Supplementary Materials

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

Manasa KP, Darius Wlochowitz, Edgar Wingender, Tim Beißbarth,Alexander Kel

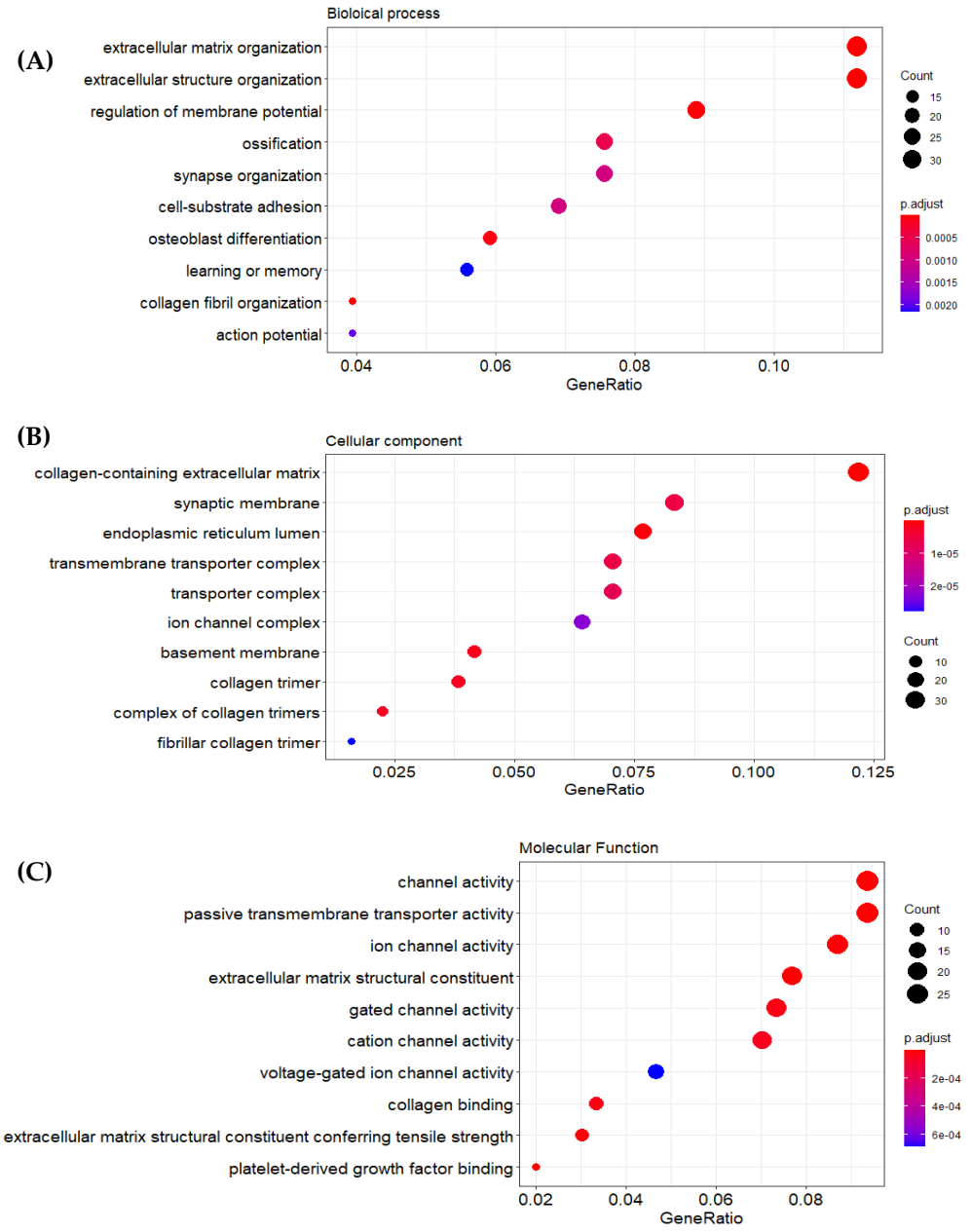
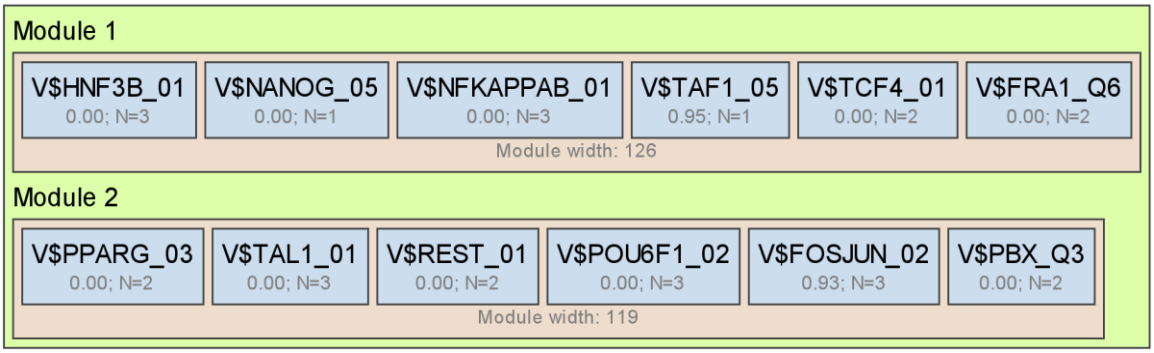
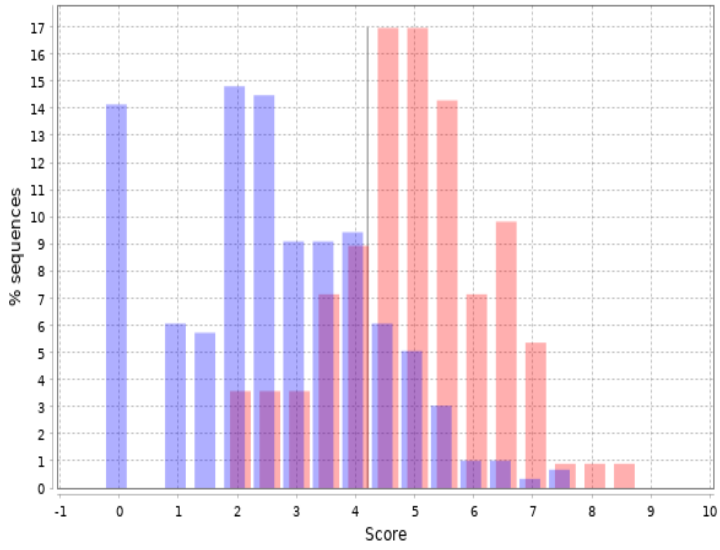


Figure S1. Functional enrichment analysis for differentially expressed genes abs (log-Fold change) of 0.5. (A) Dot plot for enriched GO Biological Process. Y-axis represents enriched ontology categories and X-axis represents GeneRatio, the percentage of all DEGs belonging to the corresponding category. The dots are sized based on gene ratio and are coloured according to their adjusted p-value. (B) Enrichment for GO Cellular Components similar to A (C) Enrichment for GO Molecular Function similar to A and B.

(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)

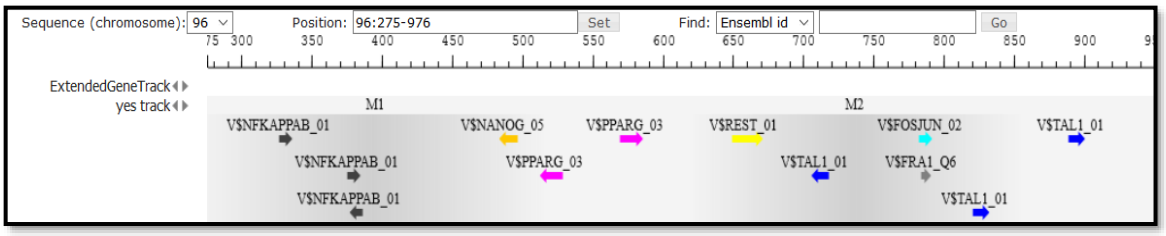
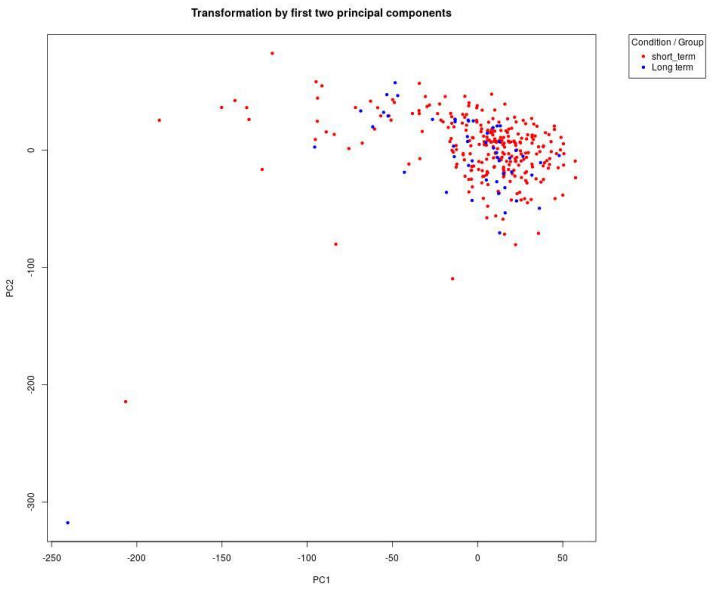


Figure S2. Results of CMA analysis of upregulated genes in short-term survivors. (A) Combination of 12 PWMs (Position weigh matrices) with their optimized cut-offs identified by genetic algorithm. (cut-off 0.00 means that the algorithm chose the default cut-off; parameter N represent maximal number of 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

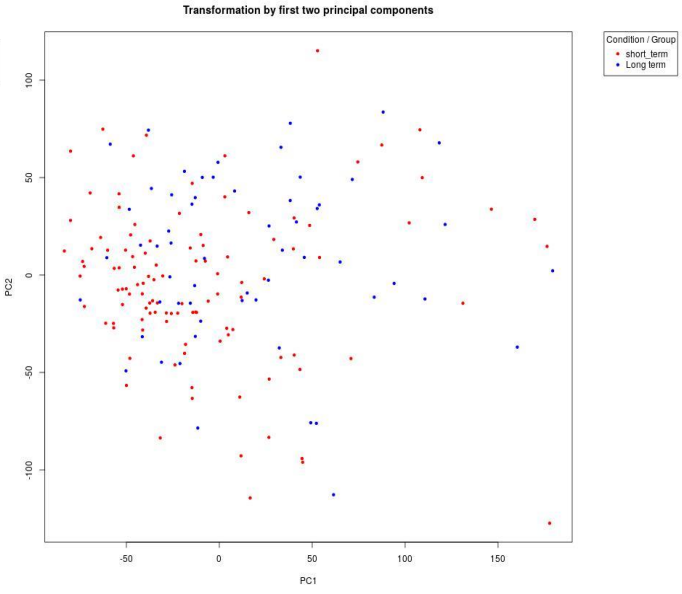
Table S2. Table of the transcription factors identified in CMA analysis of GSE dataset, LogFC (STS vs LTS), pvalue and adj_pvalue in STS across 3 datasets GSE, TCGA-GBM microarray and GSE16011

Transcription factors	GSE			TCGA Dataset			GSE16011		
	LogFC (ST vs LT)	pvalue	adj_pvalue	LogFC (ST vs LT)	pvalue	adj_pvalue	LogFC (ST vs LT)	pvalue	adj_pvalue
FOSL1	0.2353	0.00833	0.093832	0.022	0.866477	0.942691	0.461	0.032989	0.170584
PBX3	0.173	0.12895	0.372164	0.163	0.151272	0.411335	0.558	0.016908	0.113002
NFKB1	0.083	0.155435	0.408934	-0.004	0.956529	0.982133	0.209	0.115989	0.355865
PBX2	0.082	0.102006	0.33059	-0.22	0.001109	0.029677	-0.278	0.019596	0.12327
PPARG	0.0699	0.222946	0.491011	0.129	0.375068	0.643568	0.463	3.60E-04	0.008414
TAL1	0.027	0.384873	0.647536	-0.33	1.80E-05	0.003086	0.115	0.459738	0.692082
REST	0.021	0.578374	0.787431	0.176	0.028054	0.163735	0.288	0.034068	0.174094
RELA	0.007	0.90247	0.957246	0.072	0.281728	0.557129	-0.142	0.304378	0.576816
FOXA2	-0.05	0.336166	0.604469	-0.199	0.015938	0.120938	0.166	0.322979	0.592225
NANOG	-0.07	0.111174	0.344927	-0.189	0.038579	0.196264	-0.344	0.108064	0.344113
PBX1	-0.102	0.166715	0.42262	0.06	0.564011	0.77815	0.029	0.887276	0.94975
TCF7L2	-0.12	0.147798	0.399134	-0.084	0.406285	0.671782	-0.531	0.00498	0.051104

(a)



(b)



(c)

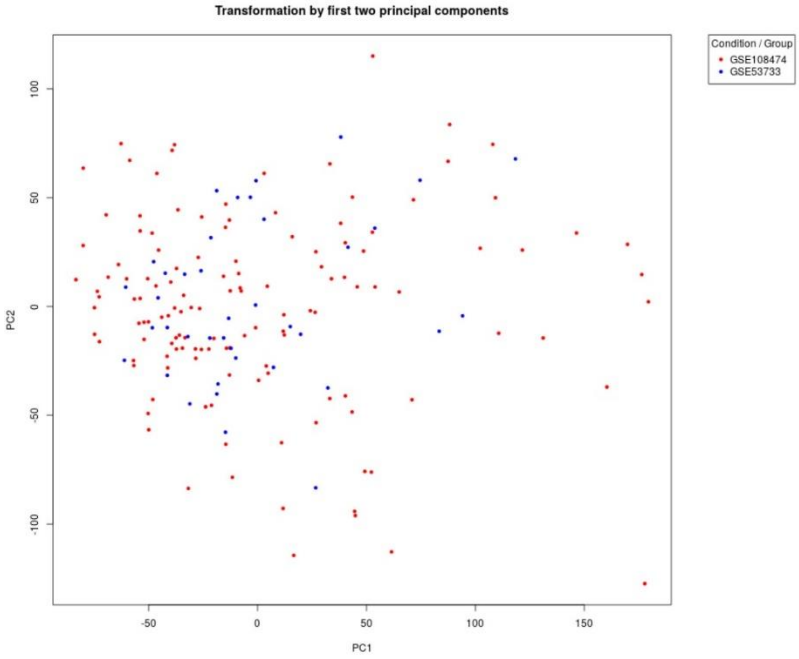


Figure S3. PCA plot for normalized (a) TCGA-GBM data (b) GSE (GSE108474, GSE53733) datasets show no significant batch effects across survival groups colored as red (STS) and blue (LTS) (c) PCA plot for normalized GSE data shows no batch effects between GSE108474 colored in red and GSE53733 colored in Blue.

Table S3. 16 Transcription factors predicted from enhancer models, potentially regulating the genes upregulated in STS TCGA (LogFC > 0.5 and adj_pvalue <0.05)

Gene description	Gene symbol	Yes-No ratio	Regulatory score
E2F transcription factor 1	E2F1	1.249664	1.714286
LYL1 basic helix-loop-helix family member	LYL1	1.165219	0
nescient helix-loop-helix 2	NHLH2	1.165219	0
glial cells missing transcription factor 1	GCM1	1.294249	1.275613
high mobility group AT-hook 1	HMGA1	1.409148	1.295815
heart and neural crest derivatives expressed 1	HAND1	1.295599	0
nuclear receptor subfamily 3 group C member 1	NR3C1	1.388077	1.44531
TAL bHLH transcription factor 1, erythroid differentiation factor	TAL1	1.648248	1.88456
transcription factor 3	TCF3	1.500366	1.703319
HNF1 homeobox A	HNF1A	2.255652	1.043357
HNF1 homeobox B	HNF1B	1.178953	0
myeloid zinc finger 1	MZF1	1.300681	0
Kruppel like factor 4	KLF4	1.289327	1.492063
lymphoid enhancer binding factor 1	LEF1	2.299881	1.460317
glial cells missing transcription factor 2	GCM2	1.294249	0
high mobility group AT-hook 2	HMGA2	1.409148	1.212121

Table S4. Table of the master regulators identified, their description, LogFC in STS and number of TFs regulated.

Molecule Name	logFC	CMA score	Reached from set
FGFR3	1.073	12.142	10
AEBP1	0.972	11.164	10
IGFBP2	1.098	10.191	10
CNR1	0.745	7.323	10
TRIM22	0.638	11.303	10
SPRY2	0.584	9.602	10
CASP1	0.524	8.102	10
DUSP6	0.705	8.742	10
CXCL8,CXCR1,CXCR2,GNAI2	1.050	10.031	10
CD14,IRAK1,IRAK2,LBP,LY96,MYD88,TIRAP,TLR4	0.741	8.277	10
IL1B,IL1R1,IL1RAP,IRAK1,IRAK2,IRAK4,MYD88,TOLLIP	0.950	9.157	10
PDGFA	0.825	7.948	10

Table S5. Table containing Yes/No ratio and corresponding P-value of enrichment for FRA1 transcription factor binding sites in GSE and TCGA dataset. There are 4 binding sites enriched for FRA-1 transcription factor and are all found to be significantly enriched in GSE and is validated in TCGA dataset.

ID	GSE dataset		TCGA dataset	
	Yes/No Ratio GSE	**P-value_GSE**	**Yes/No Ratio TCGA**	**P-value_TCGA**
V$FRA1_Q5	2.45536	0.00117	6.20968	5.8419E-5
V$FRA1_Q6	2.45536	0.00117	2.17735	3.9016E-6
V$FRA1_Q6_01	2.90434	4.1107E-4	2.47487	5.9147E-4
V$JUNBFRA1_01	2.60357	7.4892E-7	1.30521	2.1143E-6

Table S6. GBM subtypes across survival groups and across 4 datasets used in this study.

GBM Subtype	GSE108474		GSE53733		GSE16011		TCGA_GBM	
Verhaak_2010 subtyping	**STS**	**LTS**	**STS**	**LTS**	**STS**	**LTS**	**STS**	**LTS**
Classical	50	10	04	07	46	01	66	08
Mesenchymal	16	04	08	04	18	02	84	12
Proneural	13	14	04	14	13	13	67	07
Neural	18	07	00	00	16	00	41	21
Total	97	35	16	25	93	16	258	48

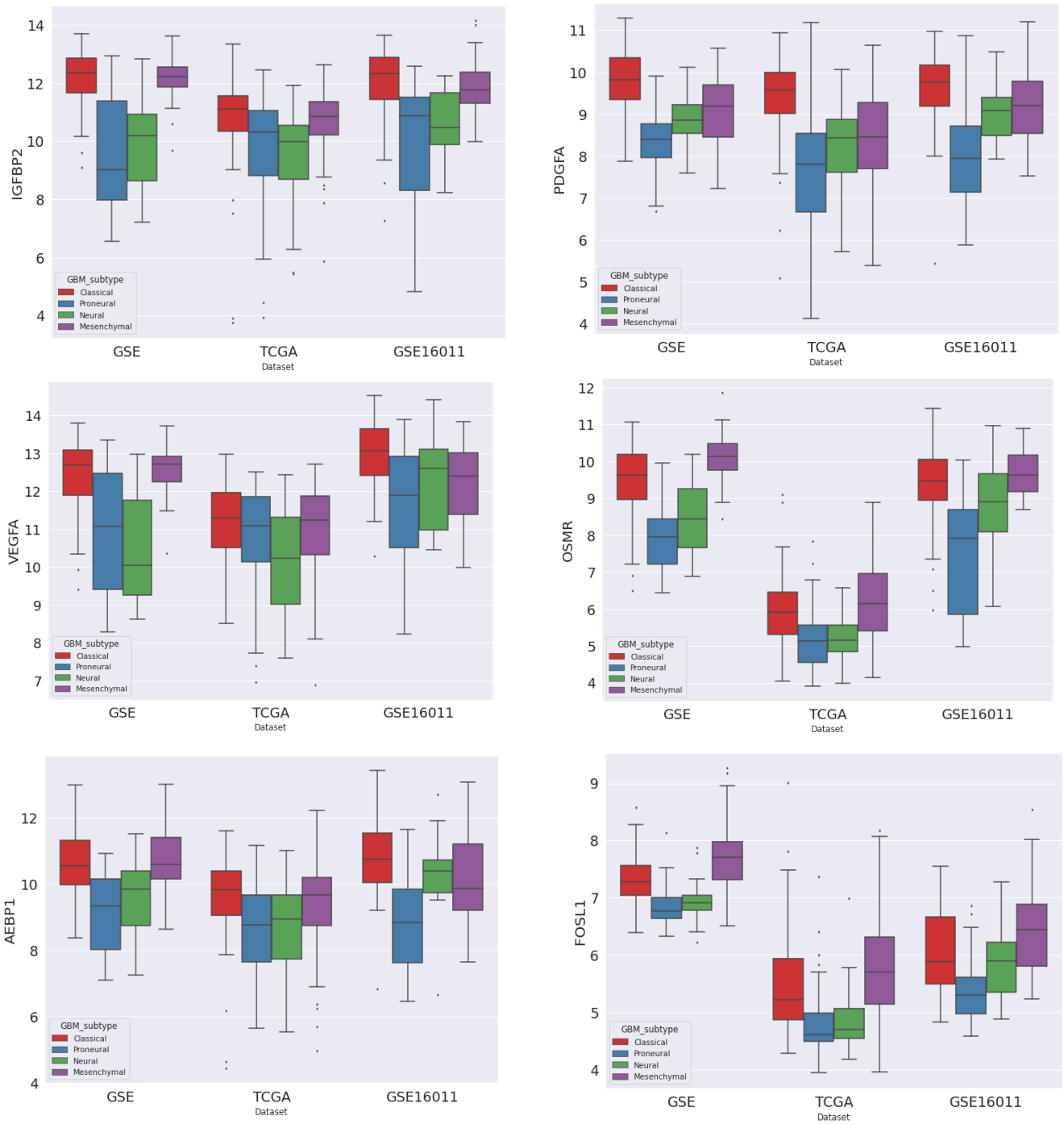


Figure S4. Expression levels of five master regulators and FRA-1 transcription factor across subtypes of GBM in GSE, TCGA and GSE16011 cohorts.