Appendix In:

SHOX2 is a Potent Independent Biomarker to Predict Survival of WHO Grade II-III Diffuse Gliomas

Yu-An Zhang[a,i,1], Yunyun Zhou,[b,c,1], Xin Luo[d], Kai Song[e], Xiaotu Ma[f], Adwait Sathe[g], Luc Girard[a], Guanghua Xiao[b,d,h], Adi F Gazdar[a,h,i,*]

[a]The Hamon Center for Therapeutic Oncology Research, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

[b]Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

[c]Center of Biostatistics and Bioinformatics, University of Mississippi Medical Center, Jackson, MS 39216, USA

[d]Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

[e]School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China

[f]Department of Computational Biology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN, USA

[g]The Center for Systems Biology, Department of Molecular and Cell Biology, The University of Texas at Dallas, Richardson, TX, 75080, USA

[h]The Simmons Comprehensive Cancer Center; University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

[i]Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

[1]These authors contributed equally.

[*] Corresponding Author

Adi F Gazdar
Hamon Center for Therapeutic Oncology
The University of Texas Southwestern Medical Center
6000 Harry Hines Blvd
Dallas, TX 75390-8593
USA
(fax) 214-648-4940
(phone) 214-648-4921
Email: adi.gazdar@utsouthwestern.edu

Notes on Datasets Acquisitions and Constructions

All publically available methylation and expression datasets used in this study were described in Appendix Table S1. The original beta values of samples were used in all 450K methylation datasets constructed. The values of the RNA rsem.genes.normalized raw data were used and transformed by log2 [RNA Seq V2 RSEM (Log2) in the TCGA LGG and GBM datasets constructed. To avoid any bias, the cases with multiple data entries or from recurrent tumors were excluded from further analyses in the TCGA datasets.

For GSE16011 and REMBRANDT datasets, the raw data of Affymetrix GeneChip Human Genome U133 Plus 2.0 [HG-U133_Plus_2] microarray were downloaded on the dates as described in Table S1, and processed by us into MAS5.0 and quantile normalized expression levels, using MATRIX 1.508 program. Among three probes (probe 210135_s_at, 208443_x_at, 210134_s_at) detecting *SHOX2* mRNA expressions, *SHOX2* mRNA probe_210135_s_at had the highest specificity and the highest detection signal intensity overall (see GSE16011 dataset) (GeneCards: http://genecards.weizmann.ac.il/geneannot/index.shtml, accession date on March 9 2016). Thus *SHOX2* mRNA probe_210135_s_at was selected and represented in survival analysis of *SHOX2* marker.

Table S1. Lists of datasets[a].

ID of Dataset	Source	Type of Tissues	No. of Cases	Data Types	Latest Date of Accession	Accession link	Source for Clinical and Markers Data	Ref.
GSE53162	NCBI-GEO	Non-M brain	Non-M n=40	Illumina 45K microarray	Dec 14 2015	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53162	NCBI-GEO	1
GSE41826	NCBI-GEO	Non-M brain	Non-M (bulk) n=20	Illumina 45K microarray	Dec 2 2015	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41826	NCBI-GEO	2
TCGA LGG and GBM	TCGA	Non-M brain	Non-M n=4	Illumina 45K microarray	Nov 9 2014	https://tcga-data.nci.nih.gov/docs/publications/tcga/?	TCGA	
GSE44684	NCBI-GEO	Non-M brain	Non-M n=2 (adult)	Illumina 45K microarray	Nov 11 2015	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44684	NCBI-GEO	
GSE44684	NCBI-GEO	PA	PA n=61	Illumina 45K microarray	Nov 11 2015	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44684	NCBI-GEO	3
TCGA LGG	TGCA	LGG	LGG n=516	Illumina 45K microarray	Nov 9 2014	https://tcga-data.nci.nih.gov/docs/publications/tcga/?	TCGA and ref. on the right	4
				RNA-Seq	Feb 23 2016	https://tcga-data.nci.nih.gov/docs/publications/tcga/?	TCGA and ref. on the right	4
TCGA GBM	TGCA	GBM	GBM n=129	Illumina 45K microarray	Nov 6 2014	https://tcga-data.nci.nih.gov/docs/publications/tcga/?	TCGA and ref. on the right	4
				RNA-Seq	April 6 2016	https://tcga-data.nci.nih.gov/docs/publications/tcga/?	TCGA and ref. on the right	4
GSE16011	NCBI-GEO	Non-M, PA, LGG, GBM	Non-M n=8; PA n=8; LGG n=109; GBM n=159	Affymetrix Human Genome U133 Plus 2.0 Array	March 31 2016	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16011	NCBI-GEO and Ref. on the right	5
GSE30336	NCBI-GEO	LGG, other	LGG n=48; other/non-glioma: n=4	Affymetrix Human Genome U133 Plus 2.0 Array	March 8 2016	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30336	NCBI-GEO and Ref. on the right	4, 6
GSE30338	NCBI-GEO	LGG, GBM, other	LGG n=68; GBM=9; other/non-glioma: n=4	Illumina 45K microarray	March 2 2016	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30338	NCBI-GEO and Ref. on the right	4, 6
REMBRANDT	ArrayExpress	Non-M, LGG, GBM	Non-M n=14; LGG n=151; GBM =178	Affymetrix Human Genome U133 Plus 2.0 Array	March 26 2016	http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3073/	NCI REMBRANDT, GSE55918, E-MTAB-3073	

GSE43378	NCBI-GEO	LGG, GBM	LGG n=18; GBM=32	Affymetrix Human Genome U133 Plus 2.0 Array and Clinical Variables	March 21 2016	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?targ=self&form=html&view=quick&acc=GSE43378	NCBI-GEO	7
GSE58218	NCBI-GEO	LGG	LGG n=228	Illumina 45K microarray	Jan 19 2016	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58218	NCBI-GEO	8
GSE61160	NCBI-GEO	LGG	LGG n=46	Illumina 45K microarray	March 20 2016	http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61160	Ref. on the right	4, 9

Notes: [a]The additional information for the datasets used in this study can be available upon requests.

1. Non-M: non-malignant; PA: pilocytic astrocytomas; LGG: lower-grade gliomas; GBM: glioblastomas.
2. Illumina 450K microarray: Illumina HumanMethylation450 BeadChip array
3. REMBRANDT dataset was constructed from the following source: (1) NCI REMBRANDT study, https://wiki.nci.nih.gov/display/caIntegrator/caIntegrator+Directory (Access date: march 31, 2016); (2) REMBRANDT Project Betastasis, http://www.betastasis.com/glioma/rembrandt/ (access date: march 14, 2016); (3) GEO GSE55918: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55918 (access date: march 21, 2016) ; (4)E-MTAB-3073 – Molecular Analysis of Brain Neoplasia, ArrayExpress, http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3073/ (access date: march 26, 2016).
4. GSE16011 dataset (Gravendeel L.A.M et al) from the study of glioma samples which were collected from the Erasmus University Medical Center tumor archive (*n* = 276) from patients (1989–2005), Rotterdam, Netherlands;
5. GSE58218 dataset (Wiestler B et al) from the study of primary tumor samples of anaplastic gliomas for DNA methylation analysis (n = 228), which were collected at the Heidelberg University Hospital (Heidelberg, Germany, n = 113) and from the NOA-04 trial (n = 115).
6. GSE61160 dataset: 24 out of 46 LGG cases with clinical information available.

References for Table S1

1. Ladd-Acosta C, Hansen KD, Briem E. et al Molecular psychiatry 2014; 19(8): 862-71.
2. Guintivano J, Aryee MJ, Kaminsky ZA. Epigenetics 2013; 8(3): 290-302.
3. Lambert SR, Witt H, Hovestadt V, et al. Acta neuropathologica 2013; 126(2): 291-301.
4. Ceccarelli M, Barthel FP, Malta TM, et al. Cell 2016; 164(3): 550-63.
5. Gravendeel LA, Kouwenhoven MC, Gevaert O, et al. Cancer research 2009; 69(23): 9065-72.
6. Turcan S, Rohle D, Goenka A, et al. Nature 2012; 483(7390): 479-83.
7. Kawaguchi A, Yajima N, Tsuchiya N, et al. Cancer Sci 2013; 104(9): 1205-10.
8. Wiestler B, Capper D, Sill M, et al. Acta neuropathologica 2014; 128(4): 561-71.
9. Mur P, Mollejo M, Ruano Y, et al. Acta neuropathologica 2013; 126(2): 277-89.

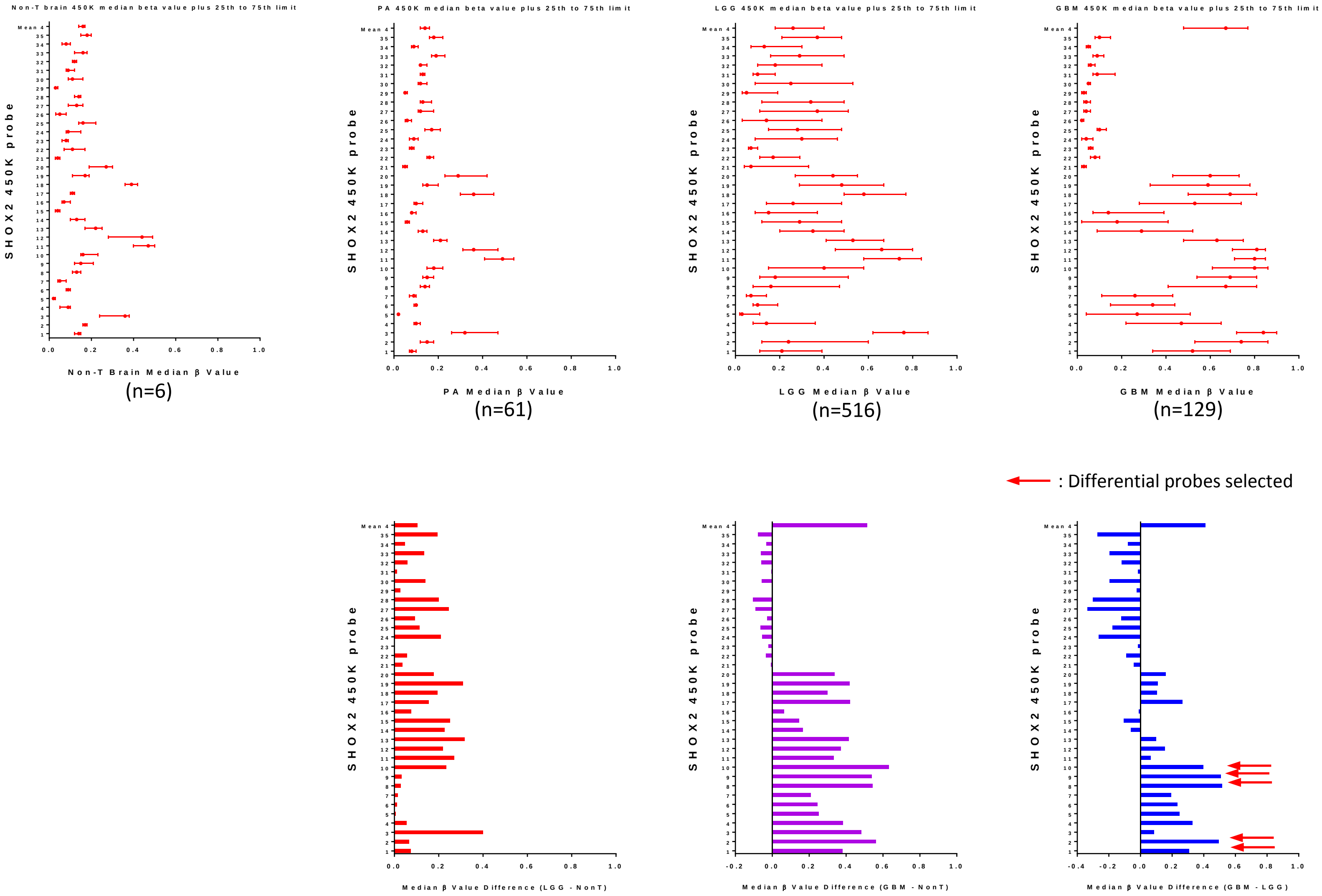


Figure S1. Identifying the highly differential methylation regions in *SHOX2* gene between LGG and GBM through analyses of the 450K methylation datasets. See Appendix Tables S1 and S2 for data source and 450K probe numbers designated and their original array ID. Non-M: non-malignant brain tissues; PA: pilocytic astrocytomas; LGG: lower grade gliomas; GBM: gliobastomas. Bars and lines (upper panel): median and interquartile range.

Table S2A. Part I: Analyses and comparison of *SHOX2* methylations in TCGA LGG and GBM 450K methylation datasets and pilocytic astrocytoma (PA) and non-malignant (Non-M) brain tissues.

Probe No.	SHOX2 450K probe ID	LGG (n=516)				GBM (n=129)				PA (n=61)				Non-M brain (n=6)			
		No. of samples	Median	Mean	Std	No. of samples	Median	Mean	Std	No. of samples	Median	Mean	Std	no. of samples	Median	Mean	Std
1	cg25506747	513	0.21	0.28	0.20	128	0.52	0.51	0.22	60	0.08	0.09	0.04	6.00	0.14	0.15	0.05
2	cg21552242	516	0.24	0.36	0.26	129	0.74	0.67	0.23	61	0.15	0.16	0.07	6.00	0.17	0.17	0.04
3	cg12315391	516	0.76	0.72	0.17	129	0.84	0.79	0.14	61	0.32	0.36	0.13	6.00	0.36	0.32	0.16
4	cg03458172	516	0.14	0.24	0.20	129	0.47	0.45	0.26	61	0.10	0.11	0.04	6.00	0.09	0.08	0.04
5	cg05401764	516	0.03	0.11	0.17	129	0.27	0.32	0.27	60	0.02	0.02	0.01	6.00	0.02	0.02	0.01
6	cg09095222	516	0.10	0.16	0.12	129	0.34	0.32	0.17	61	0.10	0.10	0.01	6.00	0.09	0.09	0.02
7	cg00089486	516	0.07	0.13	0.13	129	0.26	0.29	0.20	61	0.09	0.08	0.02	6.00	0.05	0.06	0.03
8	cg21437028	516	0.16	0.28	0.24	129	0.67	0.59	0.26	61	0.14	0.16	0.08	6.00	0.13	0.13	0.03
9	cg21847036	515	0.18	0.31	0.25	129	0.69	0.65	0.21	61	0.15	0.17	0.07	6.00	0.15	0.16	0.05
10	cg15726154	516	0.40	0.40	0.26	129	0.80	0.71	0.21	61	0.18	0.21	0.09	6.00	0.16	0.18	0.06
11	cg00010946	516	0.74	0.71	0.15	129	0.80	0.77	0.12	61	0.49	0.48	0.10	6.00	0.47	0.44	0.07
12	cg17619405	516	0.66	0.61	0.22	129	0.81	0.76	0.14	61	0.36	0.39	0.12	6.00	0.44	0.40	0.14
13	cg18194945	516	0.53	0.54	0.18	129	0.63	0.59	0.21	61	0.21	0.23	0.07	6.00	0.22	0.22	0.06
14	cg04521004	516	0.35	0.36	0.18	129	0.29	0.31	0.22	61	0.13	0.15	0.07	6.00	0.13	0.15	0.08
15	cg12993163	516	0.29	0.32	0.23	129	0.18	0.24	0.24	61	0.06	0.07	0.03	6.00	0.04	0.04	0.02
16	cg09542210	516	0.15	0.25	0.22	129	0.14	0.26	0.24	61	0.08	0.09	0.01	6.00	0.07	0.08	0.02
17	cg20501518	516	0.26	0.32	0.21	129	0.53	0.51	0.26	61	0.10	0.12	0.04	6.00	0.11	0.11	0.03
18	cg05512327	516	0.58	0.61	0.17	129	0.69	0.65	0.20	61	0.36	0.38	0.11	6.00	0.39	0.40	0.05
19	cg04913979	516	0.48	0.48	0.24	129	0.59	0.55	0.27	61	0.15	0.19	0.10	6.00	0.17	0.15	0.06
20	cg09220088	516	0.44	0.43	0.20	129	0.60	0.56	0.21	60	0.29	0.32	0.11	6.00	0.27	0.24	0.08
21	cg01557547	516	0.07	0.21	0.25	129	0.03	0.05	0.09	61	0.05	0.06	0.04	6.00	0.04	0.04	0.01
22	cg04532033	516	0.17	0.22	0.15	129	0.08	0.09	0.06	61	0.16	0.16	0.03	6.00	0.11	0.12	0.07
23	cg18899952	516	0.07	0.09	0.07	129	0.06	0.07	0.05	61	0.08	0.08	0.03	6.00	0.08	0.07	0.02
24	cg16703882	516	0.30	0.29	0.19	129	0.04	0.07	0.10	61	0.09	0.10	0.06	6.00	0.09	0.11	0.07
25	cg06156376	516	0.28	0.32	0.19	129	0.10	0.13	0.09	61	0.17	0.18	0.06	6.00	0.16	0.18	0.06
26	cg25977493	516	0.14	0.22	0.20	129	0.02	0.04	0.07	61	0.06	0.08	0.07	6.00	0.05	0.06	0.05
27	cg26129769	516	0.37	0.35	0.23	129	0.04	0.06	0.08	61	0.12	0.15	0.07	6.00	0.13	0.12	0.05

28	cg25694447	516	0.34	0.34	0.22	129	0.04	0.06	0.07	61	0.13	0.16	0.07	6.00	0.14	0.13	0.05
29	cg14961949	516	0.05	0.14	0.18	129	0.03	0.05	0.07	61	0.05	0.05	0.02	6.00	0.03	0.03	0.01
30	cg17191178	516	0.25	0.33	0.25	129	0.05	0.08	0.10	61	0.12	0.13	0.04	6.00	0.11	0.12	0.05
31	cg15548431	509	0.10	0.18	0.17	129	0.09	0.13	0.10	61	0.13	0.14	0.03	6.00	0.09	0.10	0.03
32	cg21503297	516	0.18	0.25	0.19	129	0.06	0.08	0.07	61	0.12	0.13	0.03	6.00	0.12	0.12	0.01
33	cg25460158	516	0.29	0.33	0.20	129	0.09	0.11	0.09	61	0.19	0.21	0.05	6.00	0.16	0.16	0.05
34	cg00591153	516	0.13	0.21	0.19	129	0.05	0.06	0.05	61	0.09	0.10	0.03	6.00	0.08	0.08	0.03
35	cg24317285	516	0.37	0.36	0.18	129	0.10	0.13	0.10	61	0.18	0.19	0.04	6.00	0.18	0.17	0.06
	Mean of 4 450K probes (cg25506747, cg21552242,cg21437028 and cg21847036)	516	0.26	0.31	0.18	129	0.67	0.60	0.21	61	0.14	0.15	0.04	6.00	0.16	0.15	0.03

Notes:

1. See Appendix Table S1 for lists of dataset.
2. The 450K data for non-malignant adult brain samples were from these six samples (TCGA-FG-5963, TCGA-FG-5964,TCGA-74-6573, TCGA-06-AABW from TCGA; GSM1088933, GSM1088934 from GEO), based on availability at time of analyses.

Table S2B. Part II: Analyses and comparison of *SHOX2* methylations in TCGA LGG and GBM 450K methylation datasets and pilocytic astrocytoma (PA) and non-malignant (Non-M) brain tissues.

Probe No.	SHOX2 450K probe ID	Median beta value difference and p values (student t)											
		GBM - LGG	p (GBM -LGG)	LGG - PA	p (LGG - PA)	LGG - Non-M Brain	p (LGG - Non-M Brain)	GBM - PA	p (GBM - PA)	GBM - Non-M Brain	p (GBM - Non-M Brain)	PA - Non-M Brain	p (PA - Non-M Brain)
1	cg25506747	0.308	5.51E-27	0.128	3.34E-12	0.073	1.20E-01	0.436	7.41E-32	0.381	1.40E-04	-0.055	2.79E-03
2	cg21552242	0.494	2.09E-30	0.089	1.66E-08	0.067	7.46E-02	0.583	6.60E-39	0.561	4.59E-07	-0.022	9.53E-01
3	cg12315391	0.084	2.01E-05	0.441	3.29E-48	0.400	1.05E-08	0.526	1.14E-49	0.484	2.11E-13	-0.042	4.35E-01
4	cg03458172	0.328	1.04E-22	0.044	7.54E-07	0.055	5.24E-02	0.372	9.70E-20	0.384	6.45E-04	0.012	7.37E-02
5	cg05401764	0.247	1.47E-25	0.006	6.03E-05	0.005	2.12E-01	0.253	1.57E-15	0.252	7.15E-03	-0.001	5.31E-01
6	cg09095222	0.232	3.61E-32	0.006	4.11E-05	0.012	1.31E-01	0.238	9.12E-20	0.245	1.05E-03	0.006	7.72E-02
7	cg00089486	0.193	2.56E-26	-0.020	9.37E-03	0.015	2.31E-01	0.173	3.50E-14	0.208	5.35E-03	0.035	1.64E-02
8	cg21437028	0.515	4.11E-35	0.023	8.59E-05	0.028	1.11E-01	0.539	2.97E-27	0.543	2.21E-05	0.004	3.09E-01
9	cg21847036	0.506	1.85E-40	0.031	1.49E-05	0.033	1.41E-01	0.538	2.22E-40	0.539	1.05E-07	0.001	7.44E-01
10	cg15726154	0.397	7.60E-33	0.216	1.36E-08	0.234	4.13E-02	0.613	3.75E-41	0.631	2.22E-08	0.018	5.43E-01
11	cg00010946	0.064	4.83E-06	0.245	2.61E-26	0.269	2.82E-05	0.309	1.51E-39	0.333	1.47E-10	0.024	3.43E-01
12	cg17619405	0.153	8.67E-13	0.299	5.34E-14	0.219	1.67E-02	0.451	1.15E-41	0.372	6.18E-09	-0.079	9.22E-01
13	cg18194945	0.098	3.27E-03	0.325	2.45E-35	0.317	2.85E-05	0.424	1.16E-29	0.415	2.51E-05	-0.008	9.53E-01
14	cg04521004	-0.061	7.53E-03	0.224	7.28E-18	0.227	4.76E-03	0.163	1.54E-07	0.166	8.07E-02	0.003	9.53E-01
15	cg12993163	-0.106	9.39E-04	0.229	5.83E-17	0.251	2.75E-03	0.123	4.01E-08	0.145	3.93E-02	0.021	3.05E-02
16	cg09542210	-0.012	8.11E-01	0.068	4.35E-09	0.075	5.35E-02	0.056	1.15E-07	0.063	7.66E-02	0.008	3.78E-01
17	cg20501518	0.266	1.25E-17	0.156	2.92E-13	0.155	1.66E-02	0.422	1.41E-23	0.421	3.54E-04	-0.001	7.70E-01
18	cg05512327	0.105	4.04E-02	0.228	2.34E-22	0.195	3.13E-03	0.333	7.30E-18	0.300	3.72E-03	-0.033	6.67E-01
19	cg04913979	0.110	3.58E-03	0.325	1.09E-19	0.310	6.47E-04	0.435	3.57E-19	0.419	4.10E-04	-0.016	3.24E-01
20	cg09220088	0.160	4.93E-11	0.153	2.92E-05	0.177	1.89E-02	0.313	1.02E-14	0.337	2.48E-04	0.024	9.33E-02
21	cg01557547	-0.042	3.53E-12	0.018	2.81E-06	0.035	1.02E-01	-0.024	5.67E-01	-0.007	8.12E-01	0.017	3.08E-01
22	cg04532033	-0.091	4.79E-21	0.010	3.12E-03	0.057	1.03E-01	-0.081	1.15E-16	-0.034	2.23E-01	0.047	1.85E-03
23	cg18899952	-0.016	2.63E-05	-0.007	1.99E-01	-0.004	4.45E-01	-0.022	1.85E-02	-0.020	7.76E-01	0.003	3.72E-01
24	cg16703882	-0.266	3.51E-33	0.215	9.06E-14	0.210	1.76E-02	-0.051	2.10E-02	-0.056	4.06E-01	-0.005	9.34E-01

25	cg06156376	-0.177	3.49E-26	0.108	4.34E-08	0.113	6.71E-02	-0.069	6.84E-05	-0.064	2.21E-01	0.005	8.15E-01
26	cg25977493	-0.122	1.24E-20	0.087	8.77E-08	0.093	6.08E-02	-0.034	3.38E-03	-0.028	5.29E-01	0.006	6.25E-01
27	cg26129769	-0.336	8.85E-40	0.250	1.50E-11	0.246	1.31E-02	-0.086	2.62E-11	-0.091	1.03E-01	-0.005	2.85E-01
28	cg25694447	-0.303	2.50E-39	0.211	2.25E-10	0.200	2.09E-02	-0.092	6.92E-15	-0.104	2.65E-02	-0.011	3.61E-01
29	cg14961949	-0.025	3.37E-09	0.003	9.95E-05	0.026	1.27E-01	-0.022	4.66E-01	0.000	5.58E-01	0.023	7.90E-04
30	cg17191178	-0.197	2.35E-25	0.125	3.37E-09	0.139	4.58E-02	-0.072	2.28E-04	-0.058	3.49E-01	0.014	5.30E-01
31	cg15548431	-0.016	1.61E-03	-0.028	6.27E-02	0.012	2.57E-01	-0.044	4.95E-01	-0.005	4.55E-01	0.040	5.37E-03
32	cg21503297	-0.119	1.94E-23	0.056	1.83E-06	0.059	9.09E-02	-0.063	1.25E-09	-0.060	1.13E-01	0.003	2.70E-01
33	cg25460158	-0.195	7.42E-31	0.094	1.20E-06	0.133	3.19E-02	-0.102	1.90E-13	-0.062	2.15E-01	0.039	2.62E-02
34	cg00591153	-0.080	1.34E-18	0.036	5.69E-06	0.047	9.80E-02	-0.043	3.54E-09	-0.033	2.19E-01	0.011	1.86E-01
35	cg24317285	-0.272	1.29E-37	0.185	1.89E-12	0.194	1.14E-02	-0.086	1.46E-05	-0.078	3.50E-01	0.008	2.71E-01
	Mean of 4 450K probes (cg25506747, cg21552242,cg21437028 and cg21847036)	0.411	9.38E-50	0.121	1.43E-11	0.103	3.46E-02	0.532	3.97E-40	0.514	3.85E-07	-0.018	8.17E-01

Notes:

1. See Appendix Table S1 for lists of dataset.
2. The 450K data for non-malignant adult brain samples were from these six samples (TCGA-FG-5963, TCGA-FG-5964,TCGA-74-6573, TCGA-06-AABW from TCGA; GSM1088933, GSM1088934 from GEO), based on availability at time of analyses.

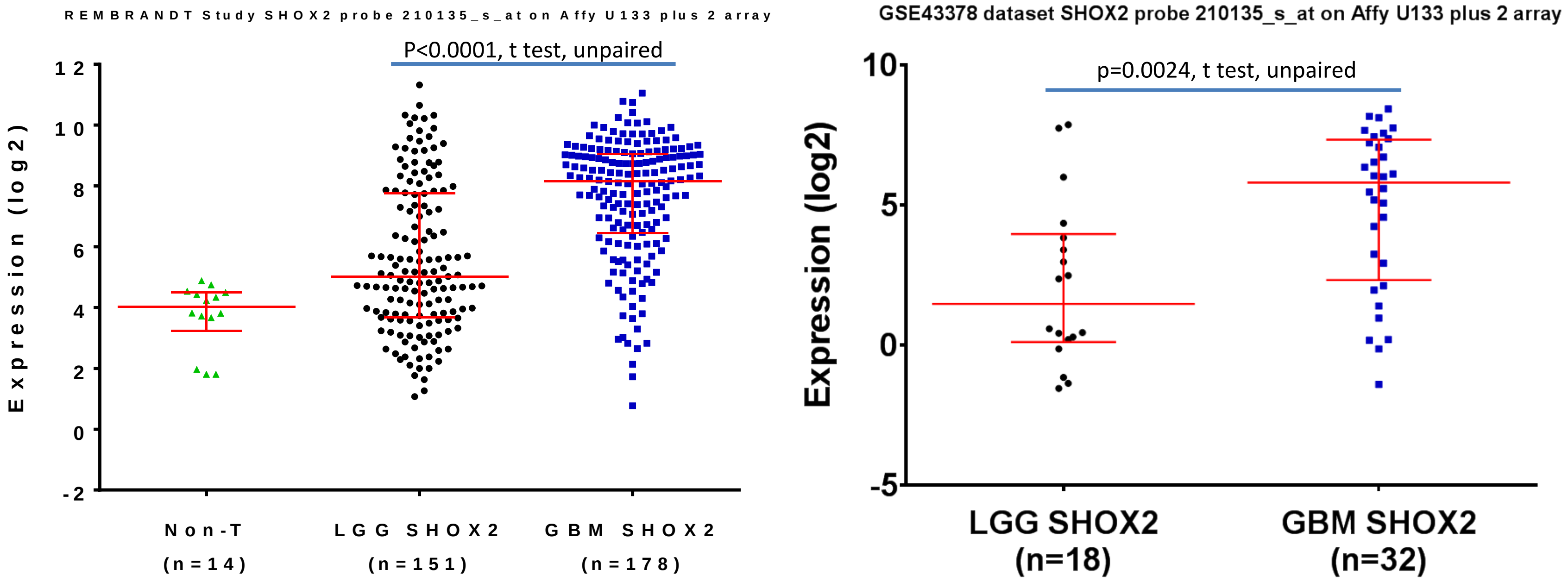


Figure S2. Validation of significantly higher *SHOX2* mRNA expression in GBM than in LGG in additional external datasets (REMBRANDT and GSE43378) besides TCGA LGG and GBM datasets (see Appendix Tables S1for lists of datasets and data source). Non-T: non-tumor brain.

Table S3. Kaplan-Meier survival curve and univariate Cox proportional-hazards model analyses of *SHOX2* 450K probes and MGMT methylation marker in predicting overall survival of LGG patients (TCGA LGG Dataset).

Variable	Cutoff beta value	Events/Patients	Median OS Years (95% CI)	Log-rank test p value	Hazard ratio (95% CI)	p value (Cox model)
MGMT						
Methylated	NA	87/425	7.77 (6.26 to 10.90)	reference	reference	reference
Unmethylated	NA	39/90	3.31 (2.08 to 8.77)	5.89E-07	2.5610 (1.75 to 3.75)	1.40E-06
SHOX2 450K P1						
Low	<0.175	33/224	11.14 (8.77 to NA)	reference	reference	reference
High	≥0.175	91/288	5.25 (4.34 to 7.29)	2.86E-06	2.5239 (1.69 to 3.77)	6.05E-06
SHOX2 450K P2						
Low	<0.225	40/251	8.19 (6.67 to NA)	reference	reference	reference
High	≥0.225	86/264	5.18 (4.18 to 7.96)	5.41E-06	2.3313 (1.60 to 3.39)	1.00E-05
SHOX2 450K P8						
Low	<0.155	45/254	8.77 (7.88 to 12.09)	reference	reference	reference
High	≥0.155	81/261	4.56 (4.08 to 7.59)	1.17E-06	2.4169 (1.67 to 3.49)	2.44E-06
SHOX2 450K P9						
Low	<0.205	46/277	7.88 (6.12 to 12.18)	reference	reference	reference
High	≥0.205	80/236	5.18 (3.03 to 8.19)	4.85E-07	2.4722 (1.72 to 3.56)	1.12E-06
SHOX2 450K P10						
Low	<0.205	26/175	11.59 (6.67 to NA)	reference	reference	reference
High	≥0.205	100/340	5.48 (4.56 to 8.77)	2.95E-05	2.4598 (1.59 to 3.80)	5.18E-05
SHOX2 Ave of 450K Probes 1, 2, 8 & 9						

Low	<0.357	65/354	8.19 (7.29 to 11.19)	reference	reference	reference
High	≥0.357	61/161	2.83 (2.52 to 4.56)	3.45E-12	3.3278 (2.33 to 4.76)	4.30E-11

Notes

1. See Appendix Table S2 for 450K probes numbers designated and their corresponding 450K probe ID.
2. See Appendix Tables S1 for lists of datasets and data source.
3. MGMT methylation status data were based on the Supplemental Table S1A as described previously by Ceccarelli M et al Cell (2016)164:3 (see Appendix Table S1). NA: not available.

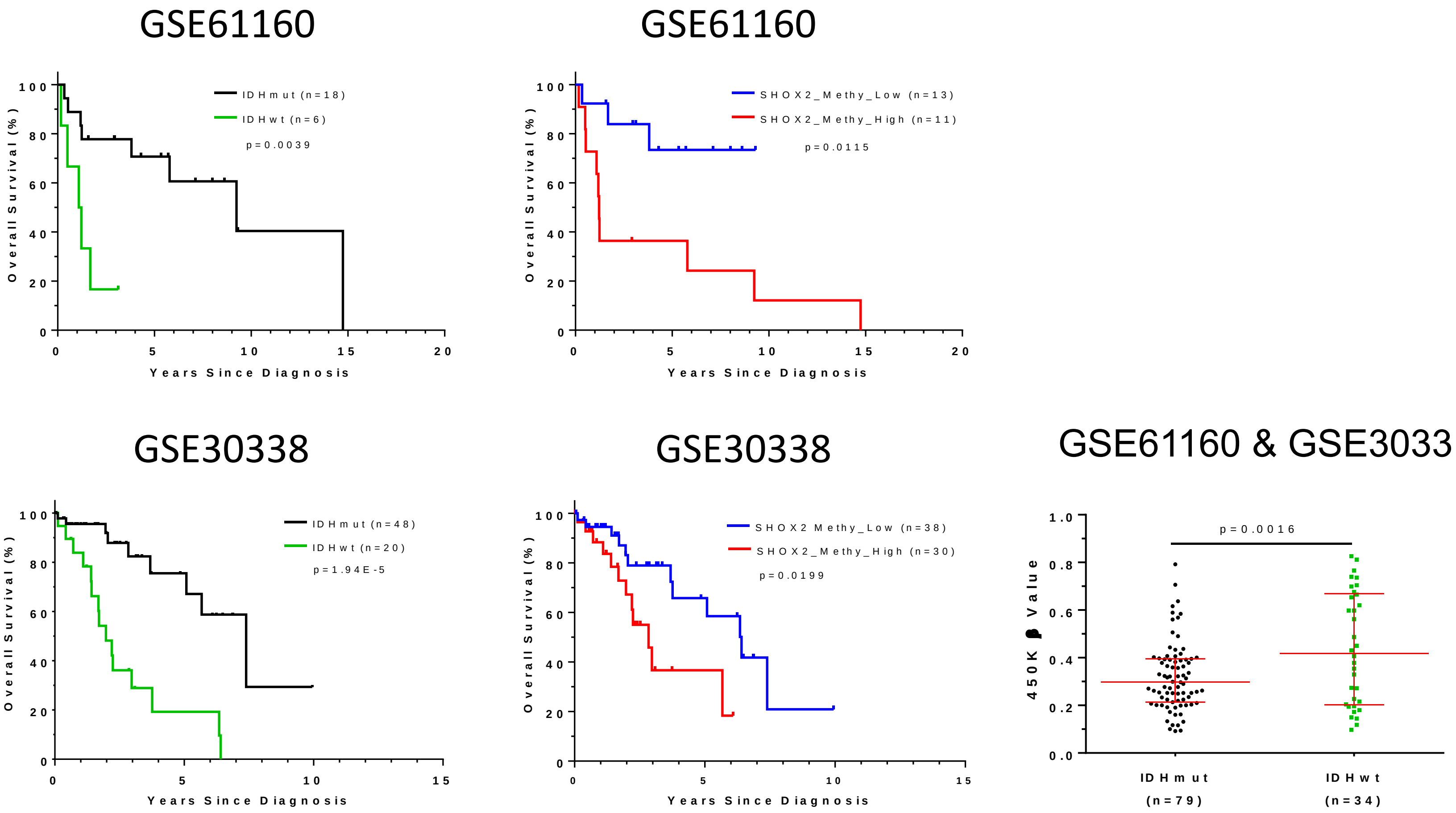


Figure S3. Kaplan-Meier overall survival curve (OS) analyses of *SHOX2* methylation in additional external 450K microarray methylation datasets besides the TCGA LGG dataset. *SHOX2* methylation and *IDH* marker data from GSE61160 and GSE30338 datasets were analyzed and compared individually. The beta cutoff value of pooled mean *SHOX2* probes (P1,2,8,9): 0.357 (see Methods/Materials and Appendix Table S2B). p: Log-rank test. Wt: wild type; mut: mutation. See Appendix Tables S1for lists of datasets and data source.

Table S4. External methylation datasets: The Kaplan-Meier survival curve analyses and univariate Cox proportional-hazards model analyses of *SHOX2* methylation and *IDH* status in predicting overall survival of LGG patients.

Datasets	**Variable**	**IDH Genotype**	**Events/Patients**	**Median OS Years (95% CI)**	**Log-rank test p value**	**Hazard ratio (95% CI)**	**p value (Cox model)**
GSE30338 GSE61160	IDH status						
	IDHmut	mut	17/66	9.24 (5.78 to NA)	reference	reference	reference
	IDHwt	wt	20/26	1.68 (1.40 to 6.35)	7.43E-08	5.59 (2.78 to 11.26)	1.42E-06
GSE30338 GSE61160	SHOX2 expression						
	Low	wt or mut	15/52	7.39 (5.08 to NA)	reference	reference	reference
	High	wt or mut	22/40	2.24 (1.4 to NA)	0.0014	2.88 (1.47 to 5.65)	0.0022
	Low	wt	7/10	3.76 (1.68 to NA)	reference	reference	reference
	High	wt	13/16	1.40 (1.09 to 2.97)	0.0374	3.15 (1.01 to 9.82)	0.0477
	Low	mut	8/42	NA (7.39 to NA)	reference	reference	reference
	High	mut	9/24	5.78 (2.84 to NA)	0.1055	2.21 (0.83 to 5.94)	0.1144

Note: The data from the combined 450K microarray datasets (GSE30338 and GSE61160) were analyzed with increased sample size. The beta cutoff value of pooled mean *SHOX2* probes (P1,2,8,9): 0.357 (see Methods/Materials and Appendix Table S2B). Wt: wild type; mut: mutation. See Appendix Tables S1for lists of datasets and data source. See Appendix Tables S1 for lists of datasets and data source.

Table S5. External expression datasets: The Kaplan-Meier survival curve analyses and univariate Cox proportional-hazards model analyses of *SHOX2* methylation and *IDH* status in predicting overall survival of LGG patients.

Datasets	**Variable**	**IDH Genotype**	**Events/Patients**	**Median OS Years (95% CI)**	**Log-rank test p value**	**Hazard ratio (95% CI)**	**p value (Cox model)**
GSE30336	IDH status						
	IDHmut	mut	8/36	7.39 (5.67 to NA)	reference	reference	reference
	IDHwt	wt	10/12	1.71 (1.42 to NA)	1.49E-04	5.40 (2.04 to 14.29)	6.88E-04
GSE30336	SHOX2 expression						
	Low	wt or mut	12/40	6.41 (5.67 to NA)	reference	reference	reference
	High	wt or mut	6/8	1.54 (1.09 to NA)	2.95E-06	9.91 (3.09 to 31.78)	1.16E-04
	Low	wt	4/6	6.35 (1.71 to NA)	reference	reference	reference
	High	wt	6/6	1.54 (1.09 to NA)	0.0467	4.54 (0.90 to 22.99)	0.0677
	Low	mut	8/34	7.39 (5.67 to NA)	reference	reference	reference
	High	mut	0/2	NA			
GSE16011	IDH status						
	IDHmut	mut	41/47	3.76 (2.37 to 5.97)	reference	reference	reference
	IDHwt	wt	32/41	1.81 (1.14 to 2.87)	3.60E-01	1.25 (0.48 to 1.99)	3.61E-01
GSE16011	SHOX2 expression						
	Low	wt or mut	54/70	5.62 (3.49 to 8.0)	reference	reference	reference
	High	wt or mut	38/39	1.19 (0.79 to 1.80)	2.00E-05	2.48 (1.61 to 3.80)	3.48E-05
	Low	wt	13/21	7.52 (3.71 to 15.23)	reference	reference	reference
	High	wt	19/20	0.66 (0.32 to 1.33)	1.77E-03	3.19 (1.50 to 6.82)	2.67E-03

	Low	mut	27/33	5.62 (2.95 to 10.72)	reference	reference	reference
	High	mut	14/14	2.02 (1.06 to 3.85)	1.43E-02	2.26 (1.16 to 4.41)	1.70E-02
REMBRANDT study	SHOX2 expression by Gene						
	Low	wt or mut	60/104	66.4 (45.3 to 94.5)	reference	reference	reference
	High	wt or mut	46/47	15.3 (12.7 to 22.3)	6.03E-10	3.32 (2.23 to 4.96)	4.12E-09
REMBRANDT study	SHOX2 expression by probe						
	Low	wt or mut	81/126	49.2 (42.5 to 74.9)	reference	reference	reference
	High	wt or mut	25/25	12.0 (9.0 to 17.7)	9.37E-10	4.08 (2.51 to 6.62)	1.31E-08

Notes: The cutoff values by mclust analyses for *SHOX2* mRNA expression by probe (probe_210135_s_at) in each dataset were 5.922 (GSE16011), 5.678 (GSE30336) and 6.575 (REMBRANDT), respectively. The cutoff value by mclust analyses for *SHOX2* expression in the dataset by gene available from REMBRANDT study Betastasis source was 7.315. See Materials/Methods and Appendix Tables S1for lists of datasets and data source. NA: not available.

Table S6. Grade II-III diffuse gliomas distributions classified by *SHOX2* expression and IDH status (TCGA LGG RNA Seq datasets).

	IDHmut	IDHwt	Row Total	p value
All cases				
SHOX2_Low	385 (0.75)	27 (0.05)	412	
SHOX2_High	33 (0.06)	67 (0.13)	100	
Sub-column Total	418	94	512	<1.0E-15
AS grade II-III tumors				
SHOX2_Low	123 (0.64)	13 (0.07)	136	
SHOX2_High	12 (0.06)	44 (0.23)	56	
Sub-column Total	135	57	192	<1.0E-15
OD grade II-III tumors				
SHOX2_Low	155 (0.82)	9 (0.05)	164	
SHOX2_High	14 (0.07)	12 (0.06)	25	
Sub-column Total	169	21	190	4.38E-07
OA grade II-III tumors				
SHOX2_Low	107 (0.82)	5 (0.04)	112	
SHOX2_High	7 (0.05)	11 (0.08)	18	
Sub-column Total (n)	114	16	130	1.58E-08

The RNA-Seq expression data (cutoff value of 4.135 for *SHOX2* expression) from TCGA LGG dataset were analyzed.
Numbers in parentheses: fraction of total each group; p value: Fisher's exact test, Two-sided.
AS: astrocytomas; OD: oligodendrogliomas; OA: oligoastrocytomas.
See Appendix Tab le S1 for detailed sample source and data.

Table S7. The Kaplan-Meier overall survival (OS) curve analyses and univariate Cox proportional-hazards model analyses of *SHOX2* expression by histology and tumor grades (TCGA LGG dataset).

Datasets	**Variable**	**Events/Patients**	**Median OS Years (95% CI)**	**Log-rank test p-value**	**Hazard ratio (95% CI)**	**p value (Cox model)**
Astrocytoma	SHOX2 expression					
	Low	28/138	7.77 (5.62 to NA)	reference	reference	reference
	High	30/56	1.99 (1.62 to 2.83)	3.15E-14	7.28 (4.07 to 13.03)	2.31E-11
Oligodendroglioma	SHOX2 expression					
	Low	31/164	7.96 (6.52 to NA)	reference	reference	reference
	High	12/27	2.77 (2.16 to NA)	0.0055	2.55 (1.29 to 5.03)	0.0071
Grade II	SHOX2 expression					
	Low	31/228	10.90 (7.96 to NA)	reference	reference	reference
	High	7/21	5.25 (2.16 to NA)	0.0319	2.45 (1.05 to 5.70)	0.0373
Grade III	SHOX2 expression					
	Low	44/186	6.25 (4.34 to 11.60)	reference	reference	reference
	High	44/80	1.99 (1.62 to 2.70)	4.64E-14	4.83 (3.09 to 7.55)	5.43E-12

Note:

The RNA-Seq expression data (cutoff value of 4.135 for *SHOX2* expression) from TCGA LGG dataset were analyzed. The grades II or III LGG samples included all histology types in TCGA LGG RNA-Seq dataset analyzed. See Appendix Tables S1for lists of datasets and data source.

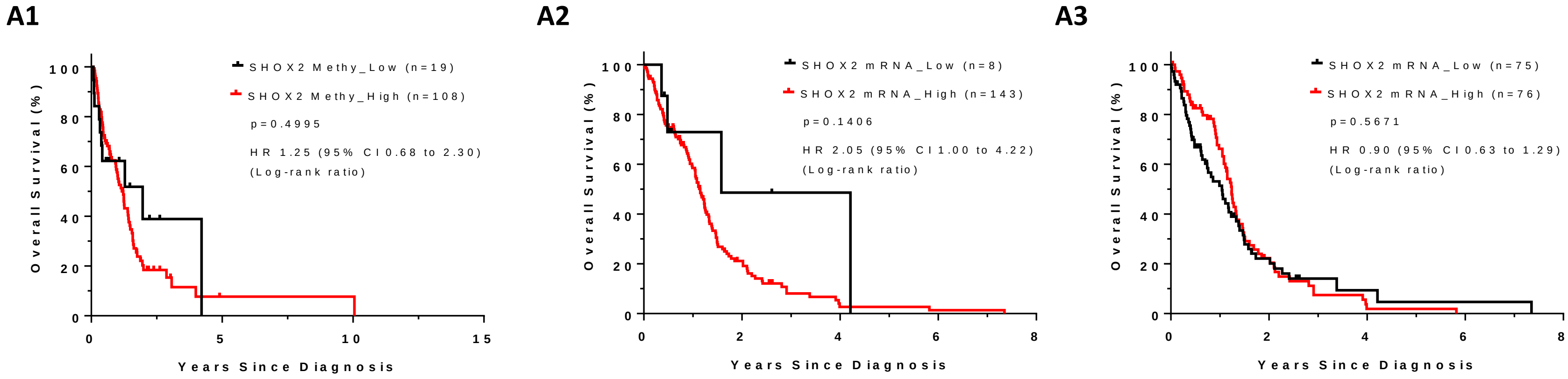
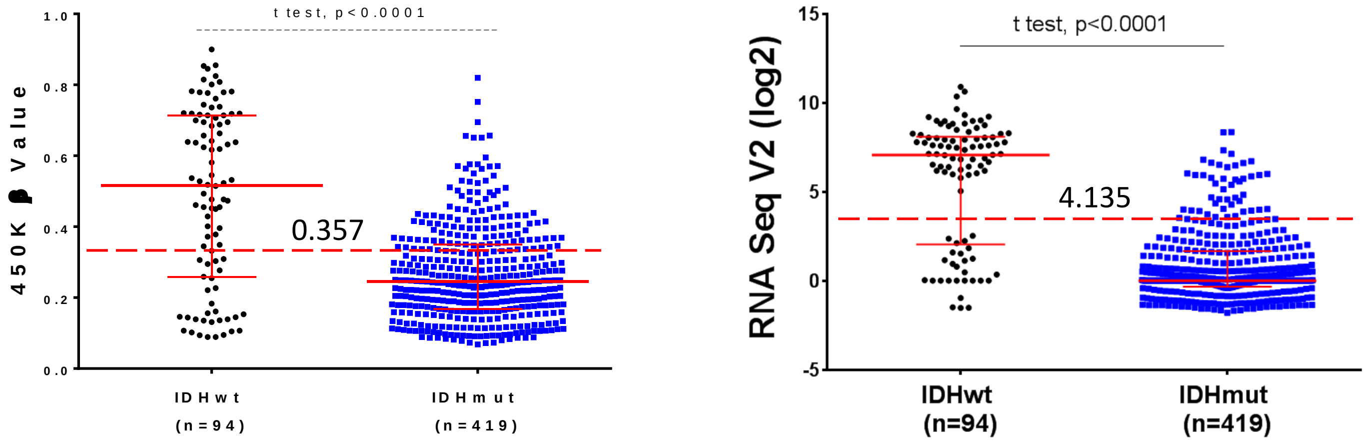


Figure S4. The Kaplan-Meier overall survival curve analyses of *SHOX2* methylation and expression marker in GBM (TCGA GBM 450K microarray and RNA Seq V2 dataset). The cutoff values were used as 0.357 of beta values (A1), 4.135 of RNA-SeqV2 (log2) (A2), respectively, based analyses of large dataset of TCGA LGG dataset (n=516). Alternatively, the analyses data were presented in A3 using 7.835 of median value RNA-SeqV2 of TCGA GBM dataset. p: Log-rank test.. HR: hazard ratio. See Appendix Tables S1for lists of datasets and data source.



	IDHwt	IDHmut
Median β Value	0.52	0.25
% (β>0.357)	66.0	23.6

	IDHwt	IDHmut
Median R.Seq Value	7.1	0.0
% (R.Seq>4.135)	71.3	7.9

Figure S5. Identification of significant associations between high *SHOX2* methylation or expression and the LGG tumors with *IDH* wild-type (*IDHwt*) status (TCGA LGG dataset). The cutoff values to dichotomize samples were 0.357 and 4.135, for the 450K methylation and RNA SeqV2 dataset, respectively, as determined by mclust analysis. Mut: mutation. See Appendix Tables S1 for lists of datasets and data source.

Subset	Events/Patients	Median OS Years (95% CI)	p value (Log-rank test)	HR (95% CI)	p value (Cox model)
TERT_Expressed	37/58 vs 22/166	1.54 (1.35 to 2.08) vs 11.19 (7.96 to NA)	1.32-E-27	17.23 (8.88 to 33.44)	<2.0E-16
SHOX2 mRNA High	43/67 vs 7/33	1.53 (1.19 to 1.99) vs 5.25 (2.77 to NA)	9.99E-10	13.66 (4.80 to 38.89)	9.66E-07
TERTmut	22/37 vs 17/93	1.54 (1.47 to 2.23) vs 7.96 (6.52 to NA)	3.61E-13	13.22 (5.61 to 31.14)	3.57E-09
MGMT_Methy	17/33 vs 68/389	1.99 (1.53 to NA) vs 7.96 (6.52 to 11.60)	<2.0E-16	12.97 (7.12 to 23.62)	<2.0E-16
ATRXwt	46/87 vs 33/238	1.75 (1.53 to 2.23) vs 11.19 (7.96 to NA)	1.43E-25	9.00 (5.54 to 14.62)	<2.0E-16
NES mRNA Low	14/31 vs 43/228	1.99 (1.66 to NA) vs 6.67 (5.38 to 12.20)	2.10E-11	7.00 (3.62 to 13.52)	7.09E-09
NES mRNA High	31/190 vs 36/63	1.62 (1.40 to 2.23) vs 8.77 (7.88 to NA)	2.00E-17	6.76 (4.10 to 11.13)	6.18E-14
1p/19q non-codel	50/94 vs 52/249	1.75 (1.53 to 2.23) vs 7.29 (5.62 to 10.90)	3.26E-21	5.78 (3.86 to 8.66)	<2.0E-16
MGMT_Unmethy	33/61 vs 6/29	1.75 (1.47 to 2.24) vs 8.77 (3.66 to NA)	1.01E-04	4.91 (2.035 to 11.86)	0.0004
TERTwt	7/19 vs 28/143	2.12 (1.22 to NA) vs 6.67 (5.30 to NA)	0.0013	3.77 (1.60 to 8.88)	0.0025
TERT_Not_Expressed	13/35 vs 52/250	2.24 (1.98 to NA) vs 7.29 (5.48 to 10.90)	0.0009	2.78 (1.49 to 5.20)	0.0014
ATRXmut	4/7 vs 41/180	1.54 (0.56 to NA) vs 7.29 (5.48 to10.90)	0.1068	2.38 (0.81 to 6.99)	0.1160
SHOX2 mRNA Low	7/27 vs 67/385	11.14 (2.12 to NA) vs 8.19 (7.29 to 11.6)	0.2994	1.53 (0.69 to 3.39)	0.2980
1p/19q codel	0/0 vs 22/169	NA vs 11.19 (7.96 to NA)	NA	NA	NA

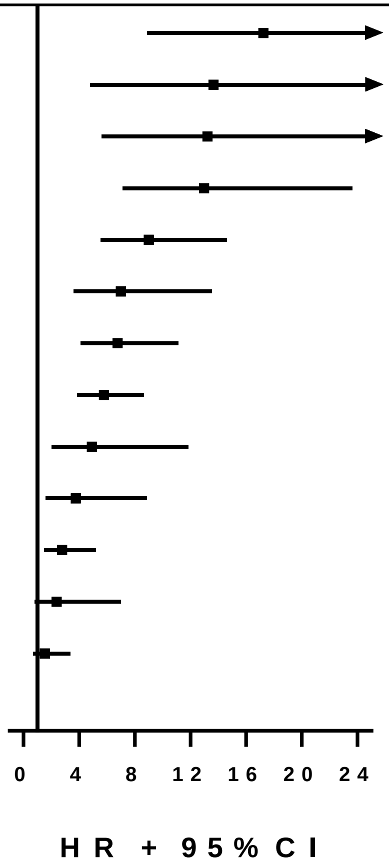
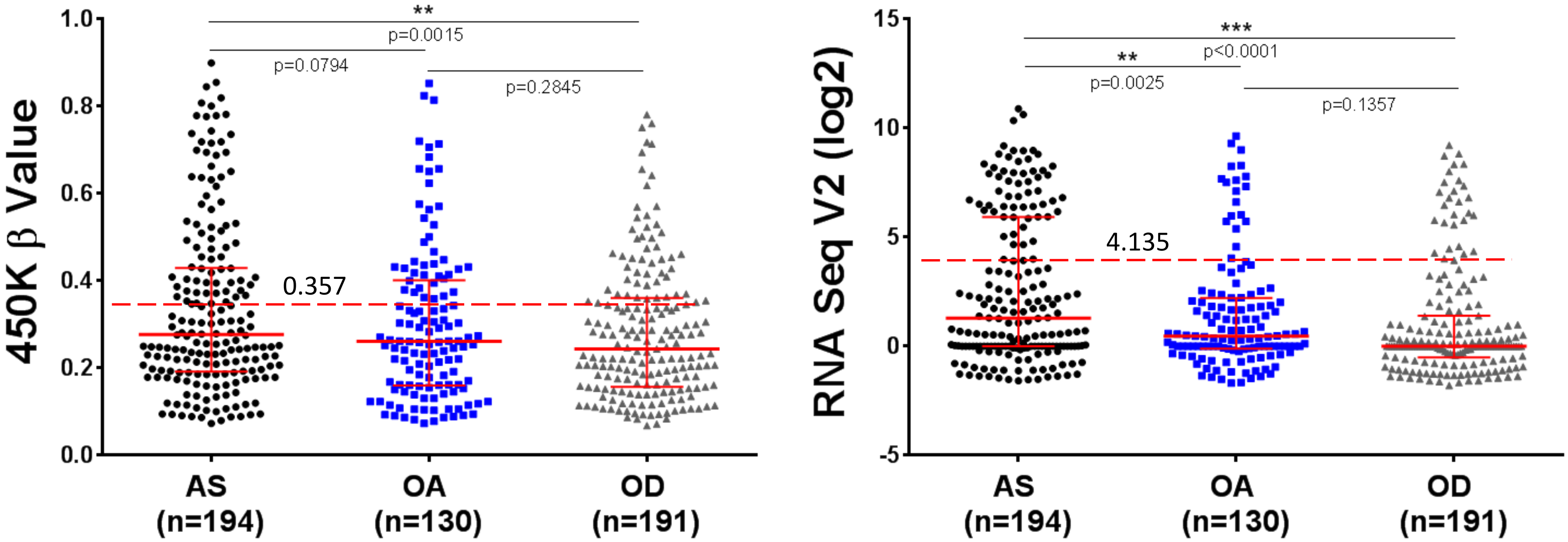


Figure S6. The univariate Cox proportional-hazards model analyses of *IDH* marker (wild-type vs mut) in the subsets of LGG sub-classified by *SHOX2* expression or other relevant markers as individual markers (TCGA LGG datasets). TCGA LGG samples were sub-classified into each subset based on *TERT* (mRNA) expressed or not-expressed, *SHOX2* mRNA expression high or low, *TERT* (promoter) mut or wt, *NES* mRNA expression high or low (using median as cutoff value and see Figure 3B), 1p/19q non-codel or codel, or *MGMT* (promoter) unmethylated or methylated, respectively. See Appendix Tables S1 for lists of datasets and data source. HR: hazard ratio. Arrow sign indicates that upper limit of HR 95% CI is > x axis limit (24). NA: not available.



	AS	OA	OD
Median β Value	0.28	0.26	0.24
% (β>0.357)	36.6	31.5	25.7

	AS	OA	OD
Median R.Seq Value	1.29	0.48	0.00
% (R.Seq>4.135)	28.9	13.8	14.1

Figure S7. Distribution of *SHOX2* methylation and mRNA expression by histology types in LGG patients (TCGA LGG dataset). The cutoff values to dichotomize samples were 0.357 and 4.135, for the 450K methylation and RNA SeqV2 dataset, respectively, as determined by mclust analysis. See Appendix Tables S1for lists of datasets and data source. AS: astrocytoma; OA: oligoastrocytoma; OD: oligodendroglioma. p value: unpaired t test (two-side). Lines and bars: median with interquartile range.

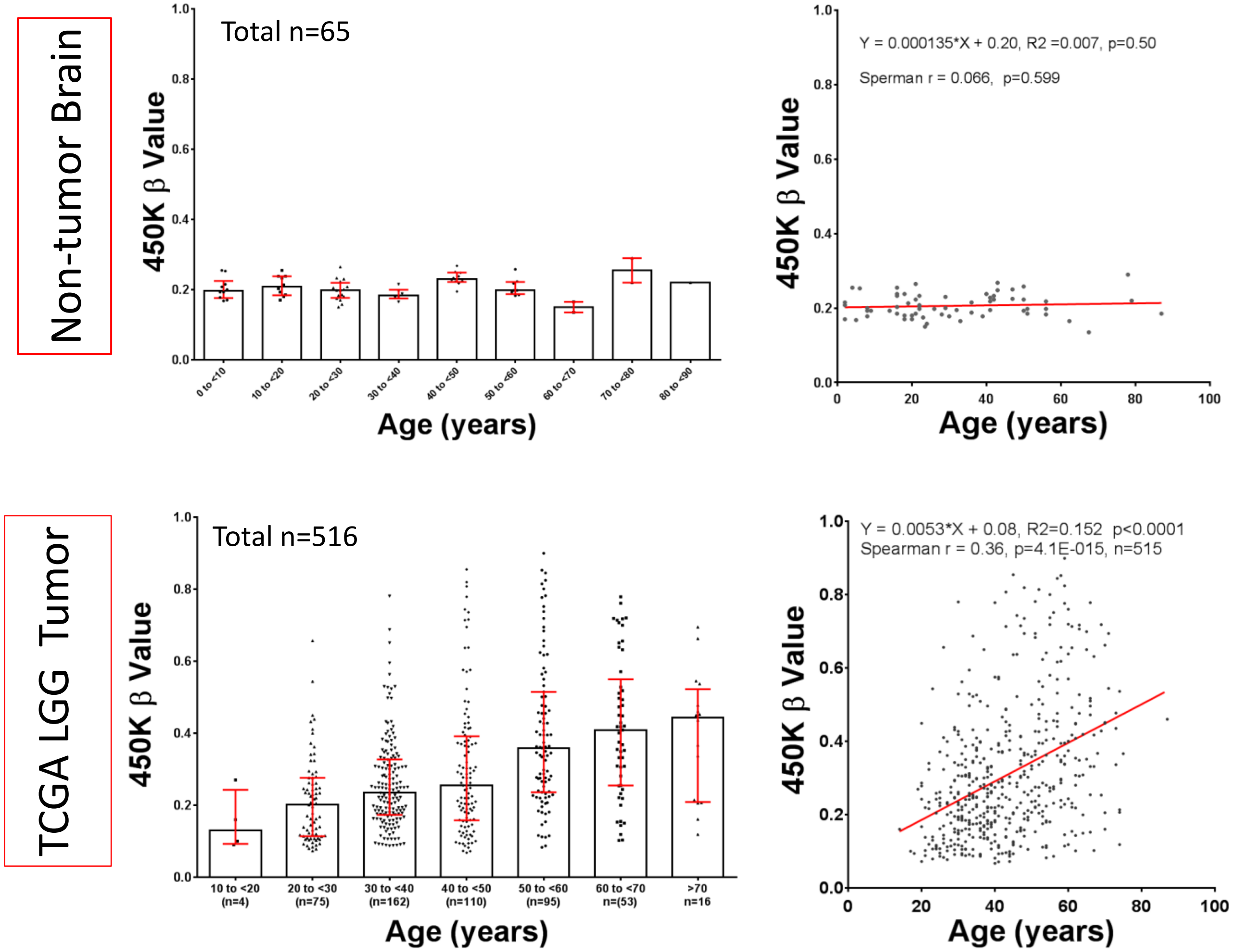


Figure S8. Relationship between *SHOX2* methylation and age in non-malignant brain tissues and LGG gliomas tumors (TCGA LGG dataset). The pooled *SHOX2* 450K methylation probes data (Probes 1, 2, 8 and 9) were presented including non-malignant brain tissues [combined samples from GSE53126, GSE41826, GSE44684, non-malignant brain samples (n=4) from TCGA LGG and TCGA GBM datasets], and TCGA LGG tumor samples. See Appendix Tables S1 and S2 for lists of datasets and data source.

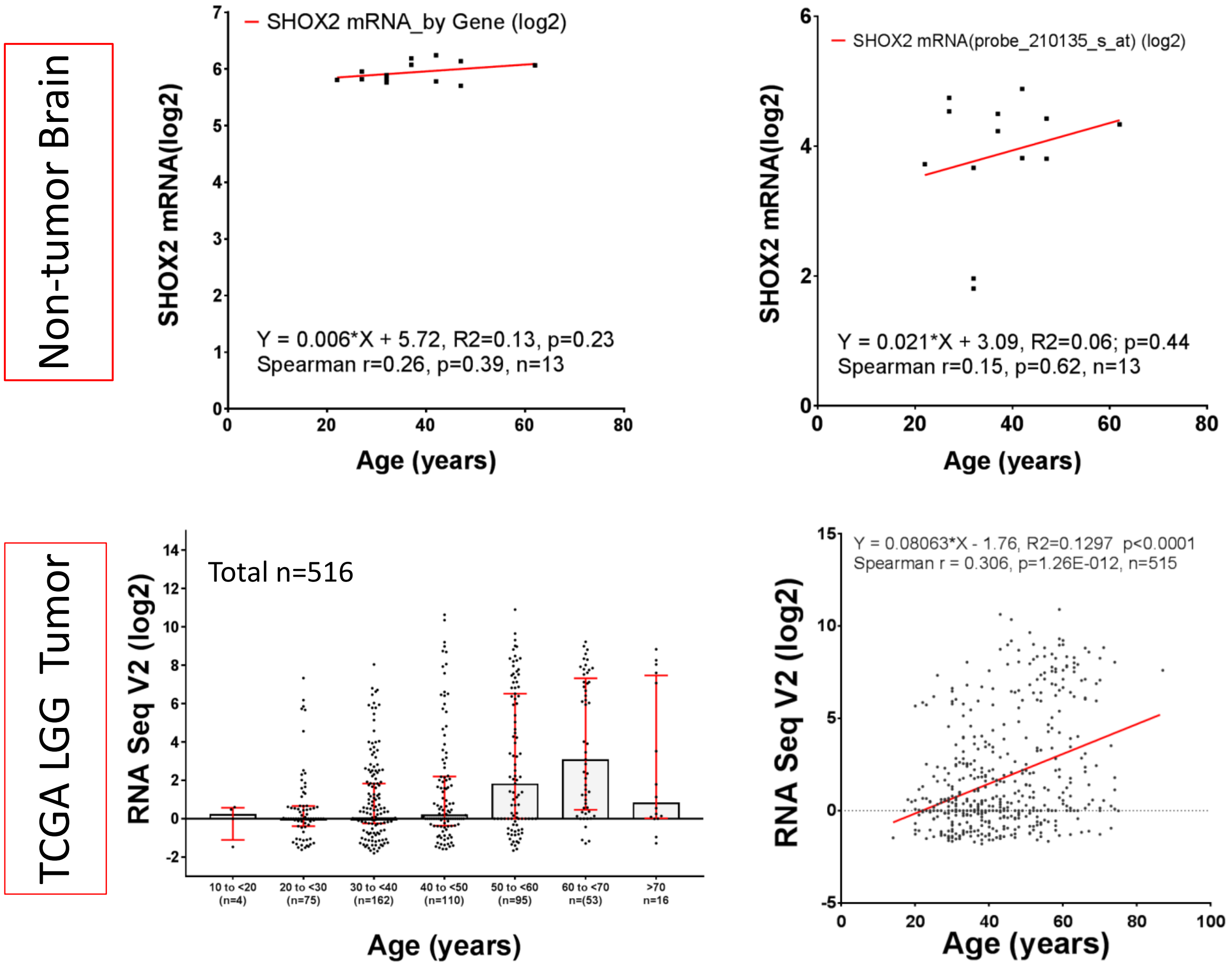


Figure S9. Relationship between *SHOX2* expression and age in non-malignant brain tissues and LGG gliomas tumors (TCGA LGG dataset). *SHOX2* mRNA expression were presented including nonmalignant brain tissues [REMBRANDT microarray dataset (n=13): analyses data by using the values (log2) by Gene or the values by specific array probe], and TCGA LGG tumor samples (RNA SeqV2). See Appendix Tables S1for lists of datasets and data source.

Table S8A. *SHOX2* expression alone: The multivariate Cox proportional-hazards models analyses for LGG patients (TCGA LGG dataset).

Variable	Cases of all age (n=515, events=126)		Cases of age <48 (n=331, events=66)		Cases of age ≥48 (n=184, events=60)	
	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
SHOX2 High vs Low	3.3830 (2.25 to 5.09)	5.12E-09	3.0030 (1.54 to 5.87)	1.29E-03	3.8658 (2.13 to 7.02)	8.76E-06
Age	1.0519 (1.04 to 1.07)	7.88E-10	1.0592 (1.02 to 1.11)	7.36E-03	1.1078 (1.07 to 1.15)	2.01E-07
Gender (male vs female)	1.2979 (0.90 to 1.87)	1.64E-01	1.3032 (0.77 to 2.20)	3.23E-01	1.2321 (0.72 to 2.12)	4.50E-01
Histology (OA vs A)	0.8292 (0.51 to 1.34)	4.45E-01	0.9126 (0.48 to 1.72)	7.78E-01	0.7651 (0.35 to 1.66)	4.98E-01
Histology (OD vs A)	0.5265 (0.34 to 0.81)	3.60E-03	0.6472 (0.35 to 1.20)	1.67E-01	0.5068 (0.27 to 0.94)	3.04E-02
Tumor Grade (3 vs 2)	2.2406 (1.47 to 3.41)	1.70E-04	1.8287 (1.06 to 3.15)	2.94E-02	3.4302 (1.55 to 7.58)	2.32E-03

Table S8B. *IDH* alone: The multivariate Cox proportional-hazards models analyses for LGG patients (TCGA LGG dataset).

Variable	Cases of all age (n=512, events=124)		Cases of age <48 (n=329, events=65)		Cases of age ≥48 (n=183, events=59)	
	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
IDH wt vs mut	3.8205 (2.47 to 5.90)	1.44E-09	2.6501 (1.19 to 5.90)	1.71E-02	5.5110 (2.79 to 10.87)	8.43E-07
Age	1.0506 (1.03 to 1.07)	1.08E-09	1.0678 (1.02 to 1.11)	1.96E-03	1.0847 (1.04 tp 1.13)	7.78E-05
Gender (male vs female)	1.1919 (0.82 to 1.72)	3.51E-01	1.3408 (0.77 tp 2.32)	2.98E-01	0.9173 (0.54 to 1.56)	0.74983
Histology (OA vs A)	0.9104 (0.56 to 1.49)	7.09E-01	1.1026 (0.56 to 2.16)	7.76E-01	0.7478 (0.35 to 1.60)	0.45255
Histology (OD vs A)	0.6758 (0.44 to 1.04)	7.44E-02	0.7245 (0.39 to 1.35)	3.09E-01	0.6939 (0.37 to 1.29)	0.24618
Tumor Grade (3 vs 2)	1.9538 (1.29 to 2.95)	1.48E-03	1.4952 (0.85 to 2.63)	1.62E-01	3.3363 (1.52 to 7.31)	0.00263

Table S8C. *SHOX2* and *IDH* comparison: The multivariate Cox proportional-hazards models analyses for LGG patients (TCGA LGG dataset).

Variable	Cases of all age (n=512, events=124)		Cases of age <48 (n=329, events=65)		Cases of age ≥48 (n=183, events=59)	
	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
SHOX2 high vs low	2.6411 (1.61 to 4.34)	1.27E-04	2.5693 (1.27 to 5.20)	8.69E-03	2.1858 (0.92 to 5.2)	7.78E-02
IDH wt vs mut	2.5258 (1.52 to 4.20)	3.47E-04	2.0250 (0.89 to 4.59)	9.13E-02	3.1987 (1.28 to 7.99)	1.27E-02
Age	1.0468 (1.09 to 1.6)	8.80E-08	1.0614 (1.02 to 1.11)	5.55E-03	1.0917 (1.05 to 1.14)	1.88E-05
Gender (male vs female)	1.3511 (0.93 to 1.97)	1.15E-01	1.4346 (0.83 to 2.49)	1.98E-01	1.0666 (0.61 to 1.86)	8.21E-01
Histology (OA vs A)	0.9541 (0.58 to 1.56)	8.51E-01	1.0950 (0.55 to 2.16)	7.94E-01	0.7953 (0.37 to 1.71)	5.56E-01
Histology (OD vs A)	0.6645 (0.43 to 1.03)	6.68E-02	0.7229 (0.38 to 1.37)	3.18E-01	0.6774 (0.36 to 1.27)	2.22E-01
Tumor Grade (3 vs 2)	1.7658 (1.15 to 2.72)	9.56E-03	1.5317 (0.86 to 2.72)	1.44E-01	2.7635 (1.22 to 6.24)	1.44E-02

Table S8D. *SHOX2* and 1p/19q comparison: The multivariate Cox proportional-hazards models analyses for LGG patients (TCGA LGG dataset).

Variable	Cases of all age (n=512, events=124)		Cases of age <48 (n=331, events=65)		Cases of age ≥48 (n=184, events=60)	
	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value

SHOX2 high vs low	2.8689 (1.87 to 4.39)	1.27E-06	3.5828 (1.79 to 7.15)	2.97E-04	4.0788 (2.03 to 8.18)	7.59E-05
Age	1.0521 (1.04 to 1.07)	5.59E-10	1.0785 (1.03 to 1.12)	3.72E-04	1.1100 (1.07 to 1.15)	2.00E-07
Gender (male vs female)	1.2321 (0.85 to 1.78)	0.26551	1.7750 (1.03 to 3.07)	4.05E-02	1.2607 (0.72 to 2.21)	4.19E-01
Histology (OA vs A)	0.9024 (0.56 to 1.46)	0.67548	1.1415 (0.60 to 2.17)	6.86E-01	0.7558 (0.35 to 1.65)	4.82E-01
Histology (OD vs A)	0.8456 (0.51 to 1.41)	0.52266	1.9511 (0.94 to 4.06)	7.37E-02	0.4740 (0.22 to 1.01)	5.45E-02
Tumor Grade (3 vs 2)	2.3642 (1.55 to 3.61)	6.79E-05	1.8095 (1.05 to 3.13)	3.39E-02	3.3787 (1.52 to 7.51)	2.82E-03
1p/19q non-codel vs codel	2.2966 (1.30 to 4.06)	0.00424	7.2398 (2.80 to 18.67)	4.32E-05	0.8757 (0.36 to 2.10)	7.66E-01

Table S8E. *SHOX2* and *MGMT* comparison: The multivariate Cox proportional-hazards models analyses for LGG patients (TCGA LGG dataset).

Variable	Cases of all age (n=515, events=126)		Cases of age <48 (n=331, events=66)		Cases of age ≥48 (n=184, events=60)	
	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
SHOX2 High vs Low	3.0808 (2.01 to 4.71)	2.09E-07	3.2982 (1.68 to 6.48)	5.29E-04	2.9967 (1.53 to 5.86)	1.33E-03
Age	1.0504 (1.03 to 1.07)	2.48E-09	1.0690 (1.02 to 1.12)	2.61E-03	1.1043 (1.06 to 1.15)	5.25E-07
Gender (male vs female)	1.2983 (0.90 to 1.87)	1.62E-01	1.4097 (0.83 to 2.34)	2.01E-01	1.1780 (0.69 to 2.02)	5.53E-01
Histology (OA vs A)	0.8814 (0.54 to 1.43)	6.08E-01	1.0372 (0.54 to 2.00)	9.14E-01	0.7704 (0.36 to 1.66)	5.06E-01
Histology (OD vs A)	0.5913 (0.38 to 0.92)	1.89E-02	0.7190 (0.38 to 1.35)	3.04E-01	0.5742 (0.31 to 1.08)	8.39E-02
Tumor Grade (3 vs 2)	2.1370 (1.40 to 3.25)	3.98E-04	1.6975 (0.98 to 2.94)	5.95E-02	3.5345 (1.59 to 7.85)	1.91E-03
MGMT unmethy. vs methy.	1.8200 (1.20 to 2.76)	4.77E-03	1.8554 (0.96 to 3.60)	6.78E-02	1.8132 (0.96 to 3.43)	6.70E-02

Table S8F. *SHOX2* and *TERT* mutation comparison: The multivariate Cox proportional-hazards models analyses for LGG patients (TCGA LGG dataset).

Variable	Cases of all age (n=292, events=74)		Cases of age <48 (n=187, events=37)		Cases of age ≥48 (n=105, events=37)	
	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
SHOX2 High vs Low	2.9859 (1.63 to 5.46)	3.85E-04	3.1961 (1.20 to 8.54)	2.05E-02	2.6255 (1.20 to 8.54)	3.89E-02
Age	1.0627 (1.04 to 1.09)	6.31E-09	1.1580 (1.08 to 1.24)	2.29E-05	1.1083 (1.08 to 1.24)	3.45E-04
Gender (male vs female)	0.9278 (0.57 to 1.52)	7.65E-01	1.0177 (0.46 to 2.28)	9.66E-01	0.8725 (0.46 to 2.28)	7.05E-01
Histology (OA vs A)	0.7580 (0.41 to 1.42)	3.85E-01	1.1557 (0.45 to 2.96)	7.63E-01	0.7421 (0.45 to 2.96)	5.23E-01
Histology (OD vs A)	0.5612 (0.30 to 1.05)	7.24E-02	0.9674 (0.34 to 2.74)	9.50E-01	0.4787 (0.34 to 2.74)	9.86E-02
Tumor Grade (3 vs 2)	2.9555 (1.64 to 5.31)	2.94E-04	3.8623 (1.60 to 9.30)	2.58E-03	2.4265 (1.60 to 9.30)	9.75E-02
TERT promoter wt vs mut	1.0876 (0.59 to 2.01)	7.89E-01	2.5300 (1.05 to 6.09)	3.84E-02	0.4577 (1.05 to 6.09)	1.41E-01

Table S8G. *SHOX2* and *TERT* expression comparison: The multivariate Cox proportional-hazards models analyses for LGG patients (TCGA LGG dataset).

Variable	Cases of all age (n=292, events=74)		Cases of age <48 (n=329, events=66)		Cases of age ≥48 (n=183, events=60)	
	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
SHOX2 High vs Low	3.3368 (2.20 to 5.06)	1.31E-08	4.1038 (2.02 to 8.32)	9.11E-05	3.6142 (1.95 to 6.68)	4.19E-05

Age	1.0516 (1.03 to 1.07)	1.40E-09	1.0745 (1.03 to 1.12)	1.07E-03	1.1065 (1.07 to 1.15)	1.53E-07
Gender (male vs female)	1.2941 (0.90 to 1.87)	1.69E-01	1.4630 (0.86 to 2.50)	1.64E-01	1.1458 (0.67 to 1.96)	6.20E-01
Histology (OA vs A)	0.8252 (0.51 to 1.34)	4.34E-01	0.9609 (0.51 to 1.81)	9.02E-01	0.7483 (0.34 to 1.63)	4.64E-01
Histology (OD vs A)	0.5095 (0.31 to 0.83)	6.29E-03	1.1168 (0.54 to 2.33)	7.68E-01	0.4025 (0.21 to 0.77)	6.26E-03
Tumor Grade (3 vs 2)	2.2292 (1.46 to 3.40)	2.03E-04	1.8227 (1.06 to 3.12)	2.90E-02	3.0542 (1.37 to 6.79)	6.19E-03
TER Not_Expressed vs Expressed	0.9363 (0.61 to 1.44)	7.63E-01	2.5675 (1.19 to 5.52)	1.58E-02	0.5064 (0.26 to 0.97)	4.01E-02

Table S8H. *SHOX2* and *NES* comparison: The multivariate Cox proportional-hazards models analyses for LGG patients (TCGA LGG dataset).

Variable	Cases of all age (n=515, events=126)		Cases of age <48 (n=331, events=66)		Cases of age ≥48 (n=184, events=60)	
	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
SHOX2 High vs Low	3.3979 (2.25 to 5.14)	7.21E-09	3.0712 (1.56 to 6.05)	1.17E-03	3.6424 (1.99 to 6.66)	2.74E-05
Age	1.0529 (1.04 to 1.07)	6.52E-10	1.0597 (1.02 to 1.11)	7.02E-03	1.1086 (1.07 to 1.15)	1.01E-07
Gender (male vs female)	1.2649 (0.88 to 1.83)	2.10E-01	1.3066 (0.77 to 2.21)	3.17E-01	1.1468 (0.66 to 1.99)	6.27E-01
Histology (OA vs A)	0.8752 (0.54 to 1.42)	5.91E-01	0.9368 (0.49 to 1.78)	8.43E-01	0.7958 (0.37 to 1.73)	5.63E-01
Histology (OD vs A)	0.5416 (0.35 to 0.84)	5.56E-03	0.6645 (0.36 to 1.24)	2.01E-01	0.5061 (0.27 to 0.94)	3.07E-02
Tumor Grade (3 vs 2)	2.1971 (1.44 to 3.35)	2.50E-04	1.8311 (1.07 to 3.15)	2.85E-02	3.3308 (1.50 to 7.41)	3.16E-03
NES High vs Low	1.3386 (0.93 to 1.93)	1.19E-01	1.1219 (0.68 to 1.85)	6.54E-01	1.4210 (0.80 to 2.54)	2.35E-01

Table S8I. *SHOX2* and *ATRX* comparison: The multivariate Cox proportional-hazards models analyses for LGG patients (TCGA LGG dataset).

Variable	Cases of all age (n=512, events=124)		Cases of age <48 (n=329, events=65)		Cases of age ≥48 (n=183, events=59)	
	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
SHOX2 High vs Low	3.9068 (2.50 to 6.10)	2.06E-09	3.3023 (1.65 to 6.60)	7.16E-04	4.0282 (2.04 to 7.95)	5.89E-05
Age	1.0551 (1.04 to 1.07)	1.62E-10	1.0663 (1.02 to 1.11)	3.36E-03	1.1079 (1.07 to 1.15)	1.65E-07
Gender (male vs female)	1.3780 (0.95 to 2.01)	9.50E-02	1.3078 (0.77 to 2.22)	3.22E-01	1.3034 (0.76 to 2.23)	3.35E-01
Histology (OA vs A)	0.8320 (0.51 to 1.35)	4.55E-01	0.9207 (0.49 to 1.74)	8.00E-01	0.7642 (0.35 to 1.66)	4.99E-01
Histology (OD vs A)	0.5613 (0.35 to 0.89)	1.48E-02	0.7762 (0.39 to 1.57)	4.81E-01	0.4759 (0.25 to 0.91)	2.48E-02
Tumor Grade (3 vs 2)	1.9844 (1.30 to 3.04)	1.59E-03	1.7471 (1.01 to 3.03)	4.64E-02	2.7779 (1.24 to 6.22)	1.30E-02
ATRX wt vs mut	0.9580 (0.61 to 1.51)	8.54E-01	0.7022 (0.38 to 1.30)	2.58E-01	1.7987 (0.72 to 4.47)	2.06E-01

Notes:

1. The cutoff value (48 years) was determined by mclust clustering analysis of age data in TCGA LGG dataset (see Section of Statistical analysis).
2. Hazard ratio for age variable: risk per 1 year. OA: oligoastrocytoma; A: astrocytoma; OD: oligodendroglioma.
3. The RNA-Seq expression data (cutoff value of 4.135 for *SHOX2* expression, cutoff value of 13.35 for *NES* expression) from TCGA LGG dataset were analyzed. See Appendix Tables S1for lists of datasets and data source.