified drivers of short survival. The data is pre-processed, adjusted for batch effects (**Figure S3)** and differential gene expression analysis is performed. Same cut-offs for log2FC and adj.*p*-value are used. We identified 171 genes upregulated in STS of GBM (log2FC>0.5 & adj.*p*-val<0.05) full list in **Table S1-E**. 49 of them were in common between GSE dataset and TCGA-GBM, full differential gene expression analysis results are given in **Table S1-G**. Composite models selected by CMA algorithm across the two datasets were expected to vary. We identified a model that includes set of 16 transcription factors **Table S3** and 12 master regulators upstream (**Table S4)** of them regulating the signal transduction and gene regulatory network in STS.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

As a result, the TCGA-GBM dataset validates IGFBP2, AEBP1(ACLP) and PDGFA as master regulators driving dysregulated gene network in STS. We also found that binding sites for FRA-1 transcription factor are statistically significantly enriched at the regulatory regions of dysregulated genes including IGFBP2 in TCGA-GBM cohort (**Table S5)**.

2.6. Impact of master regulators on survival in GBM

Univariate survival analysis was used to study the impact of these master regulators and their regulated transcription factors on overall survival in GBM based on TCGA-RNA-seq data. Patients are split into non-overlapping 50% upper and lower quantiles. Additionally, Cox regression for univariate survival analysis is performed, hazard ratio (HR) and corresponding *p*-values are shown in **Figure 4**. Univariate Cox regression analysis on other microarray datasets is given in **Table S1-I**. All master regulators were found to have significant impact upon survival except VEGFA. FRA1(FOSL1) was found to have significant cox hazard ratio (HR).

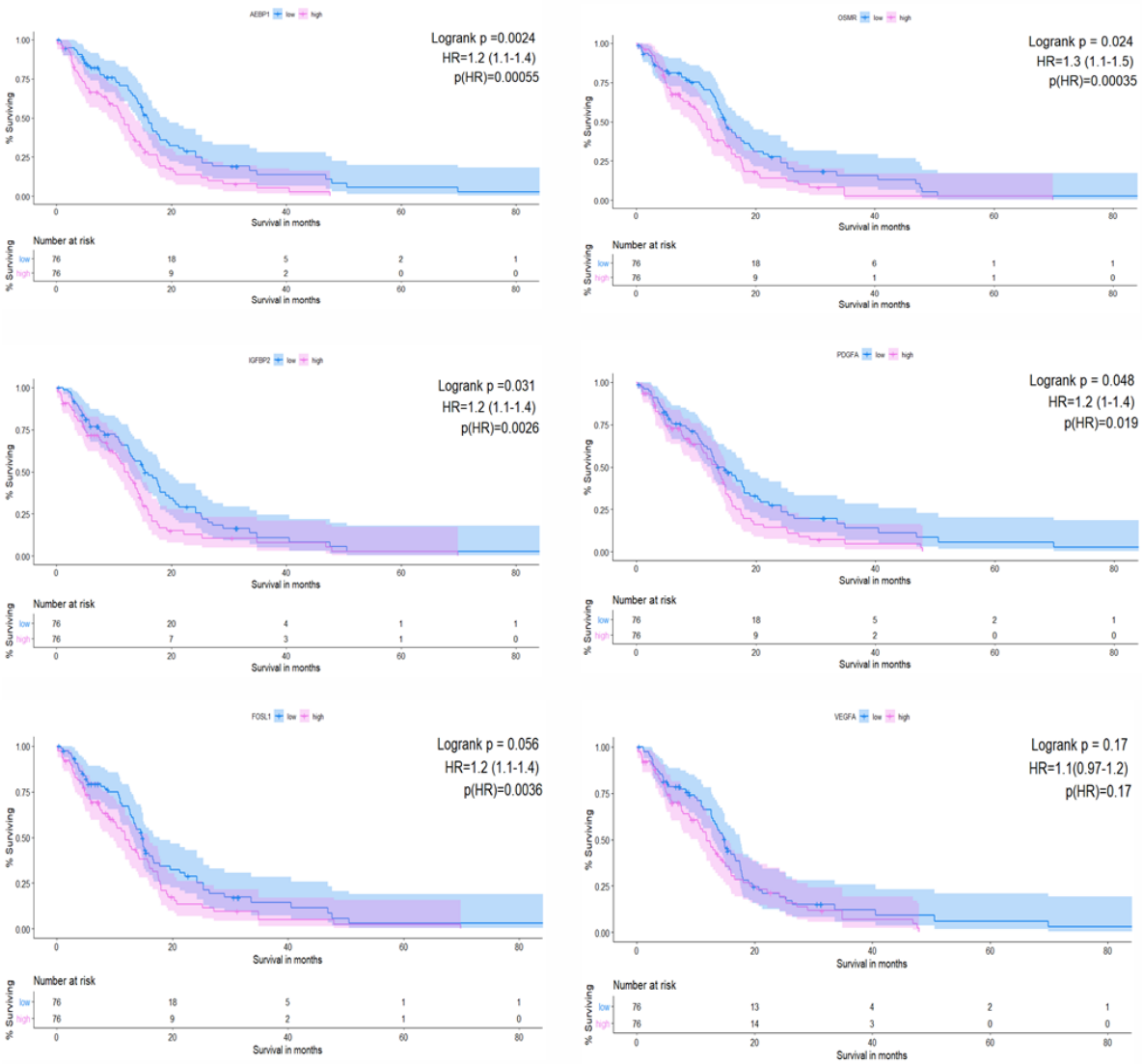

Figure 4. Survival analysis using 152 RNA-seq data of TCGA GBM cohort. Out of five master regulators, all but VEGFA (AEBP1, OSMR, PDGFA and IGFBP2) had a statistically significant impact on survival based on 50% upper and lower quantiles. Hazard ratio (HR) and statistical significance(p(HR)) according to Cox survival estimates are mentioned.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

2.7 Master regulator expression patterns across GBM Subtypes

Based on regulatory landscape of GBM, there are 4 subtypes – Classical, Mesenchymal, Proneural and Neural[44]. There is a significant level of intertumoral as well as intra-tumoral heterogeneity within each of them[44,45]. Molecular subtypes of GBM in GSE dataset is given in Table S6. Differentially expressed genes between STS and LTS within each subtype is given in Supplementary file 3. Expression patterns of master regulators across subtypes and across survival groups is depicted as boxplot in Figure S5. None of the master regulators were found to be significantly differentially expressed between survivor groups in any subtypes.

3. Discussion

Gene regulatory networks represent the causal regulatory relationships between transcription factors (TFs) and their gene targets, which enables us to discover dysregulated genes in certain biological states[46]. Comparative studies of short-term survivors and long-term survivors of GBM showed that gene-expression programs executed across survival groups vary significantly. In the light of these findings, we sought to apply an upstream analysis approach to gain insight about gene regulatory networks driving the short survival.

In the promoter analysis, we identified a set of 12 TFs in composite clusters that are enriched in the promoters regions of dysregulated genes in short term survivors (upregulated in STS). The TFs NANOG and REST have been said to be critical for self-renewal and maintenance of oncogenic signatures in glioblastoma stem-like cells[45,47], PPARG has emerged as a promising therapeutic target as its agonists increased median survival in GBM patients[48], NF-κB is implicated in several processes like invasion, epithelial-mesenchymal transition[49], resistance to radiotherapy[50] and maintenance of cancer stem-like cells[51], FRA-1/FOSL1 has been reported to be important in maintenance/progression of malignant glioma[52]. FRA-1 along with JUN-B modulates malignant feature of GBM by regulating expression of metalloproteinases like MMP-2 and MMP-9[53].Among these 12 TFs, we found that FRA-1 has a significant impact upon survival and has higher expression in STS. Debinski et al., (2005)[52] 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 was shown to be regulated by FRA-1 (supporting the feedback loop found in our work) and is a known prognostic factor in other aggressive cancers [54,55]

Graph analysis of the signal transduction network upstream of these transcription factors identified five potential master regulators that might explain gene dysregulation in STS, 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).All the identified master regulators were upregulated in STS and their expression patterns were validated in two other independent cohorts. We found that all master regulators except VEGFA has an impact on overall survival in the GBM patients. IGFBP2, AEBP1 and PDGFA master regulators driving short survival were validated in TCGA-GBM microarray cohort. Out of them, IGFBP2 had higher expression in STS. The insulin-like growth factor binding protein 2 (IGFBP2) is said to be one of most potential glioma oncogene and functions as a hub of oncogenic signalling pathways by regulating pro-tumorigenic signals of tumor initiation and progression. Earlier studies have suggested IGFBP2 to drive EMT and as a potential therapeutic target in mesenchymal GBM[49,56]. It is established that exogenous IGFBP2 promotes proliferation, invasion, and chemoresistance to temozolomide in glioma cells via integrin β1 by promoting ERK phosphorylation and nuclear translocation [57,58].IGFBP2 is considered as one of the strongest biomarkers of aggressive behavior in GBM[37,59] and also a prognostic marker for survival[37,60].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

Here, we propose that IGFBP2 can be a potential regulator of FRA-1 transcription factor. IGFBP2 induced RAF/MAPK signalling (Figure 3) activates FRA-1. It is earlier shown that IGFBP2 and FRA-1 regulated transcription of VEGF[54,55,61], which is the second most dysregulated master regulator in our network. It is reported that enhanced ERK signalling, that may be triggered by these master-regulators, may lead to mitogen-induced FRA-1 transcription [62] as well as its protection from proteasomal degradation [63]. As per the gene regulatory network developed, it is possible that FRA-1 mediates the positive feedback loop where FRA-1 activates transcription of master regulator genes in cooperation with other TFs and in turn master regulator gene products enhance/potentiate FRA-1. We find that promoters of the genes of all the five master regulators reported in the study have binding sites for FRA-1 transcription factor. Experimental evidences are available showing that IGFBP2 can enhance GBM invasion by enhancing MMP2 expression[64]. This validates our computationally predicted hypothesis that IGFBP2 be a therapeutic target. The novelty of the work is that we are able to establish through the gene regulatory networks that FRA-1 can be a downstream transcription factor of IGFBP2 in GBM. Kesari, S. et al.,2011 confirms our hypothesis that IGFBP2 can enhance GBM invasion via AP1(FOS-JUN) transcription factor. Metalloproteinases like MMP-2/MMP-9 are earlier reported [65–68] to be regulated by FRA-1 transcription factor in several cancers including GBM. Taken together, our work proposes that the regulation of IGFBP2 gene expression via AP1(FOS-JUN) transcription factor can be an important mechanism of GBM invasion. In this study, Figure 5 combines prior knowledge in the field and gene regulatory network developed in our study (Figure 3 and 5).

In summary, the current work proposes a gene regulatory network in STS regulated by five master regulators – IGFBP2, VEGFA, PDGFA, OSMR and AEBP1 which can potentially act as biomarkers of GBM prognosis and as therapeutic targets for enhancing GBM prognosis. The work also proposes a novel mechanism of gene dysregulation by IGFBP2 by modulating a key molecule of tumor invasiveness and progression - FRA-1 transcription factor. All the genes encoding these five master regulators have binding sites for FRA-1 in their promoters. FRA-1 and the master regulators are in a positive feedback loop to orchestrate a complex tumorigenic program leading to poor survival in GBM.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

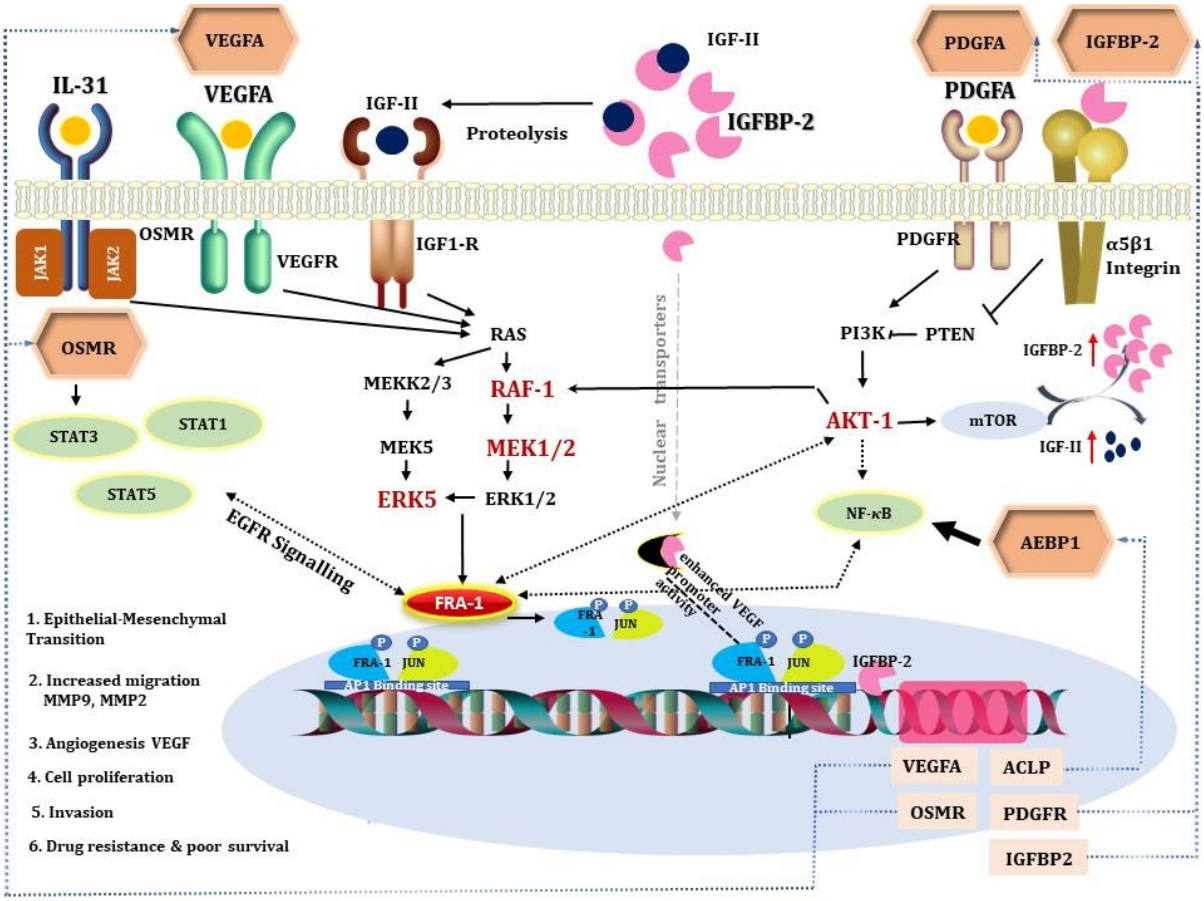

Figure 5: Diagram combining prior knowledge about of role of IGFBP2 in GBM and the gene regulatory network developed in the study. *The hexagons are master regulators identified in our analysis. All the intermediates of the gene regulatory network are colored red if upregulated in STS, and black if not present in the network. Dotted black line if the knowledge is through literature, continuous black line if known through gene regulatory network. Blue dotted lines represent gene regulatory connections between master regulators and their corresponding genes transcribed by target transcription factors.*

VEGFA, PDGFA, IGFs and IL-31 activate RAF/MEK/ERK signalling which mediates cell survival through PI3K-AKT pathway[69,70]*. MEK2/RAF1/ERK5 and AKT-1 are found to be upregulated in STS, suggestive of activated ERK signalling which can contribute to drug resistance* [71,72]*. IGFBP2 activates IGFR either by increasing bioavailability of IGFs or by direct interaction with its functional domain. Integrin acts as receptor for IGFBP2 extracellular signals* [57,58] *and modulates NF-κB signalling. IGFBP2 by nuclear translocation* [61] *is involved in transcriptional regulation of the VEGF gene and modulates angiogenesis* [61]*. STAT3 and NF-κB are said to be the two major downstream transcription factors of IGFBP2 that direct tumorigenic intracellular signalling* [73] *via EGFR signalling. Oncostatin M, receptor for cytokine IL31 is a regulator of EGFR signalling*[34]*.FRA-1 is required for AKT activation in cancers to promote AKT-dependent cell growth*[74]*, NF-κB can regulate AP1(FOS & JUN) thereby VEGF expression in pancreatic tumor cell lines*[75]*. All the 5 master regulators have binding sites for FRA-1. In the figure we depicted the possible positive feedback loop between FRA-1 and the master regulators 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 and clinical information of patients with GBM were collected from public repository of GEO database – GSE108474[76](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108474) and GSE53733[13]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

(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. Duplicates were not removed. Sample information and cleaned datasets are given in **Supplementary file 4**

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[12]	HU133 plus 2.0 arrays	16	23
GSE108474[56]	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 [77]. Batch correction of the pooled expression data was performed using empirical Bayes framework is performed [78] . 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.

4.3 Differential gene expression (DEG) analysis

LIMMA (Linear Models for Microarray Data) method was applied to identify differentially expressed genes [40]. 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. 957 genes were significantly (adj.*p*-value<0.05) differentially expressed (DEGs). 115 of them were significantly upregulated (adj.*p*-value <0.05 & log2FC>0.5) and 83 were significantly downregulated (adj.*p*-value<0.05 & log2FC<(-0.5).

4.4. 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)[79]. The master regulator search uses the TRANSPATH® database , release 2019.3 (geneXplain GmbH, Wolfenbüttel, Germany) (https://genexplain.com/transpath)[20]. A comprehensive signal transduction network of human cells is built by Genome Enhancer 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[80], release 2020.2 (https://genexplain.com/humanpsd). The Ensembl[81] database build 99.38 (hg38) (http://www.ensembl.org) was used for gene IDs representation and Gene Ontology[19] (GO) (http://geneontology.org) was used for functional classification of the studied gene set.

4.5. Functional Annotation

To explore the biological importance of gene signatures, gene set enrichment analysis is performed. All the adj.*p*-value significant genes were used. GSEA is an efficient method to determine

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

whether the genes of interest show statistically significant enrichment between different biological states. Gene ontology enrichments for cellular component, biological process and molecular functions were performed. To investigate the top enriched ontology terms 1000 random permutations were done and adj.*p*-value cutoff of 0.05 is used. The dysregulated gene networks enrichment also gives useful insight about known disease signatures [82]. The hallmark gene set of MsigDB[83] defines specific biological states or processes. Enrichment analysis is performed in R using DOSE[84] package. PANTHER pathway enrichment of the identified transcription factors was performed using EnrichR tool [85]. TRANSPATH[20] pathway enrichment was performed using geneXplain platform.

4.6. 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). Genome Enhancer is a multi-omics analysis service (https://genexplain.com/genome-enhancer/) which is build using an open source BioUML programming environment[24] (www.biouml.org) and incorporates an automated pipeline for "upstream analysis"[25,26] and newest advanced approach called "walking pathways"[27].Significantly upregulated genes in STS were used in this workflow.

The workflow 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 [31]. Enhancers typically have a length of several hundreds of nucleotides and are bound by multiple transcription factors [20].

Identifying such clusters of binding sites for such 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 that are binding to clusters of co-localized TF binding sites (composite regulatory modules)[86]

We use Composite Module Analyst (CMA) [31] method to detect such potential enhancers, as targets of multiple TFs bound in a 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. The transcription factors are ranked based on (a)The Yes-No ratio: given a set of promoter sequences of dysregulated genes, denoted as Yes set, and promoter sequences of unchanged genes in the experimental condition, denoted as No set, motifs are considered important if they have a high Yes/No ratio, the ratio of motif occurrences per promoter in Yes and No sets, and a statistically significant enrichment of occurrences in Yes sequences assessed by the binomial *p*-value". (b) Regulatory score, which is a measure of involvement of a transcription factor in controlling expression of genes that encode master regulators. 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. Master

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 17 February 2021 doi:10.20944/preprints202010.0046.v3

regulators regulating the above revealed transcription factors are not only ranked based on (a) LogFC but also (b)CMA score - which signifies how strong is the potential for this gene to be regulated by transcription factors of interest and (c) Master regulator score – which signifies how strong is the potential of this gene product to regulate activity of those transcription factors. Selected master regulators can also be visualized and with possibility to map the LogFC, p value on the created regulatory network.

4.7. Validation of observed gene signatures

The raw microarray data of 560 TCGA-GBM samples were downloaded from TCGA legacy. The GSE16011 raw. CEL data was downloaded from GEO repository. Both raw datasets were processed and analyzed independently following same steps as mentioned earlier. These two datasets are used to observe and validate the expression pattern of master regulators across the two survival groups (see Table 6). GSE16011 comprises of data generated at a single center and is used in several studies [62-64], unlike TCGA. TCGA-GBM microarray data PCA plots are given **Figure S3** and no significant batch effects in the context of survival groups were found.

Table 6. Statistics of two validation datasets.

Datasets	Platform	Short-term survivors	Long-term survivors
GSE16011[43]	HU133 plus 2.0 arrays	93	16
TCGA-GBM microarray[18]	HU133	271	49

4.8. Validation of Master regulators

The TCGA-GBM microarray data downloaded from TCGA legacy archive is processed in same fashion as GSE. Similar cutoffs (log2FC, *p*-value) and parameters are used to identify enriched transcription factors and network analysis in-order to understand drivers of gene regulatory networks in short survival.

4.9. Impact on survival

Master regulators and their target transcription factors affect the whole regulatory network and therefore can have an independent impact on survival in GBM patients. Level 3 RNA-seq data and clinical data for 152 TCGA GBM cohort is downloaded using TCGAbiolinks package in R. Survival and Survminer libraries in R are used to perform univariate survival analysis. Univariate survival analysis is used to understand the impact of individual master regulator on survival in GBM with non-overlapping 50% upper and lower quantiles. Additionally, Univariate Cox regression for survival analysis is performed using coxph function of Survival package to calculate Hazard ratio (HR) with pvalue cutoff of 0.05 for significance.

5. Conclusion

In the work presented, we have identified candidate master regulators responsible for gene dysregulation in short-term survivors. These candidates have sufficient experimental evidence towards their role in GBM. Out of reported five master regulators, IGFBP2 is established as the most promising master regulator. Through the gene regulatory network analysis, we propose that IGFBP-2 and FRA-1 are in a positive feedback loop that may lead to a pathological autocatalytic process responsible for poor survival in 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.

Availability of software, data and materials: The datasets analyzed during the current study are available from the previous publications [60,61]. The results of the analysis performed using pipeline Genome Enhancer in geneXplain platform are available here.

https://github.com/genexplain/Manasa_KP_et_al_IGFBP2_regulatory_networks_in_Gliobastoma

The source code of the open source BioUML platform used for creation of the Genome Enhancer pipeline is available for download at www.biouml.org. Data are also available in the Additional files provided in the Supplement to the publication.

Conflicts of Interest: The authors Manasa KP, Darius Wlochowitz and Tim Beißbarth are from Department of Medical Bioinformatics, University Medical Center Göttingen, Manasa KP, Alexander Kel and Edgar Wingender are employees of geneXplain GmbH, which maintains and distributes the geneXplain//BioUML platform and Genome Enhancer used in this study.

References

1. Kolmykov, S. *et al.* Gtrd: An integrated view of transcription regulation. *Nucleic Acids Research* vol. 49 D104–D111 (2021).
2. Wen, P. Y. & Kesari, S. Malignant gliomas in adults. *New England Journal of Medicine* vol. 359 492–507 (2008).
3. Krex, D. *et al.* Long-term survival with glioblastoma multiforme. *Brain* **130**, 2596–2606 (2007).
4. Das, P. *et al.* A clinicopathological and molecular analysis of glioblastoma multiforme with long-term survival. *J. Clin. Neurosci.* **18**, 66–70 (2011).
5. Shinawi, T. *et al.* DNA methylation profiles of long- and short-term glioblastoma survivors. *Epigenetics* **8**, 149–156 (2013).
6. Sonoda, Y. *et al.* Long-term survivors of glioblastoma: Clinical features and molecular analysis. *Acta Neurochir. (Wien).* **151**, 1349–1358 (2009).
7. 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).
8. 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).
9. 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).
10. 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).
11. 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).
12. 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).

13. 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).

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

15. 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).

16. Barrett, T. *et al.* NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Res.* **41**, D991–D995 (2013).

17. Athar, A. *et al.* ArrayExpress update - From bulk to single-cell expression data. *Nucleic Acids Res.* **47**, D711–D715 (2019).

18. Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **375**, 1109–1112 (2016).

19. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nature Genetics* vol. 25 25–29 (2000).

20. 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).

21. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* (2020) doi:10.1093/nar/gkaa970.

22. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. (2003).

23. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).

24. Kolpakov, F. *et al.* BioUML: an integrated environment for systems biology and collaborative analysis of biomedical data. *Nucleic Acids Res.* **47**, W225–W233 (2019).

25. Boyarskikh, U. *et al.* Computational master-regulator search reveals mTOR and PI3K pathways responsible for low sensitivity of NCI-H292 and A427 lung cancer cell lines to cytotoxic action of p53 activator Nutlin-3. *BMC Med. Genomics* **11**, (2018).

26. Koschmann, J., Bhar, A., Stegmaier, P., Kel, A. & Wingender, E. "Upstream Analysis": An Integrated Promoter-Pathway Analysis Approach to Causal Interpretation of Microarray Data. *Microarrays* **4**, 270–286 (2015).

27. Kel, A. *et al.* Walking pathways with positive feedback loops reveal DNA methylation biomarkers of colorectal cancer. *BMC Bioinformatics* **20**, (2019).

28. Matys, V. *et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108-10 (2006).

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

30. Kel, A., Voss, N., Jauregui, R., Kel-Margoulis, O. & Wingender, E. Beyond microarrays: Finding key transcription factors controlling signal transduction pathways. *BMC Bioinformatics* **7**, S13 (2006).

31. Waleev, T. *et al.* Composite Module Analyst: Identification of transcription factor binding site

combinations using genetic algorithm. *Nucleic Acids Res.* **34**, W541 (2006).

32. 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).

33. 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.

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

35. 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).

36. Liu, Y., Song, C., Shen, F., Zhang, J. & Song, S. W. IGFBP2 promotes immunosuppression associated with its mesenchymal induction and FcγRIIB phosphorylation in glioblastoma. *PLoS One* **14**, (2019).

37. 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).

38. Cai, J. *et al.* Immune heterogeneity and clinicopathologic characterization of IGFBP2 in 2447 glioma samples. *Oncoimmunology* **7**, e1426516 (2018).

39. Lindström, M. S. Expanding the scope of candidate prognostic marker IGFBP2 in glioblastoma. *Bioscience Reports* vol. 39 (2019).

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

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

42. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).

43. Gravendeel, L. A. M. *et al.* Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res.* **69**, 9065–9072 (2009).

44. Verhaak, R. G. W. *et al.* Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).

45. Bradshaw, A. *et al.* Cancer Stem Cell Hierarchy in Glioblastoma Multiforme. *Frontiers in Surgery* vol. 3 1 (2016).

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

47. Kamal, M. M. *et al.* REST regulates oncogenic properties of glioblastoma stem cells. *Stem Cells* **30**, 405–414 (2012).

48. Ellis, H. P. & Kurian, K. M. Biological Rationale for the Use of PPARÎ³ Agonists in Glioblastoma. *Front. Oncol.* **4**, 52 (2014).

49. Yamini, B. NF-κB, Mesenchymal Differentiation and Glioblastoma. *Cells* **7**, 125 (2018).

50. Avci, N. G. *et al.* NF-κB inhibitor with Temozolomide results in significant apoptosis in glioblastoma via the NF-κB(p65) and actin cytoskeleton regulatory pathways. *Sci. Rep.* **10**,

13352 (2020).

51. da Hora, C. C. *et al.* Sustained NF-κB-STAT3 signaling promotes resistance to Smac mimetics in Glioma stem-like cells but creates a vulnerability to EZH2 inhibition. *Cell Death Discov.* **5**, 72 (2019).

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

53. 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).

54. 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).

55. 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).

56. Liu, Y., Song, C., Shen, F., Zhang, J. & Song, S. W. IGFBP2 promotes immunosuppression associated with its mesenchymal induction and FcγRIIB phosphorylation in glioblastoma. *PLoS One* **14**, (2019).

57. 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).

58. 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).

59. 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).

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

61. 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).

62. 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).

63. 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).

64. Wang, H. *et al. Insulin-like Growth Factor Binding Protein 2 Enhances Glioblastoma Invasion by Activating Invasion-enhancing Genes 1*. *CANCER RESEARCH* vol. 63 www.mdanderson.org/genomics. (2003).

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

66. Kimura, R., Ishikawa, C., Rokkaku, T., Janknecht, R. & Mori, N. Phosphorylated c-Jun and Fra-1 induce matrix metalloproteinase-1 and thereby regulate invasion activity of 143B osteosarcoma cells. *Biochim. Biophys. Acta - Mol. Cell Res.* **1813**, 1543–1553 (2011).

67. Adiseshaiah, P., Vaz, M., Machireddy, N., Kalvakolanu, D. V. & Reddy, S. P. A Fra-1-dependent, matrix metalloproteinase driven EGFR activation promotes human lung epithelial cell motility and invasion. *J. Cell. Physiol.* **216**, 405–412 (2008).

68. Prywes, R. & Henckels, E. Fra-1 regulation of Matrix Metallopeptidase-1 (MMP-1) in metastatic variants of MDA-MB-231 breast cancer cells. *F1000Research* **2**, (2013).

69. Simpson, A., Petnga, W., Macaulay, V. M., Weyer-Czernilofsky, U. & Bogenrieder, T. Insulin-Like Growth Factor (IGF) Pathway Targeting in Cancer: Role of the IGF Axis and Opportunities for Future Combination Studies. *Target. Oncol.* **12**, 571–597 (2017).

70. 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).

71. 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).

72. 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).

73. 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).

74. Zhang, X., Wu, J., Luo, S., Lechler, T. & Zhang, J. Y. FRA1 promotes squamous cell carcinoma growth and metastasis through distinct AKT and c-Jun dependent mechanisms. *Oncotarget* **7**, 34371–34383 (2016).

75. Fujioka, S. *et al.* NF-κB and AP-1 Connection: Mechanism of NF-κB-Dependent Regulation of AP-1 Activity. *Mol. Cell. Biol.* **24**, 7806–7819 (2004).

76. Gusev, Y. *et al.* Data descriptor: The REMBRANDT study, a large collection of genomic data from brain cancer patients. *Sci. Data* **5**, (2018).

77. 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).

78. 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).

79. 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).

80. Integrating pathway data for systems pathology - PubMed. https://pubmed.ncbi.nlm.nih.gov/17822386/.

81. Aken, B. L. *et al.* The Ensembl gene annotation system. *Database* **2016**, baw093 (2016).

82. 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).

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

84. 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).

85. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis

 doi:10.20944/preprints202010.0046.v3

tool. *BMC Bioinformatics* **14**, (2013).

86. Kel-Margoulis, O. V., Kel, A. E., Reuter, I., Deineko, I. V. & Wingender, E. TRANSCompel®: A database on composite regulatory elements in eukaryotic genes. *Nucleic Acids Res.* **30**, 332–334 (2002).