Supplementary Methods

The essential formulas and R codes in data processing

In our study, not only the microarray data, but also the RNA-seq data were normalized to 0-1. As we all know, there is batch effect between gene expression data from different databases. Before putting the data of different databases together for analysis, batch effect should be corrected first. In our study, RNA-seq data and microarray data were put together for analysis, so an appropriate normalization method was necessary. Given that the technological principles and data processing methods of RNA-seq and microarray are different, there is no widely recognized transformation relationship between RNA-seq data and microarray data at present.

We made some attempts. Since each dataset we included had a sufficient sample size, we take each dataset as a whole. We standardized every DEIR in its own cohort, and normalized the gene expression value to lie between 0 and 1. We made some attempts on the normalization methods and found that the division-based normalization method was suitable for RNA-seq data while the subtraction-based normalization method was suitable for microarray data. Compared with other normalization methods we tried, this method significantly improved the ability of prognostic stratification and the accuracy of prediction.

1. RNA-Seq data

(1) FPKM value was converted to TPM value using the R codes as follows:

```
FPKMtoTPM <- function(FPKM) {exp(log(FPKM) - log(sum(FPKM)) + log(1e6))}
```

(2) Gene expression data (TPM value) was normalized to [0,1] using the formula and R codes as follows:

Formula: X'=log2(X+1)/log2(Xmax+1) X: the TPM value of the DEIR.

R codes: normalization1 <- function(x){log2(x+1)/log2(max(x)+1)}

2. Microarray data

Gene expression data in microarray datasets was normalized to [0,1] using the formula and R codes as follows:

Formula: X'=(X-Xmin)/(Xmax-Xmin) X: the processed microarray value of the DEIR. The microarray data were also log-transformed (on a base 2 scale) when processed in GeneSpring GX software.

R codes: normalization2 <- function(x){(x-min(x))/(max(x)-min(x))}

Figure S1. Kaplan-Meier curve of OS based on immune infiltration type in TCGA dataset. Patients with type A infiltration had better OS than those with type B infiltration.

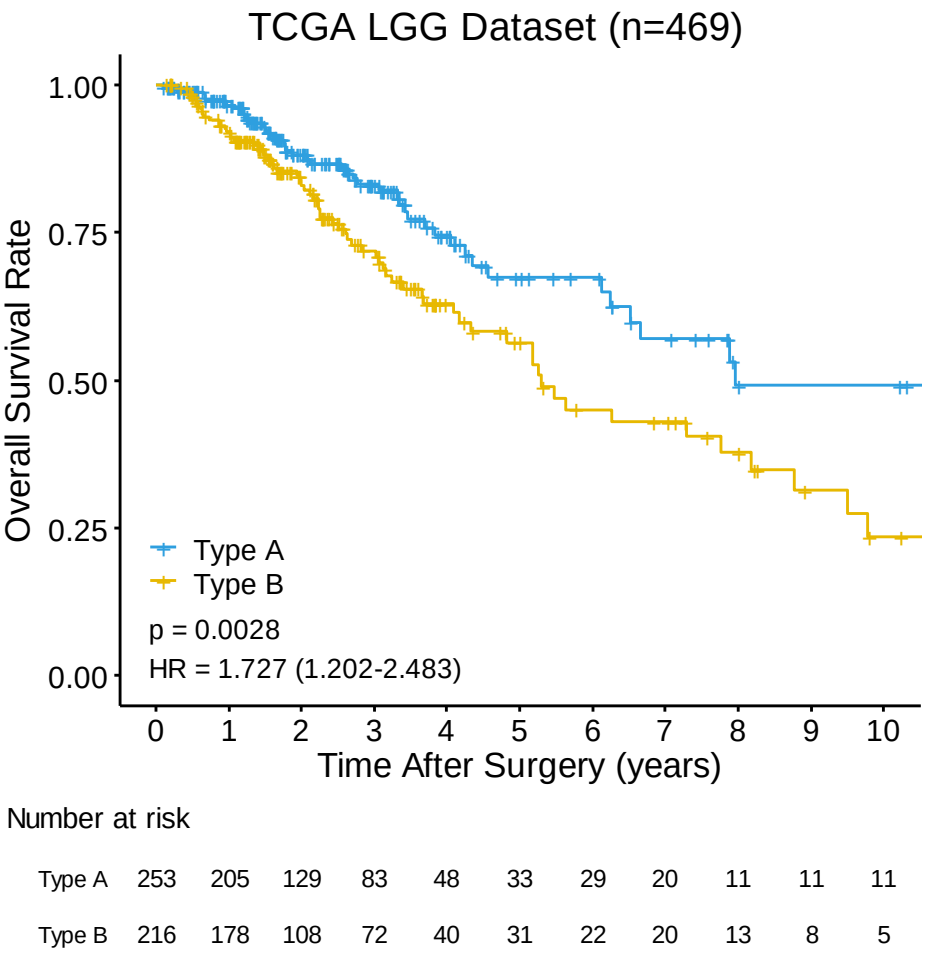

Figure S2. The use of LASSO Cox regression model to identify the most powerful prognostic markers. (A) Tuning parameter (λ) selection in the LASSO model via 10-fold cross-validation according to the minimum criteria. Partial likelihood deviance was plotted as a function of log(λ). The left dotted vertical line defines the minimum criteria, which marks the optimal value of λ. The optimal λ value of 0.0262 with log(λ) = -3.64 was chosen. (B) LASSO coefficient profiles of the 120 survival-related DEIRs. The dotted vertical line was drawn at the optimal λ value. The 20 most powerful prognostic markers with nonzero coefficients were selected.

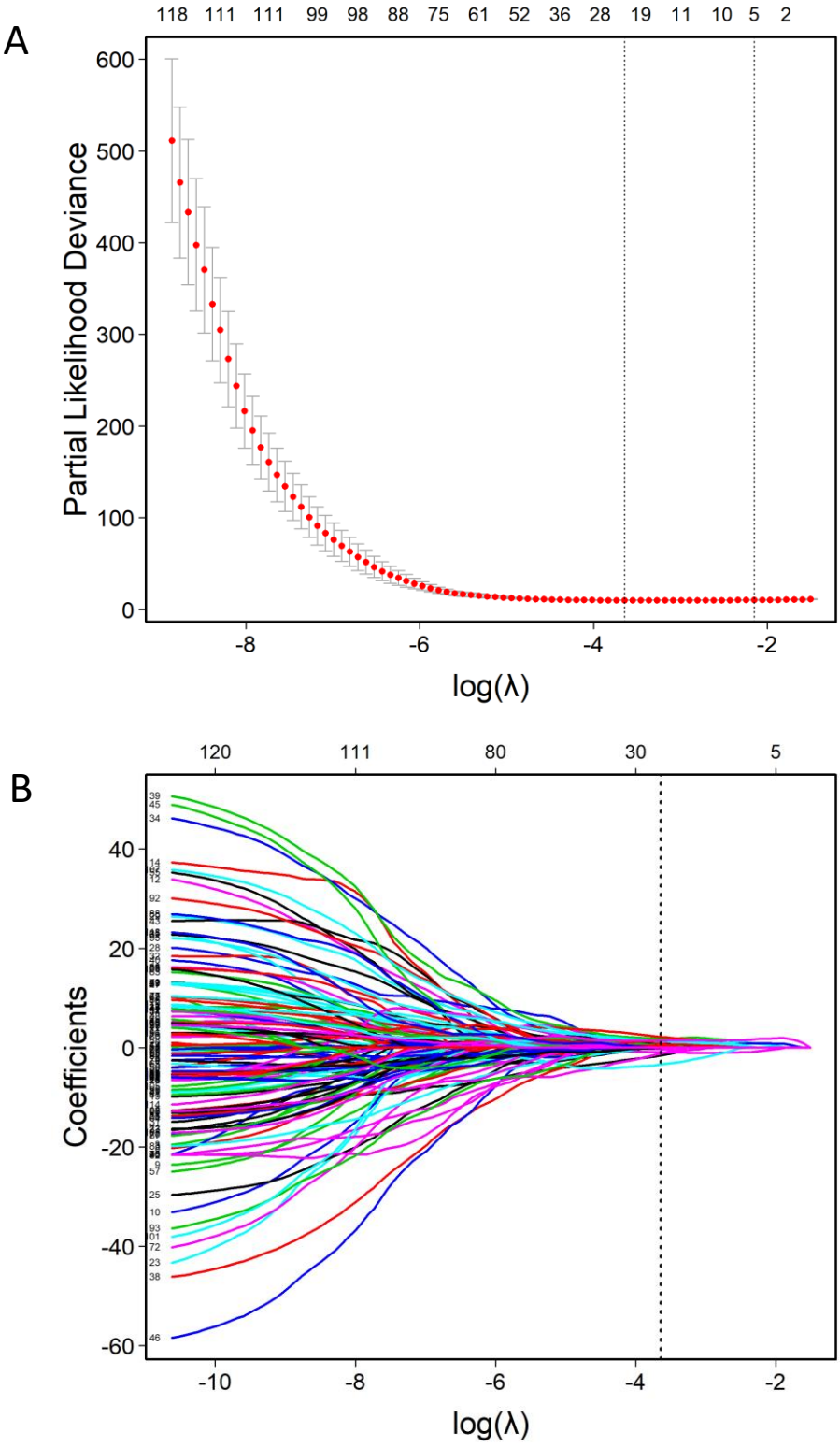

Figure S3. The formula of the risk score model.
Risk Score = (2.1627 × relative expression of IGFBP5) + (1.8334 × relative expression of CENPF) + (1.4131 × relative expression of CD101) + (1.3129 × relative expression of SIGLEC1) + (1.3071 × relative expression of TMPRSS3) + (0.8839 × relative expression of SIGLEC8) + (0.8605 × relative expression of BIRC5) + (0.8552 × relative expression of EMP1) + (0.4835 × relative expression of SPP1) + (0.4357 × relative expression of PDCD1LG2) + (0.3861 × relative expression of FABP5) + (0.2623 × relative expression of CD37) + (0.2018 × relative expression of CD300LF) + (0.0448 × relative expression of ADAMTS3) + (-0.0003 × relative expression of PROK2) + (-0.3417 × relative expression of CBX6) + (-0.9706 × relative expression of GPR27) + (-1.2091 × relative expression of CRYBB1) + (-1.5871 × relative expression of ANKRD22) + (-3.3937 × relative expression of HEY1)

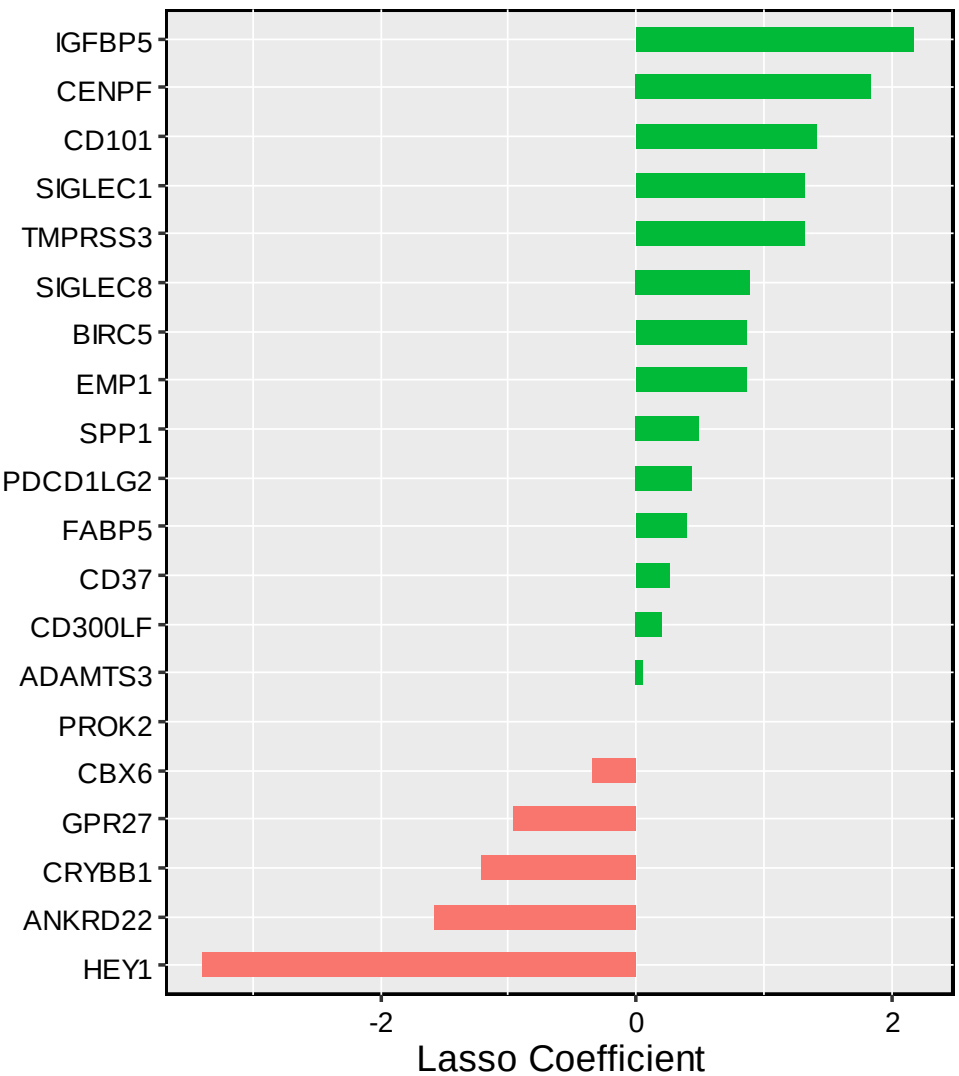

Figure S4. Overrepresented biological processes identified by the 120 survival-related immune metagenes. Most of the biological processes are related to the activation and proliferation of immune cells, indicating that intratumoral immune infiltration plays an important role in the prognosis of LGGs.

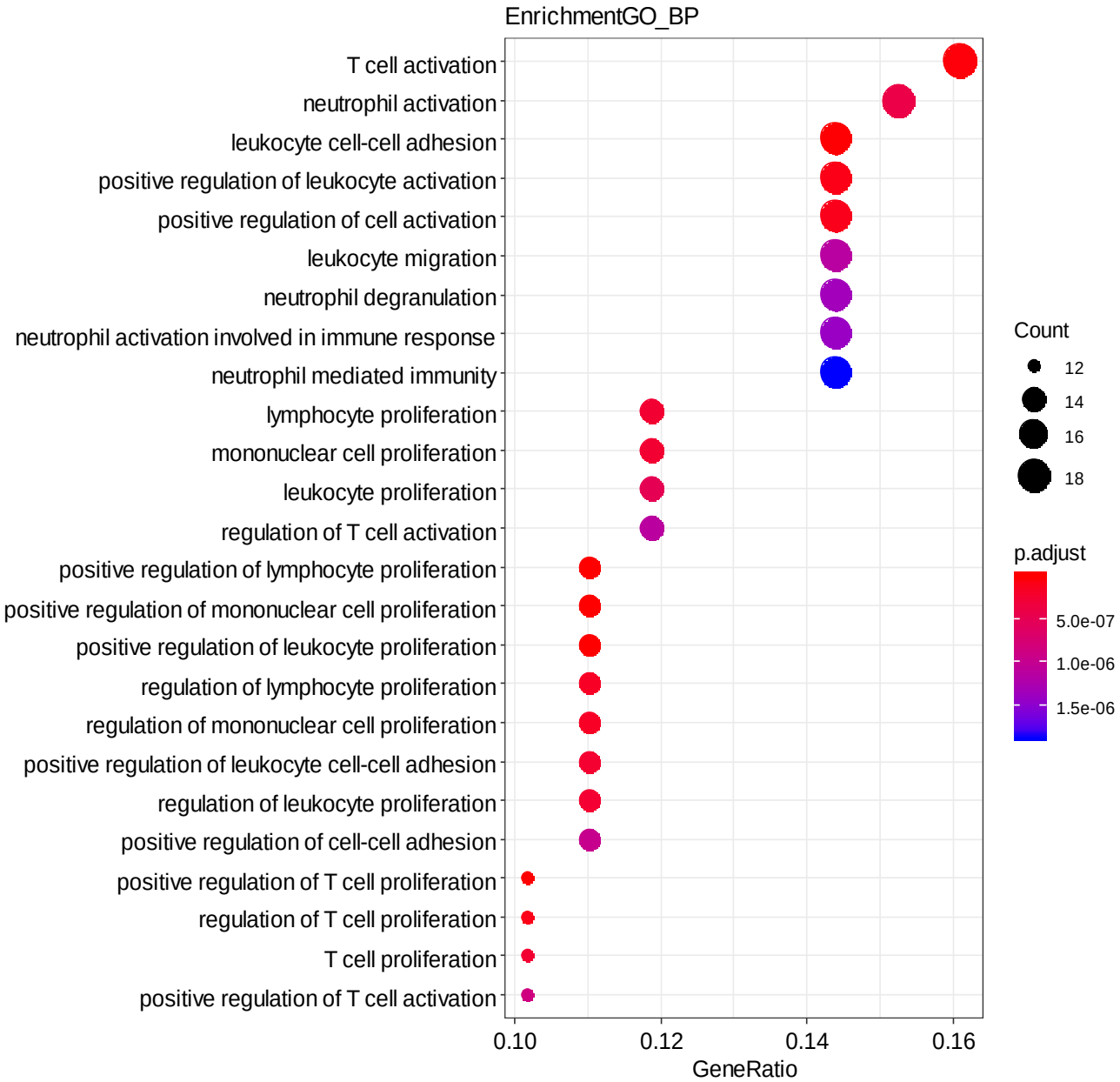

Figure S5. To compare the predictive accuracy of the risk score with traditional clinical factors, time-dependent ROC analysis was applied to the whole dataset. The risk score gave the highest 3-year and 5-year AUC (0.814 and 0.765, respectively), indicating that it was a better predictor of survival for LGG patients.

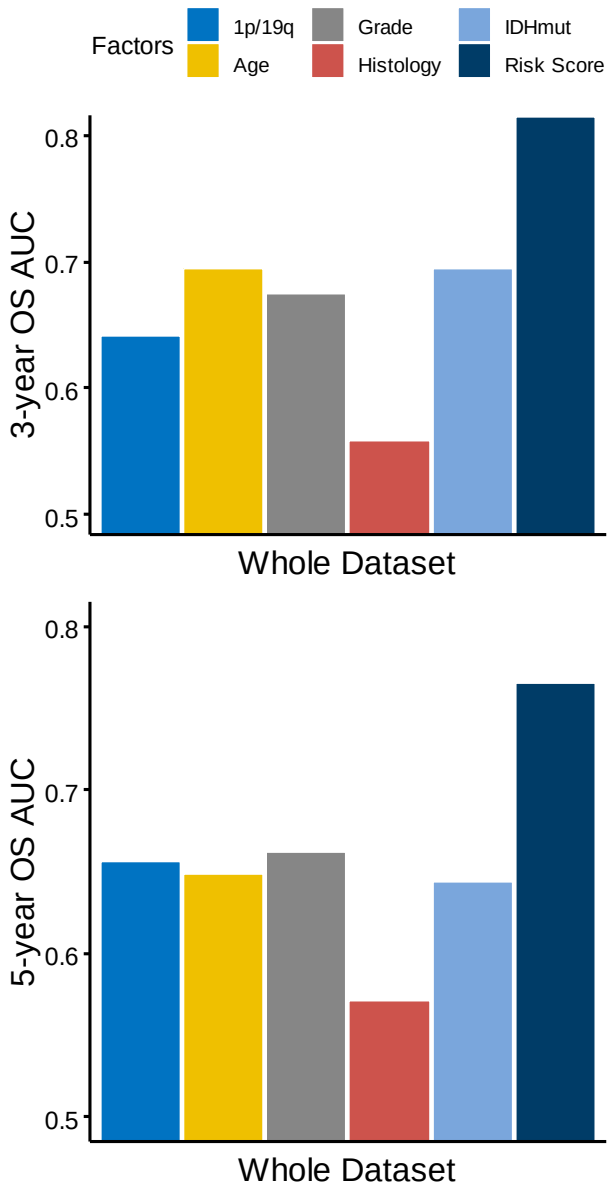

Figure S6. Kaplan-Meier survival analyses stratified by (A) immune-related risk score, (B) age, (C) grade, (D) IDH mutation, (E) 1p/19q status, and (F) histological type in the whole dataset. The HR value of immune-related risk was the highest, indicating that its stratification ability was superior to other prognostic factors.

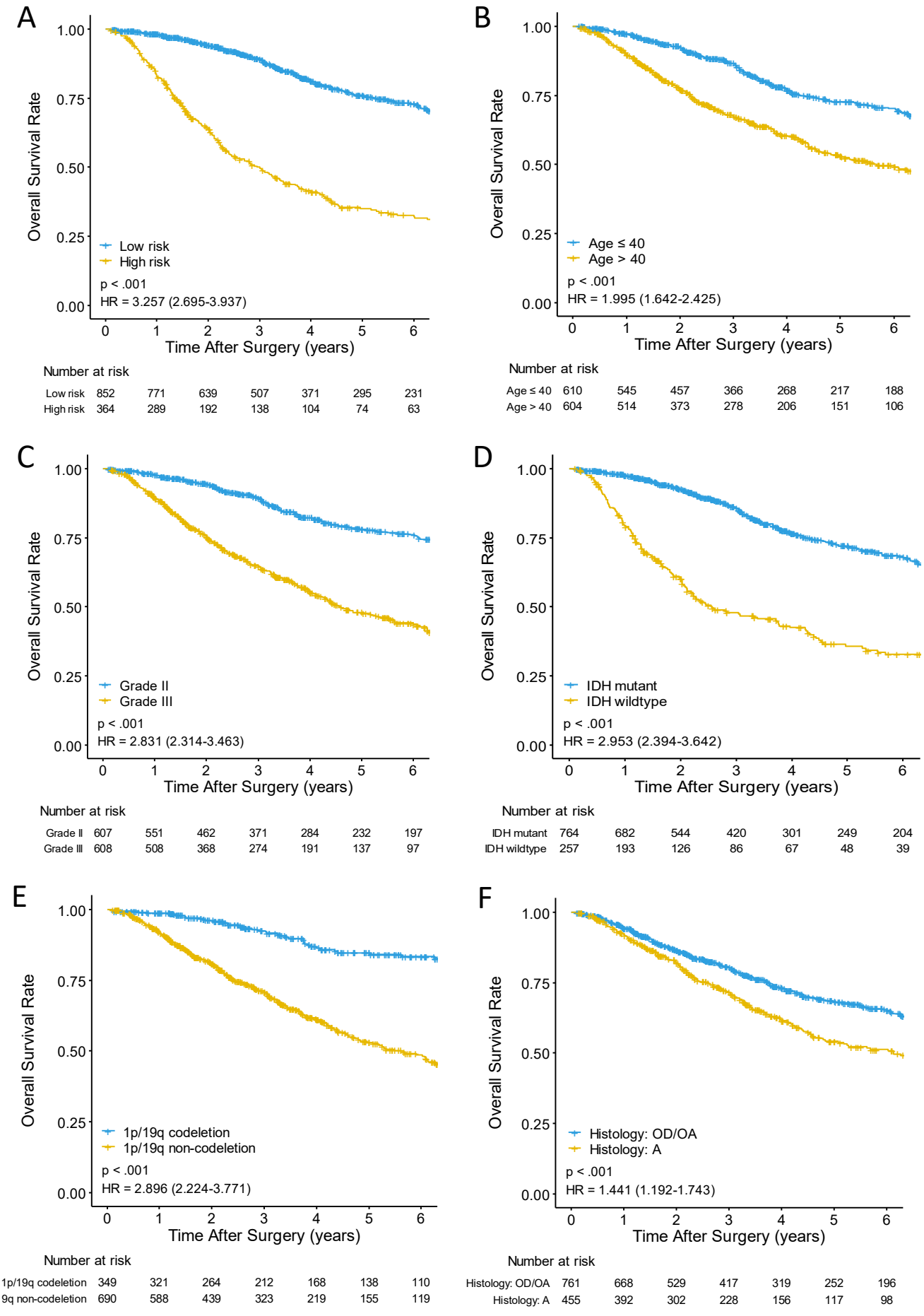

Figure S7. The correlation between the risk score and the expression of PD-L1 in TCGA and CGGA RNA-Seq datasets.

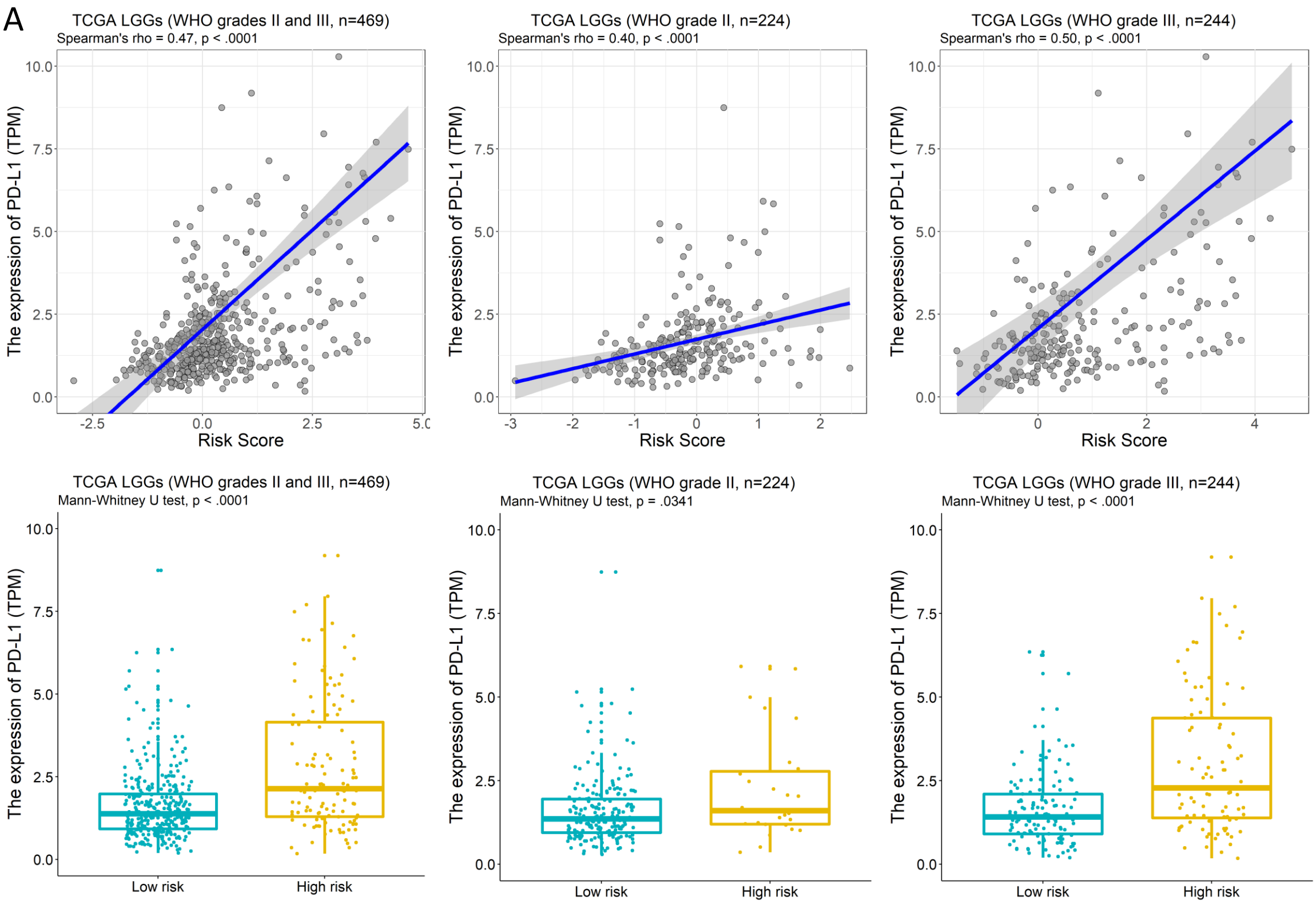

Figure S8. The correlation between the risk score and the expression of TGFβ1 in TCGA and CGGA RNA-Seq datasets.

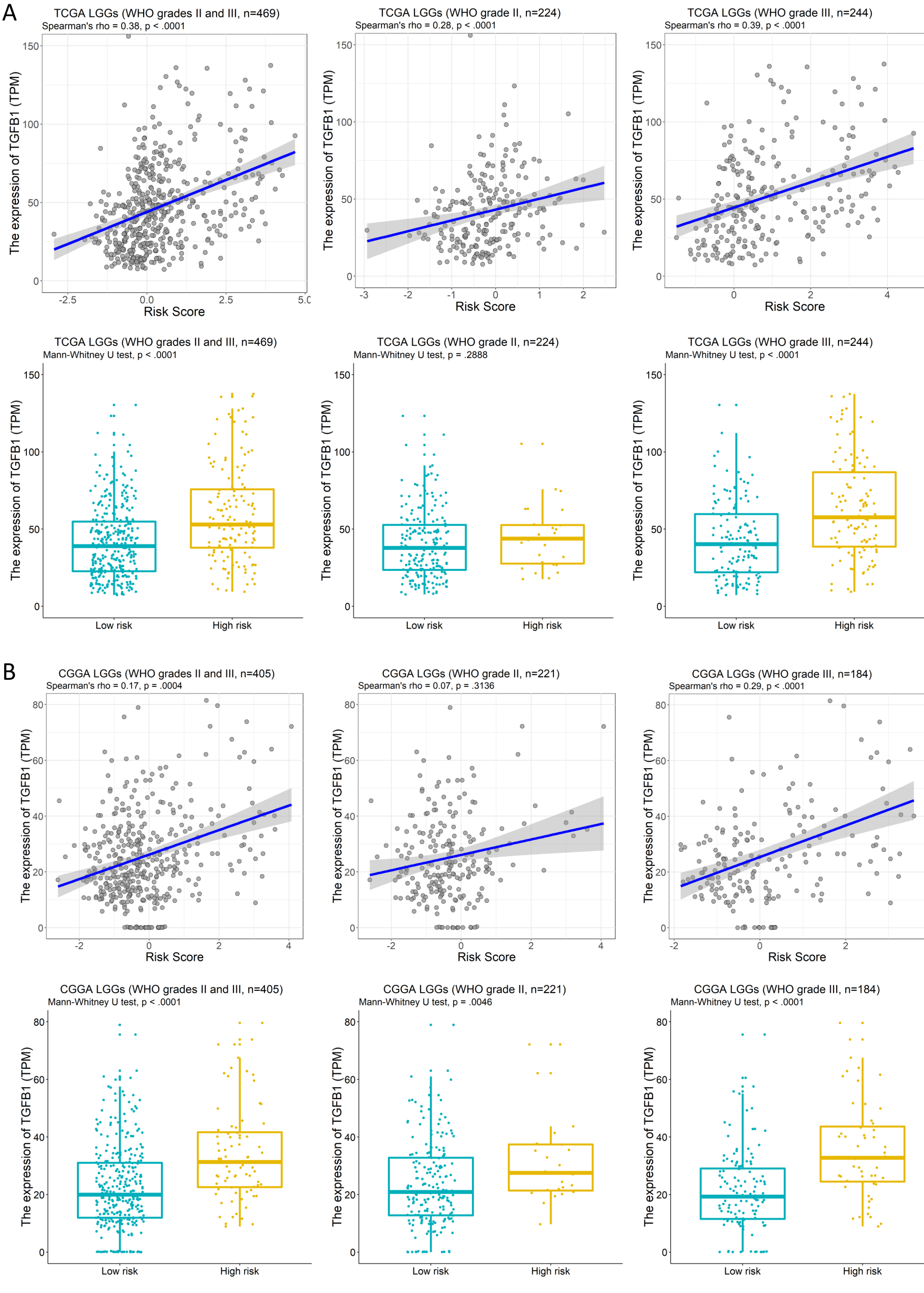

Figure S9. The correlation between the risk score and the expression of TGFβ2 in TCGA and CGGA RNA-Seq datasets.

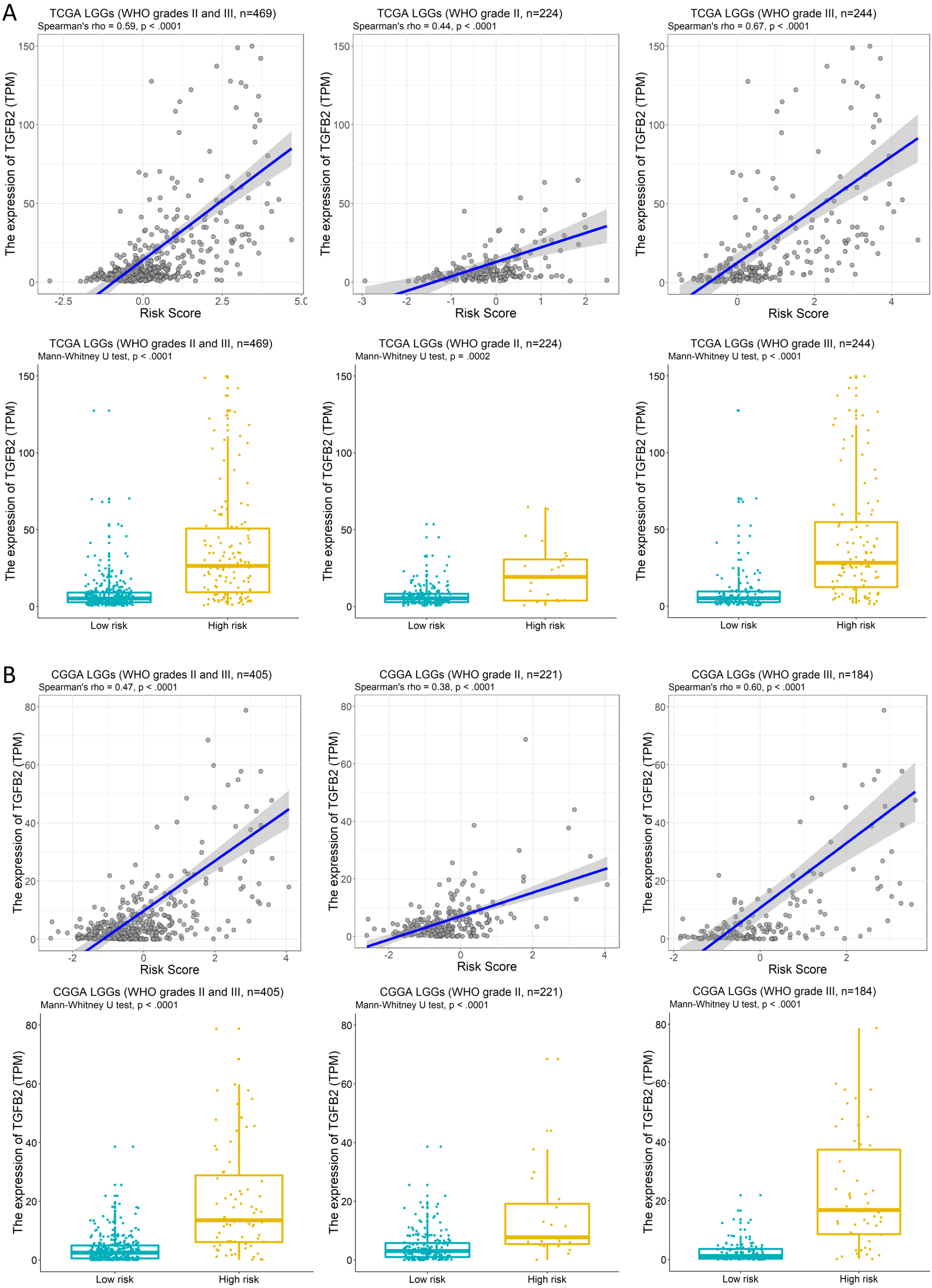

Table S1. Prognostic factors of LGG patients in the six datasets.

Factors	Training Set	Internal Validation	External Validation 1	External Validation 2	External Validation 3	External Validation 4	Whole Dataset	*P**
Total No. of patients	329	140	405	118	88	136	1216	NA
Age								0.120
>40 : ⩽40	171 : 158	69 : 71	191 : 213	49 : 68	53 : 35	71 : 65	604 : 610	
Sex								0.319
Male : Female	181 : 148	75 : 65	236 : 169	71 : 47	58 : 30	NA	621 : 459	
Grade								< .001
III : II	178 : 150	66 : 74	184 : 221	34 : 84	71 : 17	75 : 61	608 : 607	
IDH mutation								< .001
Mutant : Wildtype	256 : 68	118 : 18	276 : 99	72 : 44	42 : 28	NA	764 : 257	
1p/19q status								0.003
Codel : Non-codel	90 : 236	57 : 83	131 : 241	NA	31 : 34	40 : 96	349 : 690	
Histological type								< .001
A : (OD and OA)	120 : 209	56 : 84	111 : 294	60 : 58	22 : 66	86 : 50	455 : 761	
Chemotherapy								< .001
Yes : No	198 : 131	74 : 66	233 : 158	44 : 62	15 : 56	NA	564 : 473	
Radiotherapy								< .001
Yes : No	227 : 61	87 : 31	328 : 67	101 : 13	68 : 0	NA	811 : 172	
Risk score								< .001
High-risk : Low-risk	107 : 222	30 : 110	81 : 324	79 : 39	26 : 62	41 : 95	364 : 852	

Abbreviations: NA, not available; A, astrocytoma; OD, oligodendroglioma; OA, oligoastrocytoma.

* Pearson's chi-squared test was applied to assess the statistically significant differences in the clinical characteristics of the training and validation sets.

Table S2. Multivariate Cox analysis of OS in the whole dataset.

Factors	HR (95% CI)	*P*
Age (per 1 year increase)	1.03 (1.02 to 1.04)	< .001
Grade (III/II)	2.10 (1.56 to 2.84)	< .001
IDH mutation (mutant/wildtype)	0.88 (0.63 to 1.24)	.472
1p/19q status (codel/non-codel)	0.58 (0.40 to 0.84)	.004
Histological type (A/(OD&OA))	1.18 (0.90 to 1.54)	.233
Chemotherapy (yes/no)	1.10 (0.81 to 1.48)	.547
Radiotherapy (yes/no)	0.80 (0.50 to 1.27)	.333
Risk score (per 1 score increase)	1.65 (1.46 to 1.86)	< .001

Abbreviations: A, astrocytoma; OD, oligodendroglioma; OA, oligoastrocytoma.

Table S3. Univariate Cox survival analysis of the 28 immune cell subpopulations based on ssGSEA scores in the TCGA dataset.

Immune Cell Subpopulations	HR*	95% CI	*P* Value
Activated.B.cell	4.90E-06	(4.3e-08 to 0.00055)	< .001
Activated.CD4.T.cell	7.3	(0.49 to 110)	.150
Activated.CD8T.cell	20	(0.14 to 2700)	.230
Activated.dendritic.cell	330000	(740 to 1.5e+08)	< .001
CD56bright.natural.killer.cell	2.30E-07	(1.3e-09 to 4.1e-05)	< .001
CD56dim.natural.killer.cell	0.0063	(0.00028 to 0.14)	.002
Central.memory.CD4.T.cell	6.00E+05	(260 to 1.4e+09)	< .001
Central.memory.CD8.T.cell	400	(0.63 to 260000)	.069
Effector.memeory.CD4.T.cell	0.00017	(6e-06 to 0.0045)	< .001
Effector.memeory.CD8.T.cell	120	(1.5 to 10000)	.033
Eosinophil	4.80E-05	(1.6e-06 to 0.0015)	< .001
Gamma.delta.T.cell	8600	(45 to 1600000)	< .001
Immature.B.cell	50	(0.93 to 2800)	.055
Immature.dendritic.cell	0.011	(5.6e-07 to 230)	.380
Macrophage	0.0087	(0.00011 to 0.68)	.033
Mast.cell	27	(1.9 to 390)	.015
MDSC	24	(4.4 to 130)	< .001
Memory.B.cell	4.2	(0.1 to 170)	.450
Monocyte	6.80E-08	(2.1e-10 to 2.3e-05)	< .001
Natural.killer.cell	7.20E+07	(36000 to 1.5e+11)	< .001
Natural.killer.T.cell	3.80E+09	(1.5e+07 to 9.9e+11)	< .001
Neutrophil	0.0056	(0.00011 to 0.29)	.010
Plasmacytoid.dendritic.cell	0.057	(8.7e-05 to 37)	.380
Regulatory.T.cell	0.0029	(8e-05 to 0.11)	.002
T.follicular.helper.cell	8.00E-04	(3.8e-07 to 1.7)	.067
Type.1.T.helper.cell	3000	(5.2 to 1800000)	.014
Type.17.T.helper.cell	2.90E-06	(3.6e-08 to 0.00024)	< .001
Type.2.T.helper.cell	0.36	(0.011 to 12)	.560

* Per 1 ssGSEA score increase.

Table S4. Characteristics of the 469 LGG patients from TCGA, according to immune infiltration type.

Characteristic	Immune Infitration					Whole Dataset	
	Type A		Type B				
	No.	%	No.	%	*P**	No.	%
Total No. of patients	253	53.9	216	46.1	NA	469	100.0
Age, years					.786		
>40	128	50.6	112	51.9		240	51.2
⩽40	125	49.4	104	48.1		229	48.8
Sex					.257		
Female	121	47.8	92	42.6		213	45.4
Male	132	52.2	124	57.4		256	54.6
Grade					.010		
II	135	53.4	89	41.4		224	47.9
III	118	46.6	126	58.6		244	52.1
IDH mutation					.001		
Mutant	216	86.7	158	74.9		374	81.3
Wildtype	33	13.3	53	25.1		86	18.7
1p/19q status					< .001		
Codeletion	122	48.6	25	11.6		147	31.5
Non-codeletion	129	51.4	190	88.4		319	68.5
Histological type					< .001		
Astrocytoma	61	24.1	115	53.2		176	37.5
Oligoastrocytoma	62	24.5	59	27.3		121	25.8
Oligodendroglioma	130	51.4	42	19.4		172	36.7

Abbreviations: NA, not applicable.

* Pearson's chi-squared test.

Table S5. The associations between clinical variables, genetic variables, and immunological phenotype.

	correlation coefficient	*P*
Age	-0.012	.786
Sex	-0.052	.257
Grade	-0.119	.010
IDH mutation	0.152	.001
1p19q codeletion	0.397	< .001
Histological type	0.33	< .001
Immunological phenotype	1	< .001

Table S6. List of the 120 survival-related immune metagenes common to both TCGA and CGGA RNA-Seq datasets.

Metagene	Cell Type	Immunity
ADAM28	Activated B cell	Adaptive
CD180	Activated B cell	Adaptive
BLK	Activated B cell	Adaptive
CLECL1	Activated B cell	Adaptive
BRIP1	Activated CD4 T cell	Adaptive
CCL5	Activated CD4 T cell	Adaptive
CCNB1	Activated CD4 T cell	Adaptive
ESCO2	Activated CD4 T cell	Adaptive
EXO1	Activated CD4 T cell	Adaptive
KIF11	Activated CD4 T cell	Adaptive
KNTC1	Activated CD4 T cell	Adaptive
NUF2	Activated CD4 T cell	Adaptive
PRC1	Activated CD4 T cell	Adaptive
RTKN2	Activated CD4 T cell	Adaptive
SAMSN1	Activated CD4 T cell	Adaptive
SELL	Activated CD4 T cell	Adaptive
CD37	Activated CD8T cell	Adaptive
CD3D	Activated CD8T cell	Adaptive
CD8A	Activated CD8T cell	Adaptive
GZMK	Activated CD8T cell	Adaptive
MPZL1	Activated CD8T cell	Adaptive
ZAP70	Activated CD8T cell	Adaptive
COL4A1	Central memory CD4 T cell	Adaptive
ITGB2	Central memory CD4 T cell	Adaptive
VIM	Central memory CD4 T cell	Adaptive
FCER1G	Central memory CD8 T cell	Adaptive
SIGLEC1	Central memory CD8 T cell	Adaptive
EZH2	Effector memeory CD4 T cell	Adaptive
PTGS1	Effector memeory CD4 T cell	Adaptive
TFEC	Effector memeory CD4 T cell	Adaptive
C3AR1	Effector memeory CD8 T cell	Adaptive
CCR5	Effector memeory CD8 T cell	Adaptive
FLT3LG	Effector memeory CD8 T cell	Adaptive
HLA-DMB	Effector memeory CD8 T cell	Adaptive
HLA-DPA1	Effector memeory CD8 T cell	Adaptive
HLA-DPB1	Effector memeory CD8 T cell	Adaptive
IFI16	Effector memeory CD8 T cell	Adaptive
ACP5	Gamma delta T cell	Adaptive
BTN3A2	Gamma delta T cell	Adaptive
CD33	Gamma delta T cell	Adaptive
LMNB1	Gamma delta T cell	Adaptive
FABP5	Gamma delta T cell	Adaptive
FADD	Gamma delta T cell	Adaptive
CYBB	Immature B cell	Adaptive
HLA-DQA1	Immature B cell	Adaptive
MS4A6A	Regulatory T cell	Adaptive
CD84	T follicular helper cell	Adaptive
CEBPA	T follicular helper cell	Adaptive
CTSS	T follicular helper cell	Adaptive
PDCD1LG2	T follicular helper cell	Adaptive
SIGLEC7	T follicular helper cell	Adaptive
SIGLEC9	T follicular helper cell	Adaptive
CD48	Type 1 T helper cell	Adaptive
CD53	Type 1 T helper cell	Adaptive
CD68	Type 1 T helper cell	Adaptive
COL5A3	Type 1 T helper cell	Adaptive
EMP1	Type 1 T helper cell	Adaptive
SIT1	Type 1 T helper cell	Adaptive
SIGLEC10	Type 1 T helper cell	Adaptive
SKAP1	Type 1 T helper cell	Adaptive

P2RX5	Type 1 T helper cell	Adaptive
TLR8	Type 1 T helper cell	Adaptive
TRAF1	Type 1 T helper cell	Adaptive
DUSP14	Type 1 T helper cell	Adaptive
ABCA1	Type 17 T helper cell	Adaptive
ANK1	Type 17 T helper cell	Adaptive
ANKRD22	Type 17 T helper cell	Adaptive
IFT80	Type 17 T helper cell	Adaptive
CSRP2	Type 2 T helper cell	Adaptive
LAMP3	Type 2 T helper cell	Adaptive
TMPRSS3	Type 2 T helper cell	Adaptive
BIRC5	Type 2 T helper cell	Adaptive
CDC25C	Type 2 T helper cell	Adaptive
CENPF	Type 2 T helper cell	Adaptive
C1QC	Activated dendritic cell	Innate
C1QB	Activated dendritic cell	Innate
TREM1	Activated dendritic cell	Innate
SLA	Activated dendritic cell	Innate
HDC	CD56bright natural killer cell	Innate
HEY1	CD56bright natural killer cell	Innate
HOXA1	CD56bright natural killer cell	Innate
C1QA	CD56bright natural killer cell	Innate
C1QB	CD56bright natural killer cell	Innate
AKR7A3	CD56dim natural killer cell	Innate
GRIN1	CD56dim natural killer cell	Innate
GPR65	Eosinophil	Innate
PLAU	Immature dendritic cell	Innate
FPR1	Macrophage	Innate
GPR27	Macrophage	Innate
BASP1	Macrophage	Innate
IGSF6	Macrophage	Innate
CD4	Macrophage	Innate
NME8	Macrophage	Innate
CRYBB1	Macrophage	Innate
ADAMTS3	Mast cell	Innate
ARHGAP15	Mast cell	Innate
HSPA6	Mast cell	Innate
SIGLEC8	Mast cell	Innate
CCR2	MDSC	Innate
CD14	MDSC	Innate
CD2	MDSC	Innate
CD86	MDSC	Innate
FERMT3	MDSC	Innate
ITGAL	MDSC	Innate
IKZF1	Monocyte	Innate
CSF2RA	Natural killer cell	Innate
DLL4	Natural killer cell	Innate
FN1	Natural killer cell	Innate
GBP3	Natural killer cell	Innate
IGFBP5	Natural killer cell	Innate
CD101	Natural killer T cell	Innate
SLC7A7	Natural killer T cell	Innate
SPP1	Natural killer T cell	Innate
TREM2	Natural killer T cell	Innate
VCAM1	Natural killer T cell	Innate
ADGRE1	Natural killer T cell	Innate
MMP25	Neutrophil	Innate
CBX6	Plasmacytoid dendritic cell	Innate
PDIA4	Plasmacytoid dendritic cell	Innate
PROK2	Plasmacytoid dendritic cell	Innate
CD300LF	Plasmacytoid dendritic cell	Innate

Table S7. Multivariate Cox analysis of OS in high- and low-risk groups of the whole dataset.

Factors	High-risk group		Low-risk group	
	HR (95% CI)	*P*	HR (95% CI)	*P*
Age (per 1 year increase)	1.04 (1.02 to 1.05)	< .001	1.02 (1.00 to 1.04)	.031
Grade (III/II)	2.43 (1.48 to 3.99)	< .001	2.17 (1.48 to 3.19)	< .001
IDH mutation (mutant/wildtype)	0.50 (0.32 to 0.76)	.001	1.07 (0.56 to 2.07)	.835
1p/19q status (codel/non-codel)	0.73 (0.34 to 1.56)	.421	0.43 (0.28 to 0.68)	< .001
Histological type (A/(OD&OA))	1.16 (0.81 to 1.64)	.408	1.55 (1.02 to 2.36)	.040
Chemotherapy (yes/no)	1.02 (0.66 to 1.57)	.929	1.11 (0.75 to 1.62)	.608
Radiotherapy (yes/no)	0.44 (0.21 to 0.91)	.027	1.05 (0.57 to 1.92)	.882

Abbreviations: A, astrocytoma; OD, oligodendroglioma; OA, oligoastrocytoma.