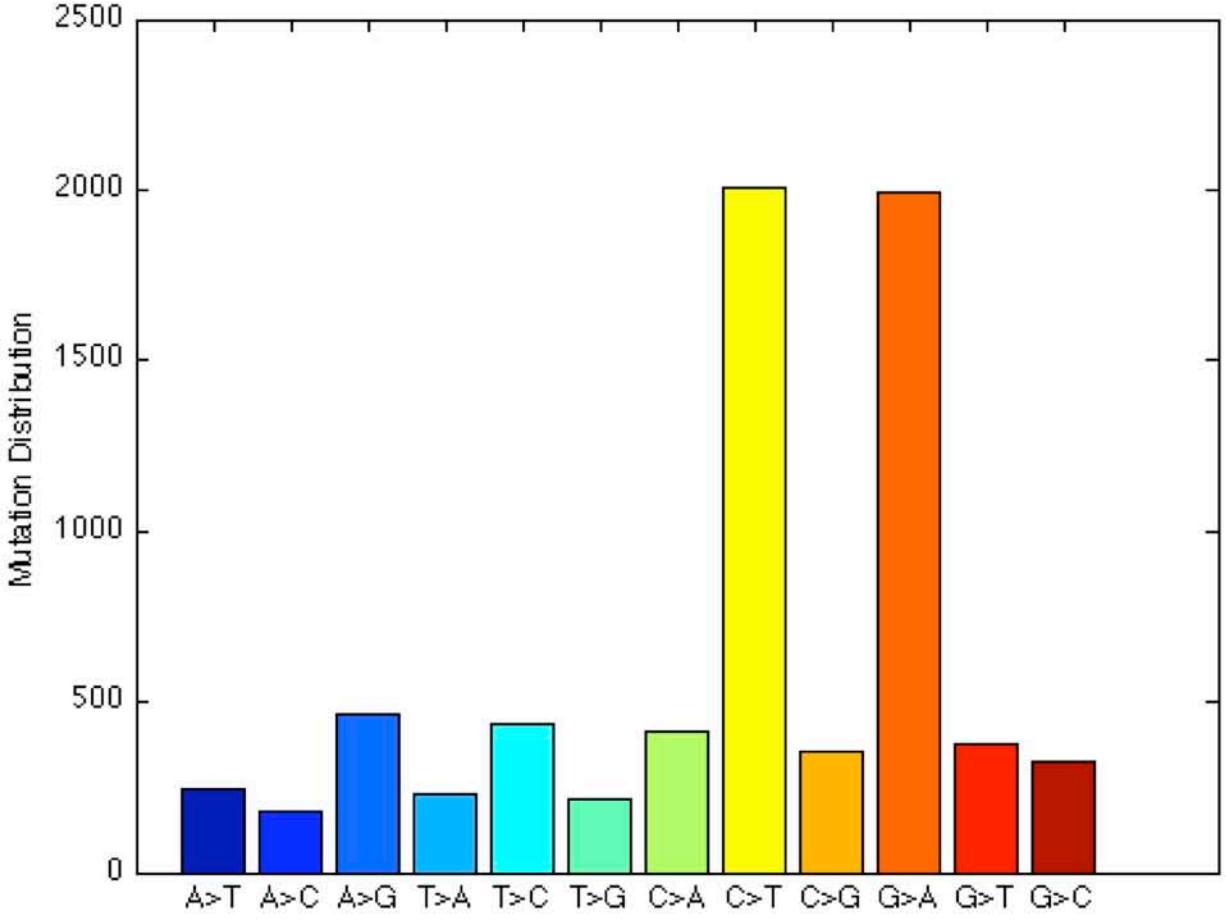
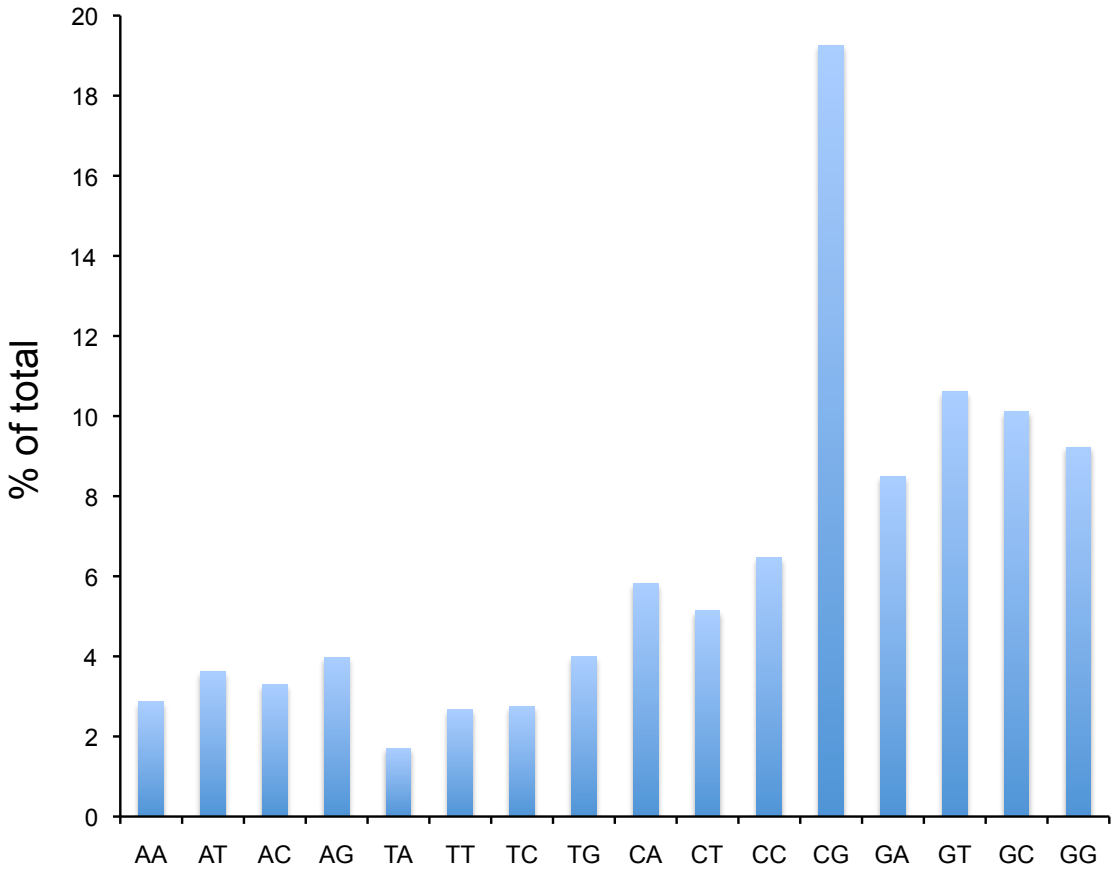
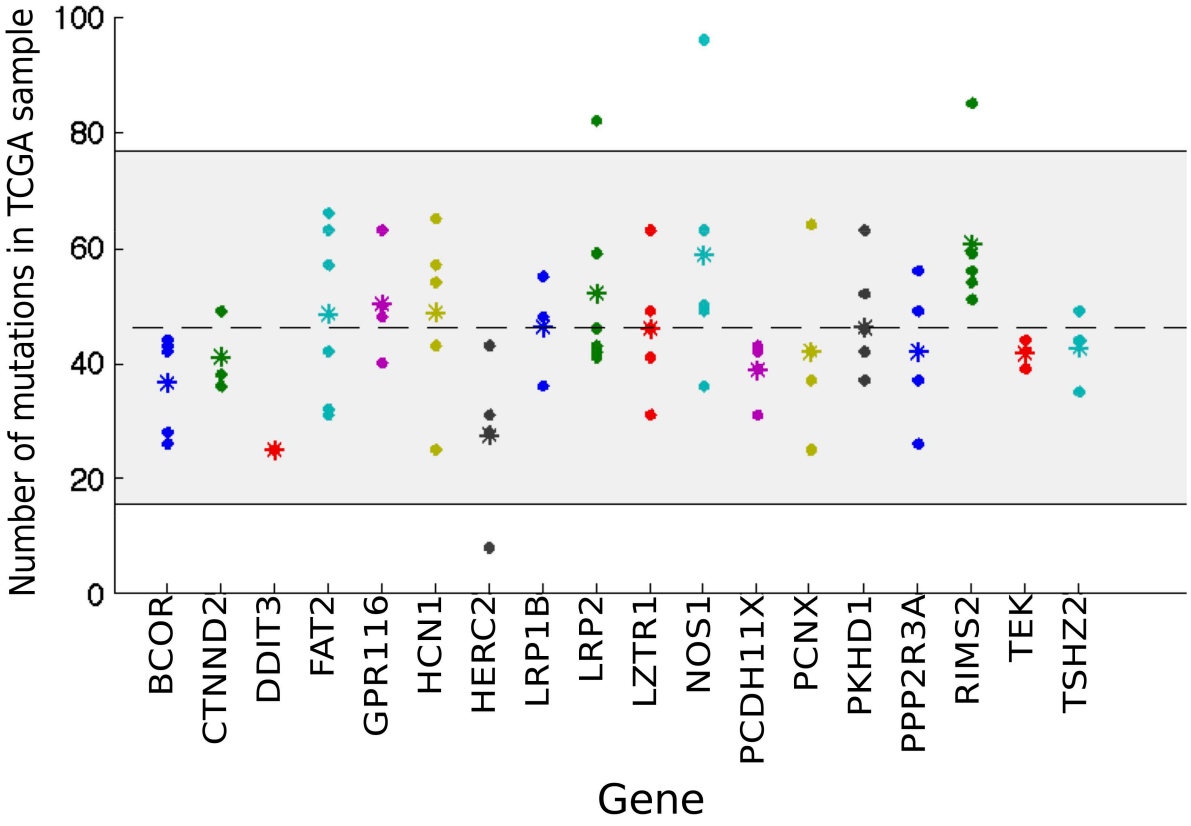


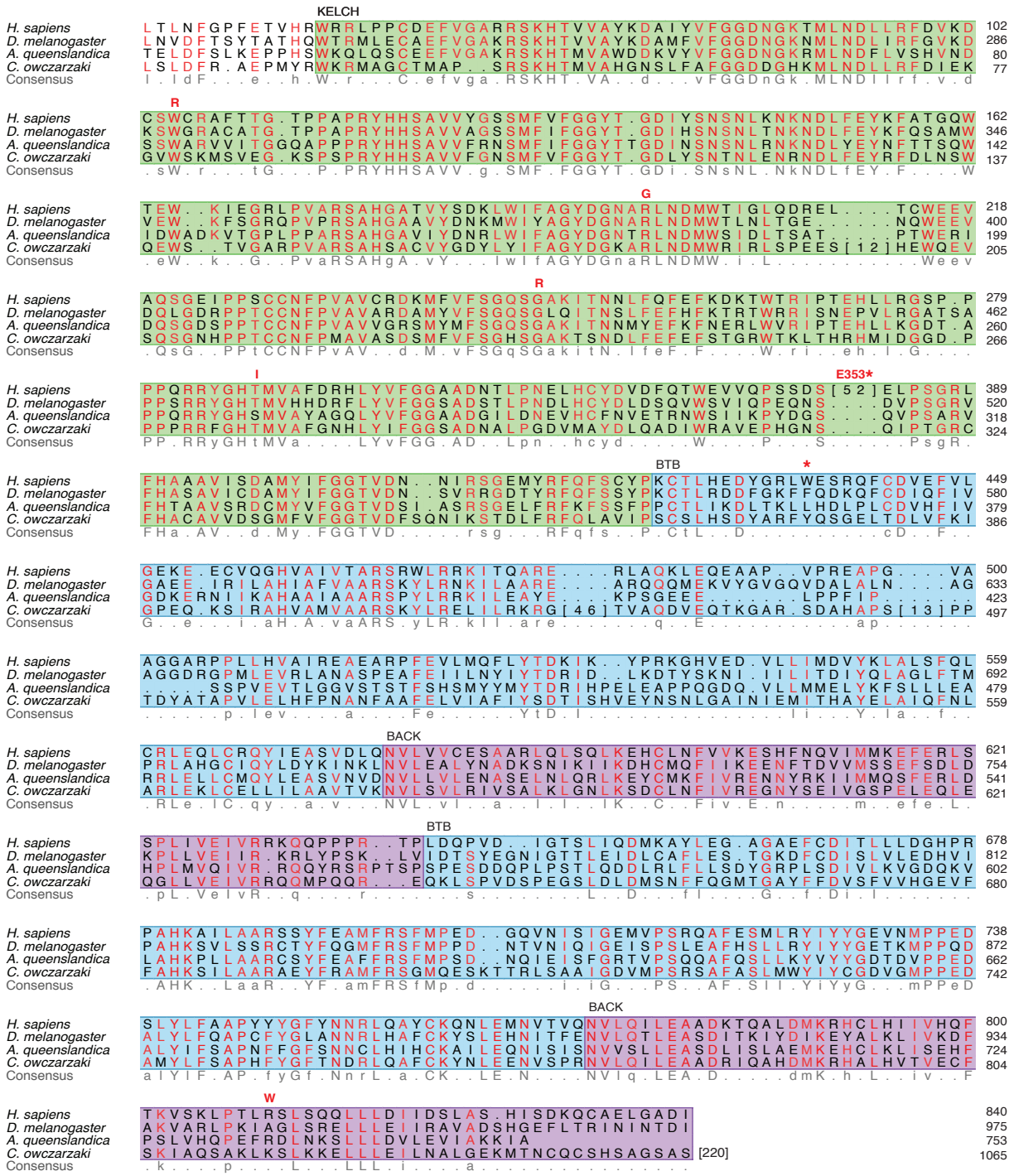
Supplementary Figure 1. Distribution of substitutions from whole exome data.



Supplementary Figure 2. Dinucleotide distribution in mutated sites.

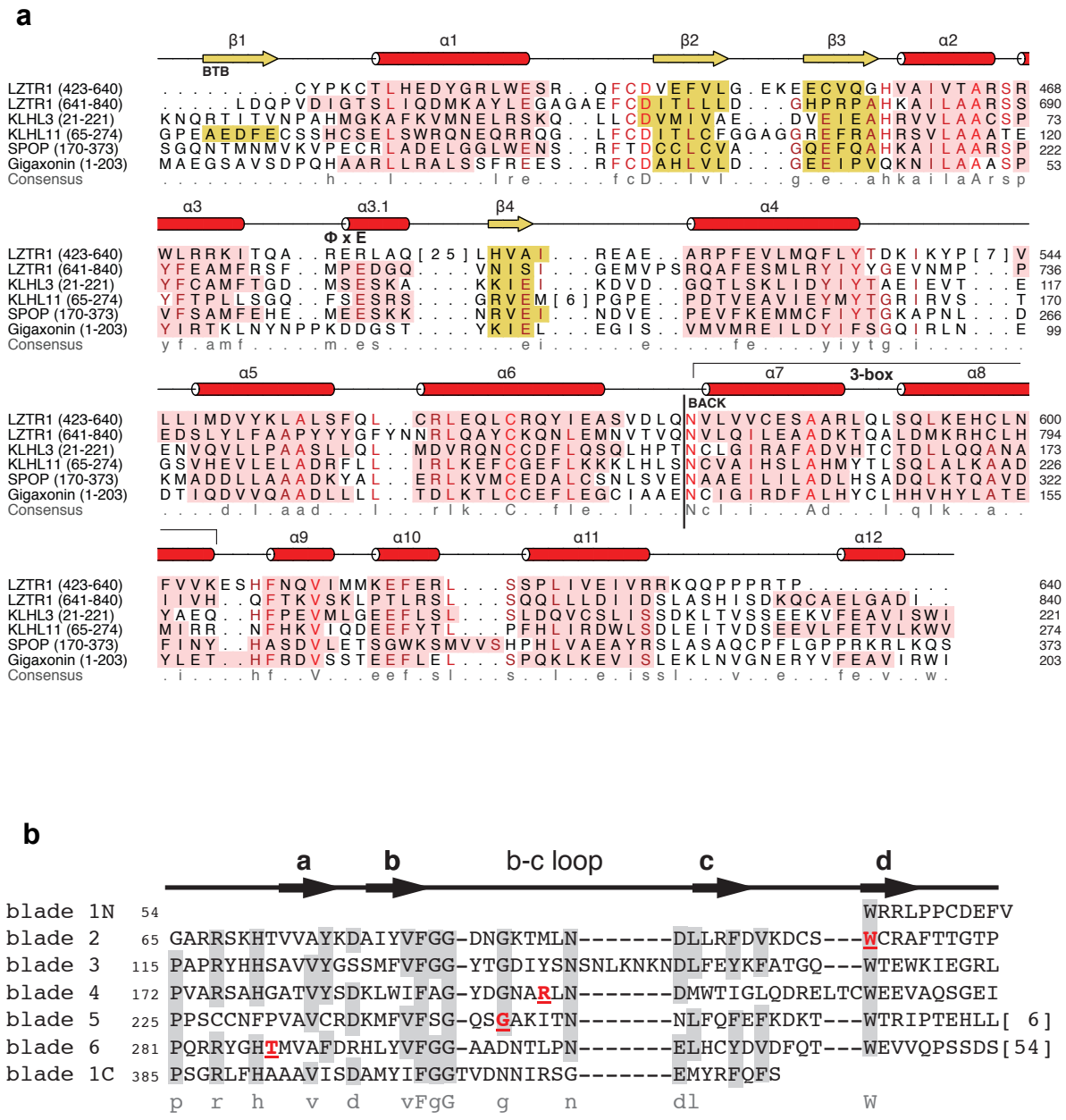


Supplementary Figure 3. Number of mutations in TCGA samples harboring MutComFocal gene candidates. For a given gene G, we plot the number of mutations M_G in samples harboring G as solid circles. We also plot the mean of M_G as asterisks. Given the mean μ and standard deviation σ of the number of mutations in all TCGA samples, we plot the 95% confidence interval of a sample being hyper-mutated (μ±1.96*σ) and show that for all G, the mean of M_G falls well within the 95% confidence interval, demonstrating that MutComFocal genes do not tend to occur in hypermutated samples.



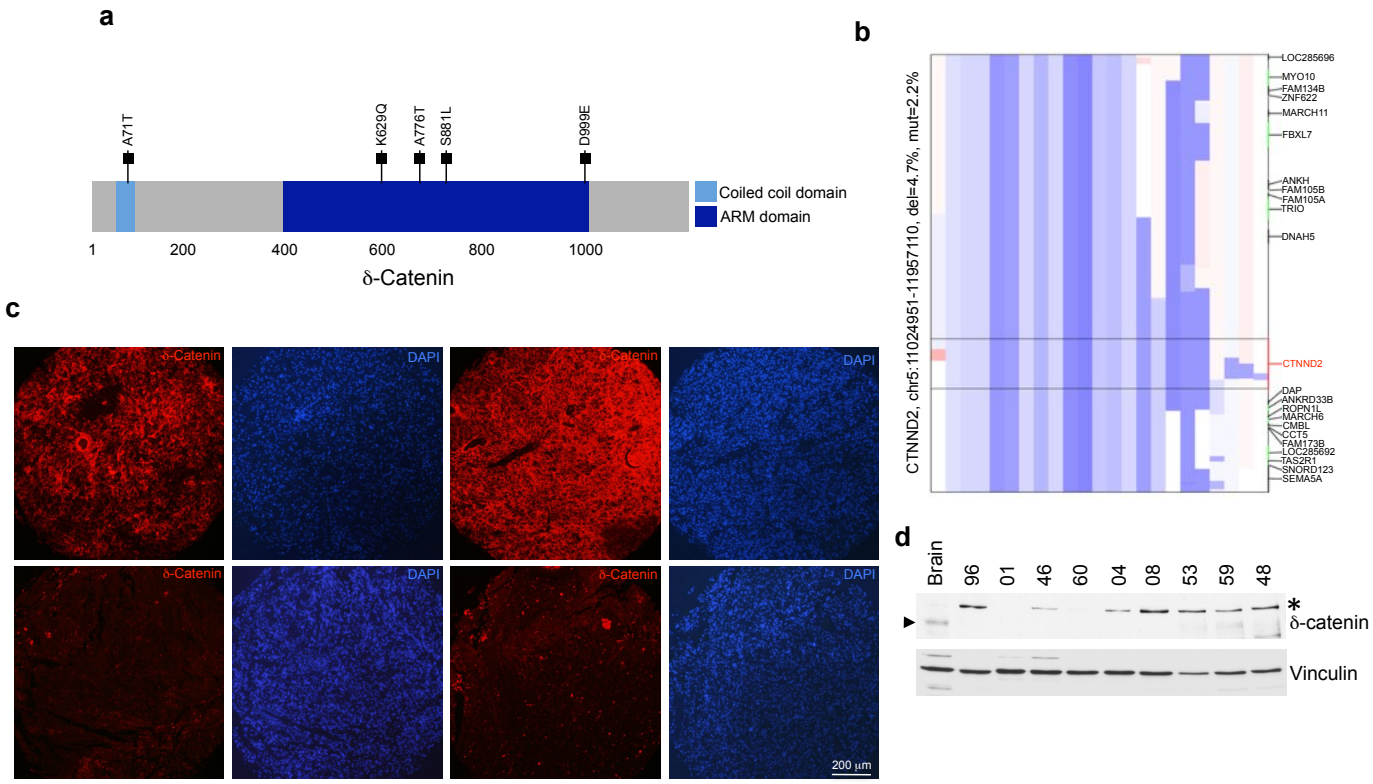
Supplementary Figure 4. Sequence alignment of selected LZTR-1 orthologs. Mutations detected in GBM are indicated in red above of the aligned sequences. The LZTR-1 gene is present in most metazoans, including the sponge *Amphimedon queenslandica*, which is generally recognized as the most ancient surviving metazoan lineage[1]. LZTR-1 is also present in some near-metazoan unicellular protists, including *Capsaspora owczarzaki.* LZTR-1 has a characteristic Kelch-BTB-BACK-BTB-BACK

domain architecture, and unlike the BTB-BACK-Kelch proteins[2], there has been little, if any, duplication of the LZTR-1 gene since its appearance. Despite its name, LZTR-1 does not contain a leucine zipper region.

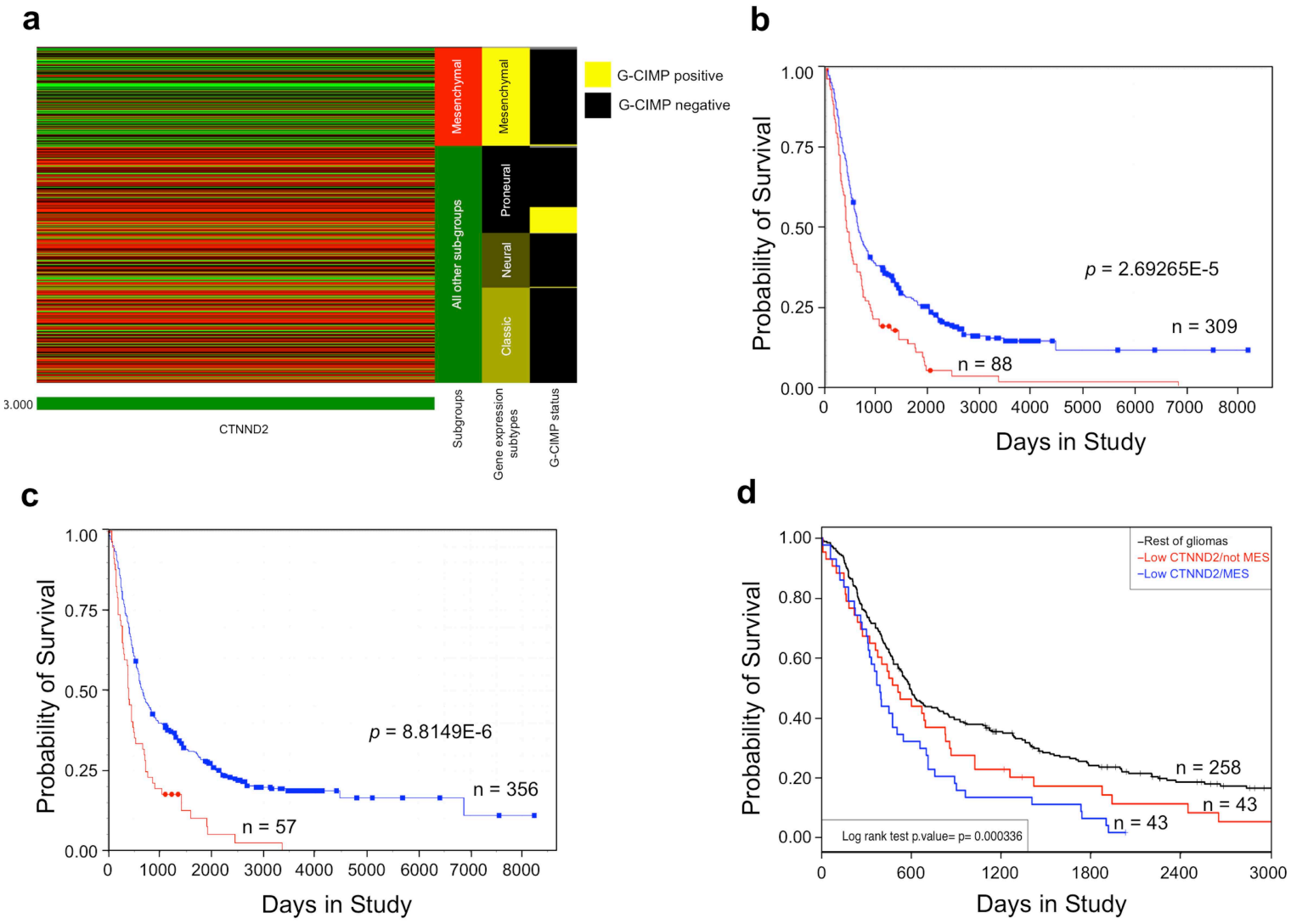


Supplementary Figure 5. Sequence alignment of BTB-BACK domains and the Kelch domain. a, The two BTB-BACK domains of LZTR-1 are included along with the predicted secondary structure from HHpred[3]. The 3-box is the Cul3 binding element within the BACK domain. The secondary structure of KLHL3 (PDB ID 4HXI), KLHL11 (PDB ID 4AP2) and Gigaxonin (PDB ID 3HVE) are indicated with shading and are based on the crystal structures. The secondary structure of SPOP is based on a crystal structure for the BTB and 3-box region (PDB ID 3HTM) and HHpred predictions from the remainder of the BACK domain. Only the N-terminal half of the BACK domain from KLHL3, KLHL11 and Gigaxonin is included, as SPOP and LZTR-1 contain truncated versions of the BACK domain. b, Sequence alignment of the six blades from the Kelch

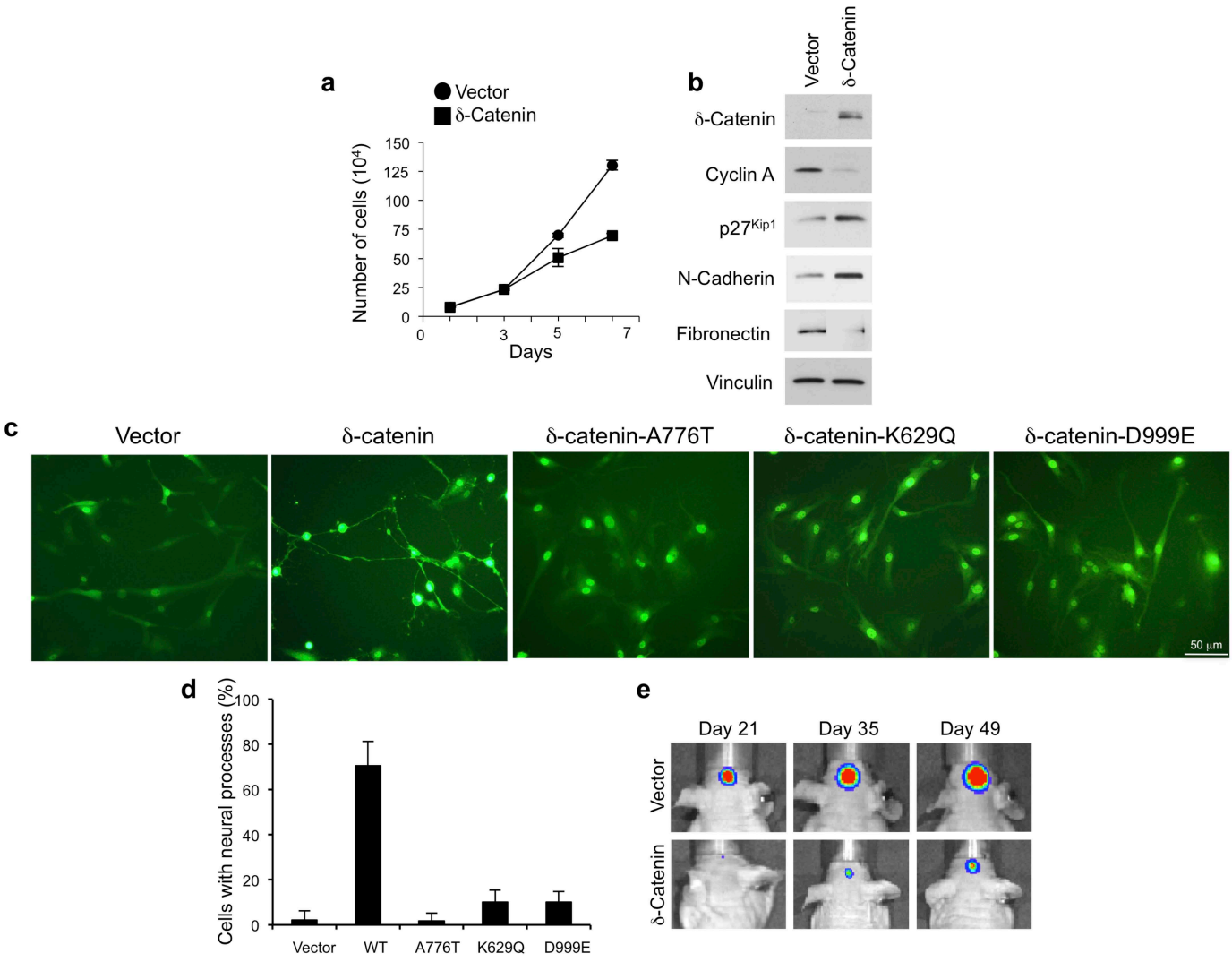
β-propeller domain. Each blade contains four core β-strands, labeled a, b, c, d. Conserved residues are highlighted in gray and residues mutated in GBM are shown in red. Insertions at the end of blades 5 and 6 are indicated in brackets. The nonsense mutation E353STOP is located between Kelch blades 6 and 1c.



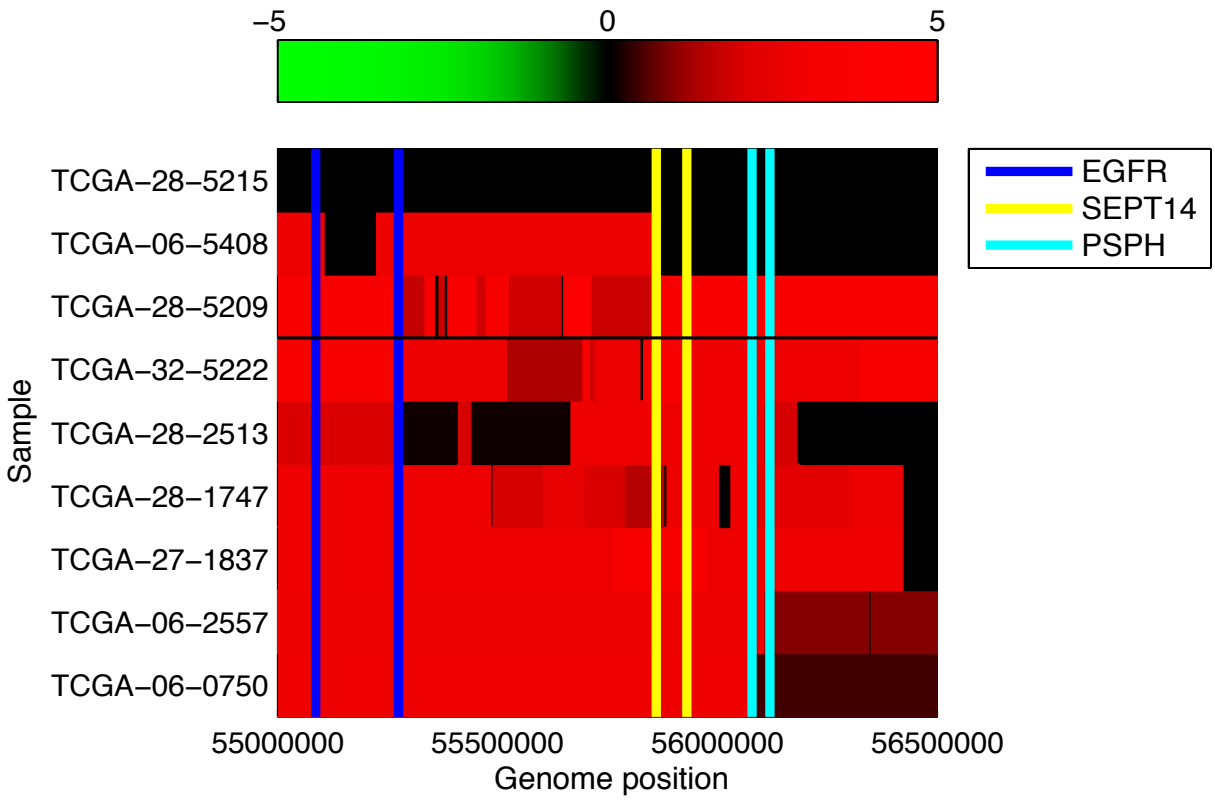
Supplementary Figure 6. Pattern of somatic mutations, CNVs and expression of CTNND2 in GBM. a, Schematic representation of identified somatic mutations in CTNND2 shown in the context of the known domain structure of the protein. Numbers refer to amino acid residues of the δ-catenin protein. b, Somatic deletions of CTNND2. Samples are sorted according to the focality of *CTNND2* deletion. In the red-blue scale, white corresponds to normal (diploid) copy number, blue is deletion and red is gain. c, Immunofluorescence staining of human primary GBM included in tissue microarrays (TMA) using δ-catenin antibody (red); Nuclei are counterstained with Dapi (blue). Two representative δ-catenin-positive and two δ-catenin-negative tumors are shown in the upper and lower panels, respectively. d, Western Blot analysis of the expression of δ-catenin in a panel of GBM-derived glioma sphere cultures. Brain, normal human brain. Arrowhead indicated δ-catenin; Asterisk, non-specific band. Vinculin is shown as control for loading.



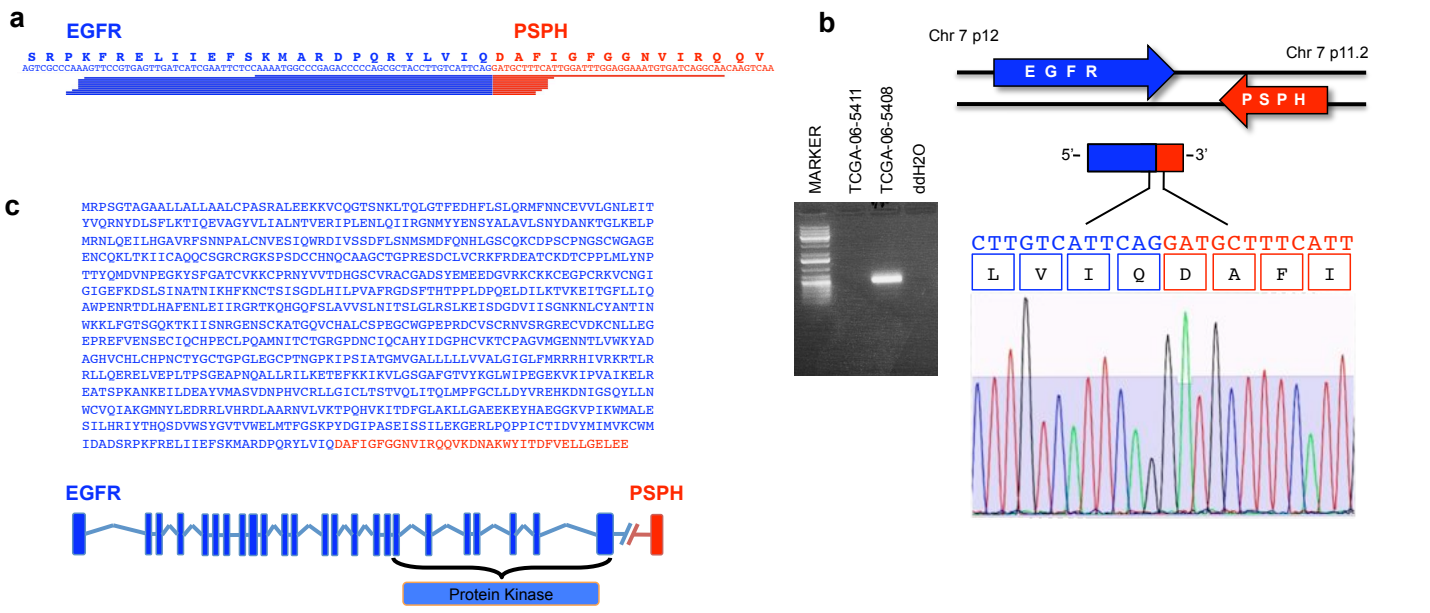
Supplementary Figure 7. a, *CTNND2* mRNA expression analysis from Atlas-TCGA samples shows that *CTNND2* is significantly down-regulated in the mesenchymal subgroup. In the green-red scale, black is the median, green is down-regulation and red is up-regulation. b, Kaplan–Meier analysis for glioma patients with low *CTNND2* mRNA expression (≤ 2-fold, red line) compared with the rest of glioma (blue line). c, Kaplan–Meier analysis for glioma patients with low *CTNND2* mRNA expression (≤2-fold) and decreased *CTNND2* gene copy number (≤1) (red line) compared with the rest of glioma (blue line). d, Kaplan–Meier analysis for patients with mesenchymal (MES, blue line), non mesenchymal (non MES, red line) GBM with low *CTNND2* mRNA expression (≤2-fold) compared with the rest of gliomas (black line).



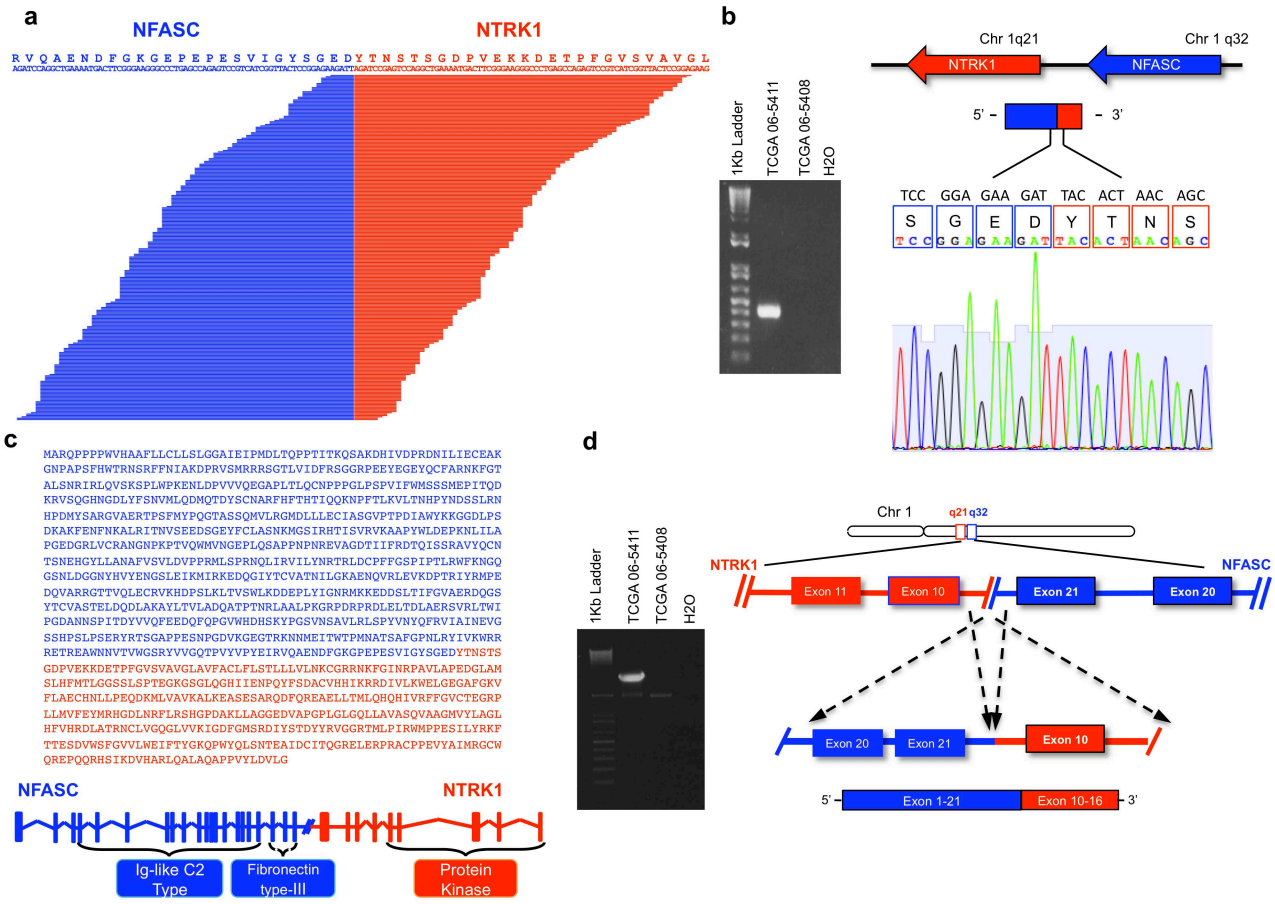
Supplementary Figure 8. Effects of expression of δ-catenin in glioma cells. a, Growth rate of U87 glioma cells transduced with a lentivirus expressing δ-catenin (squares) or the empty vector (circles, average of triplicate cultures). b, Western blot using the indicated antibodies in glioma cells expressing δ-catenin or the empty vector. Vinculin is shown as control for loading. c, U87 glioma cells transduced with a lentivirus expressing wild type δ-catenin, δ-catenin GBM-derived mutants or the empty vector were analyzed by fluorescence microscopy. d, The number of cells displaying neural processes was scored. At least 200 cells/sample were analyzed. e, Photographs of longitudinal bioluminescence imaging for one representative mouse injected intracranially with glioma sphere cells #48 transduced with lentivirus expressing *CTNND2* (lower panels) or the empty vector (upper panels).



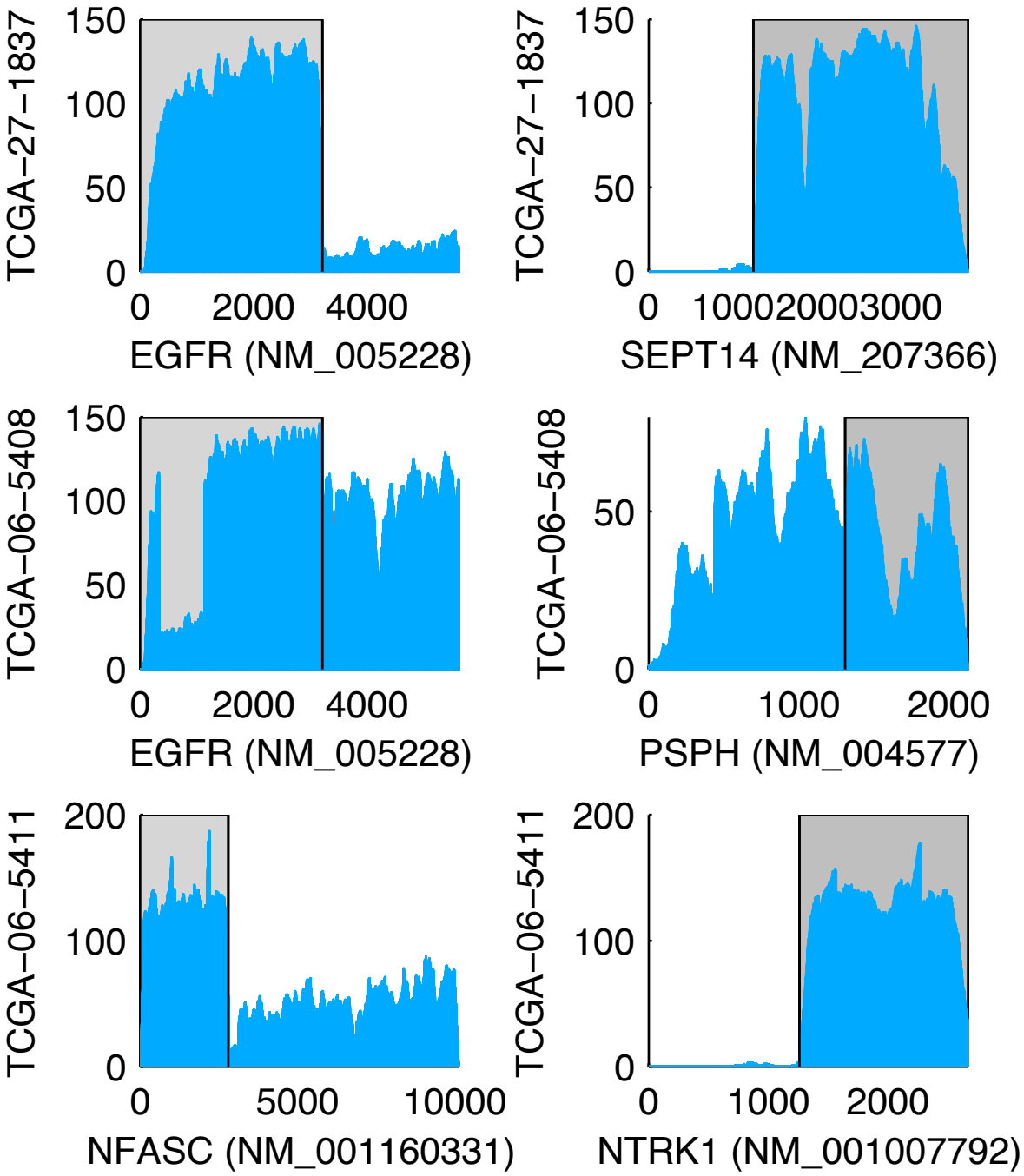
Supplementary Figure 9. Amplification surrounding the genomic neighborhood of EGFR, SEPT14, and PSPH among samples harboring EGFR fusions. We plot copy number log2 ratio across the genomic region of chr7:55000000-56500000 for samples with EGFR-PSPH (top three rows) and EGFR-SEPT14 (bottom six rows). Genomic coordinates are also plotted for EGFR (blue), SEPT14 (yellow), and PSPH (cyan).



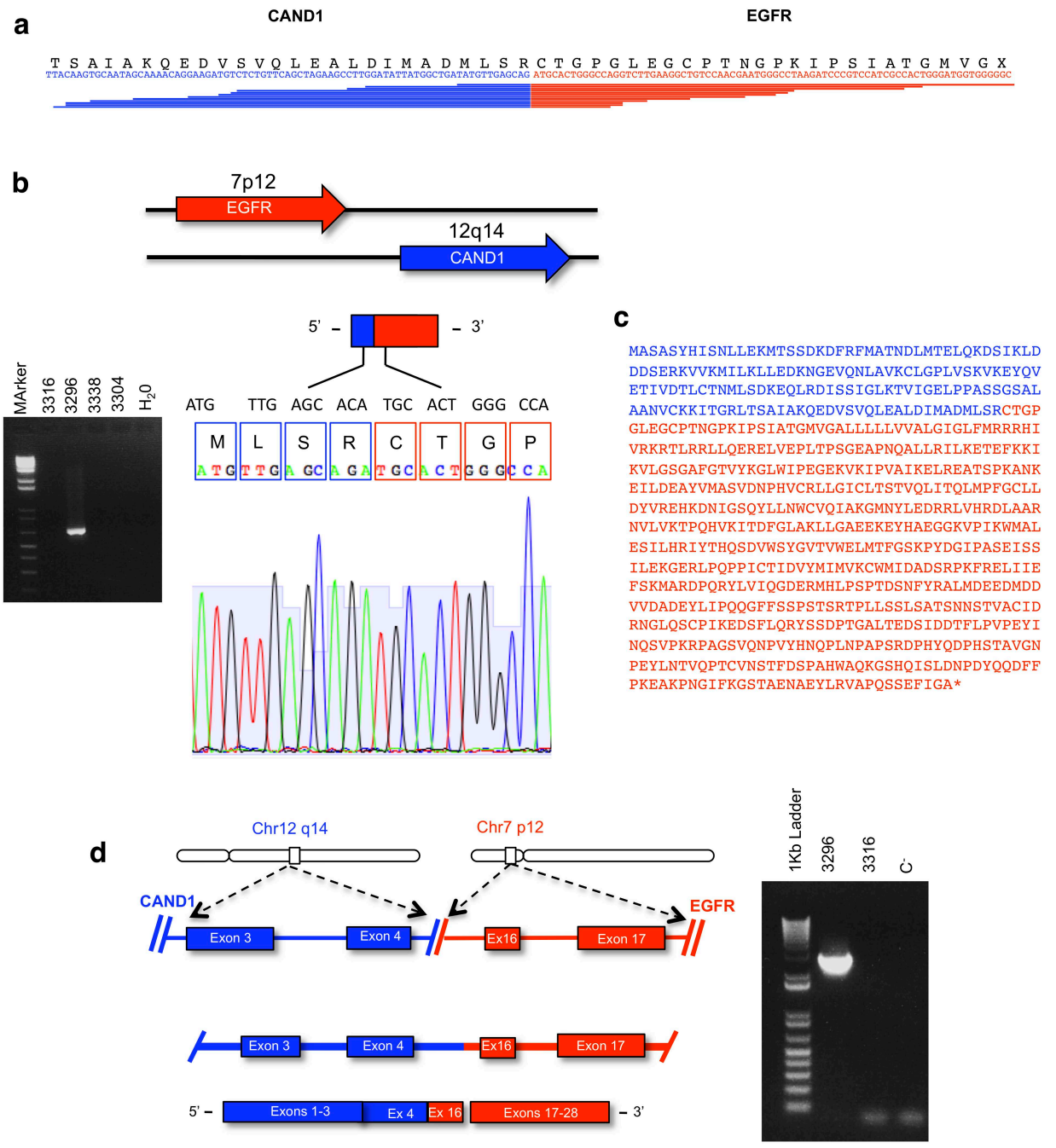
Supplementary Figure 10. *EGFR-PSPH* gene fusion identified by whole transcriptome sequencing. a, Split reads are shown aligning on the breakpoint. The predicted reading frame at the breakpoint is shown at the top with EGFR sequences in blue and PSPH in red. b, (left panel), *EGFR-PSPH* specific PCR from cDNA derived from GBMs. Marker, 1kb ladder. (right panel), Sanger sequencing chromatogram showing the reading frame at the breakpoint and putative translation of the fusion protein in the positive sample. c, EGFR-PSPH fusion protein sequence and schematics. Regions corresponding to EGFR and PSPH are shown in blue and red, respectively. The fusion includes the tyrosine kinase domain of EGFR and the last 35 amino acids of PSPH.



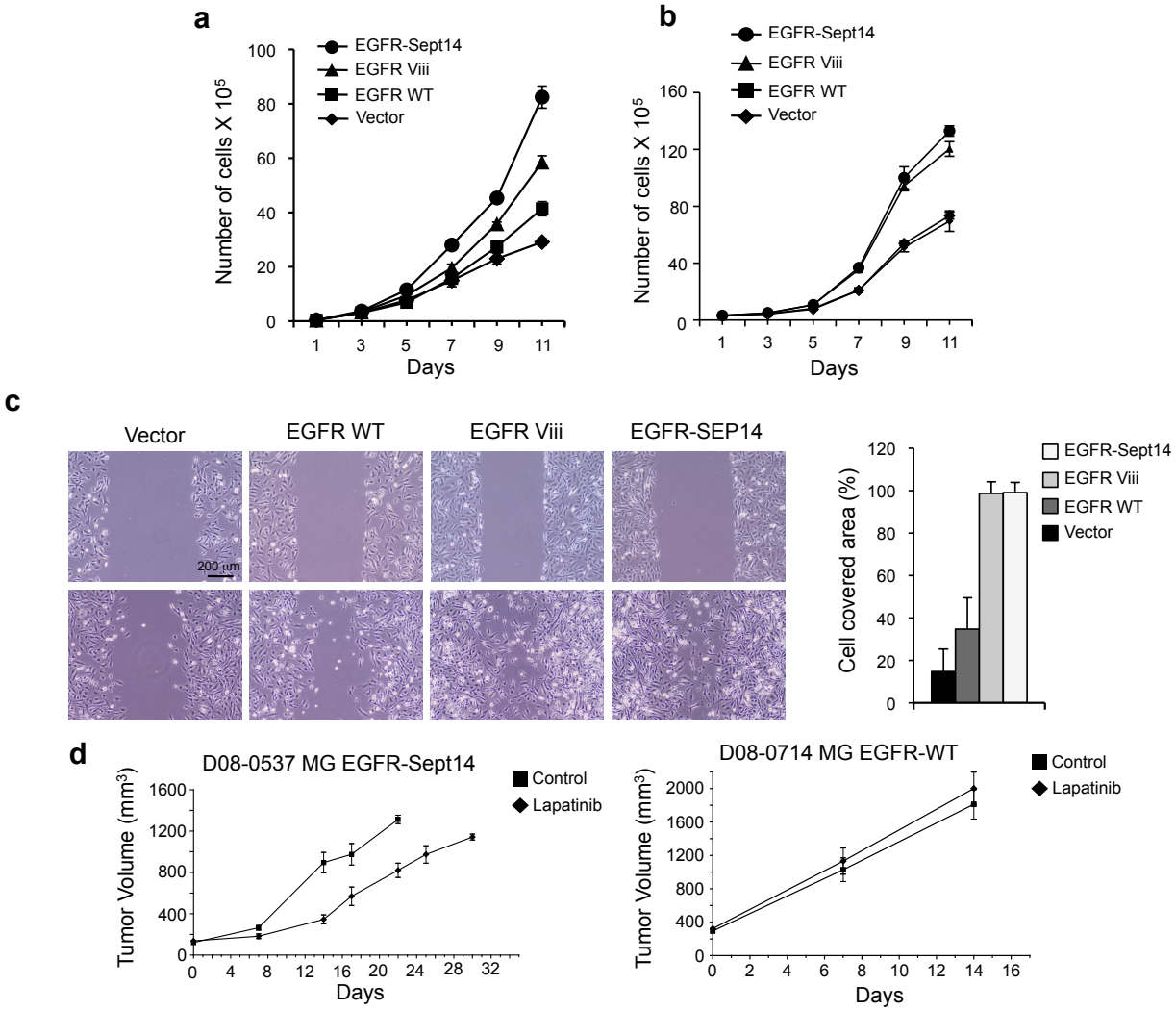
Supplementary Figure 11. *NFASC-NTRK1* gene fusion identified by whole transcriptome sequencing. a, Split reads are shown aligning on the breakpoint. The predicted reading frame at the breakpoint is shown at the top with NFASC sequences in blue and NTRK1 in red. b, (left panel), *NFASC-NTRK1* specific PCR from cDNA derived from GBMs. Marker, 1kb ladder. (right panel), Sanger sequencing chromatogram showing the reading frame at the breakpoint and putative translation of the fusion protein in the positive sample. c, NFASC-NTRK1 fusion protein sequence and schematics. Regions corresponding to NFASC and NTRK1 are shown in blue and red, respectively. The fusion includes two of the five fibronectin-type III domain of neurofascin and the protein kinase domain of NTRK1. d, Genomic fusion of *NFASC* intron 9 with intron 21 of *NTRK1*. In the fused mRNA exon 21 of *NFASC* is spliced 5' to exon 10 of *NTRK1*.



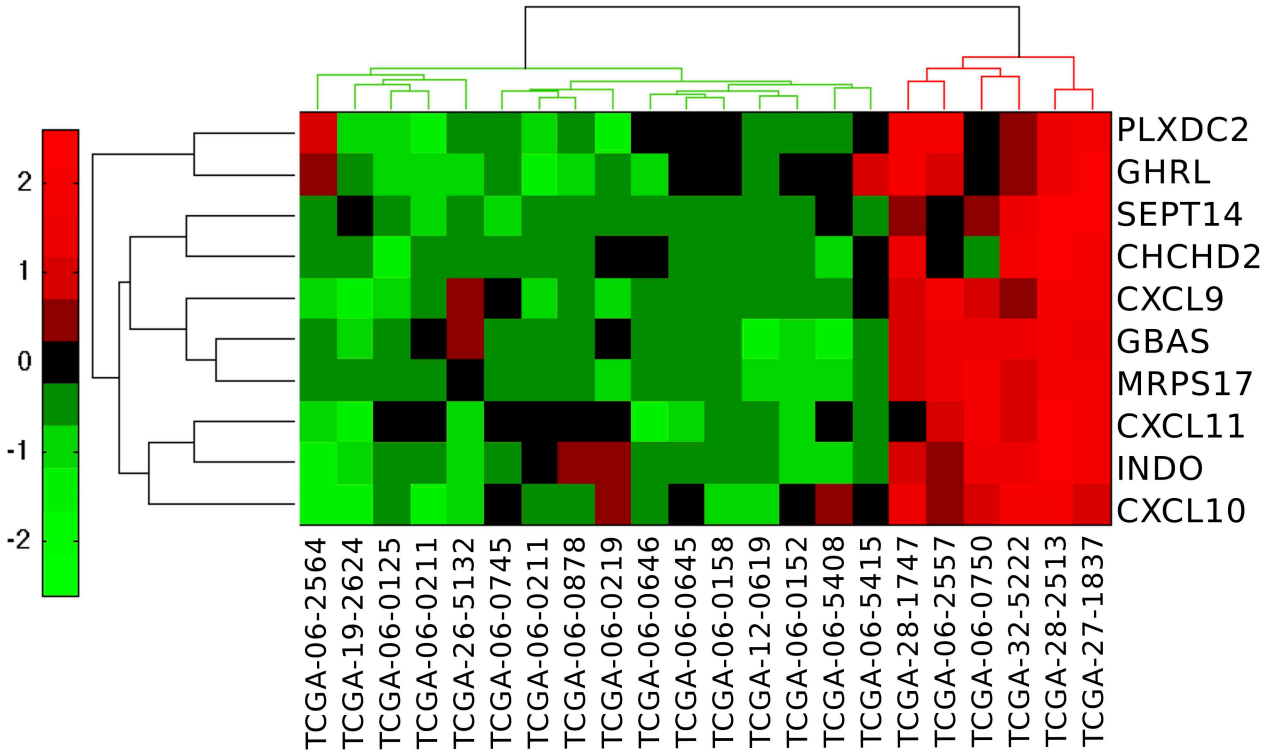
Supplementary Figure 12. Expression of thee fusion genes measured by read depth from RNA-seq data. Note the very high level of expression in the regions of the genes implicated in the fusion events.



Supplementary Figure 13. CAND1-*EGFR* gene fusion identified by whole transcriptome sequencing. a, Split reads are shown aligning on the breakpoint. The predicted reading frame at the breakpoint is shown at the top with CAND1 sequences in blue and EGFR in red. b, (left panel), *CAND1-EGFR* specific PCR from cDNA derived from GBMs. Marker, 1kb ladder; (right panel), Sanger sequencing chromatogram showing the reading frame at the breakpoint and putative translation of the fusion protein in the positive sample. c, CAND1-EGFR fusion protein sequence. Regions corresponding to CAND1 and EGFR are shown in blue and red, respectively. d, Genomic fusion of *CAND1* intron 4 with intron 15 of *EGFR*. In the fused mRNA exon 4 of *CAND1* is spliced 5' to exon 16 of *EGFR*.



Supplementary Figure 14. Expression of *EGFR-SEPT14* fusion promotes an aggressive phenotype and inhibition of EGFR kinase delays GBM growth *in vivo*. a, Growth rate of U87 glioma cells transduced with a lentivirus expressing EGFR-SEPT14, EGFR Viii, EGFR WT or the empty vector (average of triplicate cultures). b, Growth rate of SNB19 glioma cells transduced with a lentivirus expressing EGFR-SEPT14, EGFR Viii, EGFR WT or the empty vector (average of triplicate cultures). c, (Left panels) Migration assay in SNB19 glioma cells transduced with a lentivirus expressing EGFR-SEPT14, EGFR Viii, EGFR WT or the empty vector. (Right panels) Quantification of the cell covered area (average of triplicate cultures). Error bars are SD. d, Kinetics of tumor growth for glioma patient-derived xenografts carrying the EGFR-SEPT14 fusion (D08-0537 MG, left panel) or wild type EGFR (D08-0714, right panel) treated with Lapatinib or vehicle (control). Error bars are SD.



Supplementary Figure 15. Differential expression of GBM tumor samples harboring EGFR-SEPT14 fusions and EGFRvIII rearrangements. After filtering for statistical significance for differential expression, ten genes remained that characterized the EGFR-SEPT14 phenotype from the EGFRvIII phenotype. Log2 expression was plotted as a heat map. Samples were hierarchically clustered by Euclidean distance using average linkage. This clustering demonstrates clear separation between EGFR-SEPT14 samples (red) and EGFRvIII samples (green), confirming the unique molecular signature of the EGFR-SEPT14 gene fusion.

Supplementary Table 1. Somatic point mutations from 139 whole exome data from TCGA, as predicted by SAVI. This item is presented separately as an Excel file. Annotation information in each column is described below:

case	TCGA id
tf	tumor frequency
tf_lower	tumor frequency lower bound
tf_upper	tumor frequency upper bound
nf	normal frequency
nf_lower	normal frequency lower bound
nf_upper	normal frequency upper bound
tf-nf	frequency difference
tf-nf_lower	frequency difference lower bound
tf-nf_upper	frequency difference upper bound
t_qual	quality of variants nucleotides in tumor
t_vardepth	tumor variant depth
t_totdepth	tumor total depth
n_qual	quality of variants nucleotides in normal
n_vardepth	normal variant depth
n_totdepth	normal total depth
chr:pos	position
ref/var	reference/variant allele
gene	gene
CCDS	CCDS
sense	sense
exon_#	exon number
codon	condon
AA	amino acid
SNP	SNP
ref	reference genome
cntxt	context

Supplementary Table 2. List of mutations predicted by SAVI in the TCGA dataset and confirmed as somatic events (present in tumor DNA and absent in matched normal DNA) by Sanger sequencing. The mutations successfully validated by Sanger sequencing are highlighted in green, whereas those not validated by Sanger sequencing are highlighted in red.

SAMPLES	GENE	NTS VAR	aa VAR	
Non Synonymous Mutations				
TCGA-06-0166	BCOR	G/A	R810*	✓
TCGA-06-0140	BCOR	T/A	Y805F	✓
TCGA-06-0171	BCOR	A/C	V1667G,V1649G,V1701G	×
TCGA-06-2559	BCOR	C/T	W1184*	✓
TCGA-06-0188	BCORL1	G/T	E1685*	✓
TCGA-06-0188	LZTR1	C/T	R810W	✓
TCGA-06-2562	LZTR1	T/C	W105R	✓
TCGA-06-5413	LZTR1	G/A	G248R	✓
TCGA-06-2559	LRP2	A/G	L4130P	✓
TCGA-06-0188	LRP2	G/A	R4165C	✓
TCGA-06-0171	TRIML2	C/A	R265M	✓
TCGA-27-1835	GPR116	C/T	A1176T	✓
TCGA-27-1835	PTPRT	G/A	R364*	✓
TCGA-06-0140	PTPRT	C/T	V424M	✓
TCGA-06-2559	PCDH11X	G/A	R518H	✓
TCGA-06-0171	PKHD1	G/A	L1989F	✓
TCGA-06-5413	FAT2	G/A	R411*	✓
TCGA-27-1835	GABRA6	C/A	T147N	✓
TCGA-06-2562	CTNND2	C/T	A71T	✓
TCGA-06-2559	HERC2	A/C	S895A	×
TCGA-06-2562	PALM2-AKAP2	C/T	R150*	✓
TCGA-06-0171	HLTF	G/C	S29*	✓
TCGA-06-2559	MAPK1	G/A	L244F	✓
TCGA-06-0188	TACR3	C/T	V403M	✓
TCGA-27-1835	SYT9	C/T	R215W	✓
TCGA-06-0188	PTPRB	T/G	R592S-R374S	✓
TCGA-06-2559	AOX1	C/T	A507V	✓
TCGA-27-1835	MDN1	C/-	T5272-	✓
TCGA-06-0188	TMPRSS11A	G/-	Y318-,Y315-	✓
TCGA-27-1835	IFT172	C/A	R1483S	✓
TCGA-06-2562	ITGAX	G/A	R685H	✓
TCGA-06-5413	AKAP1	T/A	S690T	✓
TCGA-06-0171	CAMK2A	G/A	A384A-A395A	✓
TCGA-27-1835	KIAA1804	C/T	R760*	✓

TCGA-06-5413	HDLBP	T/C	I127V	✓
TCGA-06-2562	PIGR	C/T	V558I	✓
TCGA-06-2562	GAK	C/T	V709V	✓
TCGA-06-2559	SCNN1G	C/A	F531L	✓
Synonymous Mutations				
TCGA-06-0188	AKNA	C/T	R551R	✓
TCGA-27-1835	MAGEC2	C/T	L61L	✓
TCGA-06-2562	RNF43	T/G	R600S	✓

Supplementary Table 3. List of genes mutated in GBM ranked by enrichment of non-synonymous mutations over synonymous. This item is presented separately as an Excel file.

Supplementary Table 4. List of 67 genes prioritized by MutComFocal. In column C the gene names are highlighted in green (previously known GBM genes), red (new GBM genes, somatic mutations validated in independent GBM dataset) or blue (new GBM genes, somatic mutations not validated in independent GBM dataset).
Annotation information in each column is described below:

region	region index
type	type of the alteration
gene	gene
chr	chromsome
start	start of gene
end	end of gene
amp_mut_rank	amp-mut rank
amp_mut_tier	amp-mut tier
amp_freq	amplification frequency
amp_min1_genes	number of genes in smallest amplification
del_mut_rank	Del-Mut rank
del_mut_tier	Del-Mut tier
del_freq	deletion frequency
del_min1_genes	number of genes in smallest deletion
mut_rank	mut rank
mut_freq	mutation frequency
has_stop	does it have a nonsene mutation

region	type	gene	chr		start	end	amp_mut_ra	amp_mut_tie	amp_freq	amp_min1_g	del_mut_ran	del_mut_tier	del_freq	del_min1_ge	mut_rank	mut_freq	has_stop
225	del_mut	ADAMTS16		5	5193442	5373412	802	7	4.7	70	166	3	6	2	316	2.9	0
768	del_mut	ADAMTS18		16	75873525	76026512	903	7	2.8	6	190	3	11.5	6	725	2.2	0
149	del_mut	ADCY5		3	124486088	124650082	1022	7	4.1	7	167	3	6.8	1	441	2.2	0
106	mut_1	AOX1		2	201158975	201244462	1418	8	2.6	53	1880	8	0.9	18	140	2.9	0
966	del_mut	ATRX	X		76647011	76928375	379	6	7.2	35	112	3	14.9	5	45	2.9	1
936	del_mut	BCOR	X		39795442	39921526	681	7	7	165	50	3	14.9	1	454	3.7	2
427	amp_mut	BRAF		7	140080281	140271033	27	3	74.4	1	2104	9	0.9	5	162	0.7	0
59	mut_1	CACNA1S		1	199275262	199348317	158	5	9.2	4	1450	7	1.3	22	143	2.9	0
57	amp_mut	CFHR4		1	195123834	195154386	49	3	9.2	1	4396	26	0	0	19	2.2	0
457	del_mut	COL22A1		8	139669659	139995418	301	6	6	3	158	3	4.7	1	137	2.2	0
117	mut_1	COL6A3		2	237897393	237987589	1961	10	1.9	86	1094	6	3	6	1037	3.7	1
229	del_mut	CTNND2		5	11024951	11957110	507	6	5.1	70	82	3	4.7	1	39	2.2	0
610	amp_mut	DDIT3		12	56196637	56200567	11	3	13.6	2	1022	6	6.4	5	293	0.7	0
53	mut_1	DDR2		1	160868851	161016871	894	7	8.5	57	1473	7	0.9	3	506	2.9	0
932	del_mut	DMD	X		31047265	33267647	609	6	7	165	7	2	17.1	1	277	2.2	0
91	del_mut	DPP10		2	114916368	116318406	473	6	2.8	2	135	3	2.3	1	12	2.2	1
272	del_mut	DSP		6	7486868	7531945	1507	9	1.1	69	162	3	9	1	26	2.2	0
389	amp_mut	EGFR		7	55054218	55242525	1	1	71	1	1788	8	1.1	1	2	15.4	0
54	mut_1	F5		1	167747815	167822393	292	5	8.5	13	1652	7	1.3	24	51	2.9	1
93	mut_1	FAM123C		2	131229546	131242177	713	7	2.8	2	1308	6	1.5	11	255	2.9	0
250	mut_1	FAT2		5	150863845	150928698	1274	8	3.2	73	838	5	4.5	5	77	4.4	1
252	mut_1	GABRA6		5	161045235	161062176	908	7	3.2	1	899	5	4.5	54	41	2.9	0
288	mut_1	GPR116		6	46928203	47030634	1997	10	1.1	256	1044	6	7.7	14	392	2.9	0
237	del_mut	HCN1		5	45295108	45731977	892	7	4.5	38	214	3	3.2	1	119	3.7	1
672	del_mut	HERC2		15	26029782	26240890	1999	10	0.6	24	109	3	15.1	1	234	2.2	0
108	mut_1	IDH1		2	208809197	208828051	279	5	2.8	6	874	5	0.9	179	9	7.4	0
188	amp_mut	KIT		4	55218851	55301638	13	3	11.7	2	1929	8	3	5	2206	0.7	0
47	mut_1	KPRP		1	150997129	151001153	267	5	9.2	10	1047	6	0.6	10	1172	2.9	0
916	mut_1	KRTAP10-11		21	44890758	44891994	20563	27	6.2	67	20562	26	4.5	40	20569	2.9	0
636	del_mut	LRFN5		14	41146513	41443504	406	6	1.1	1	15	2	24.7	1	124	2.9	0
97	del_mut	LRP1B		2	140705465	142605740	1582	9	2.3	15	19	2	6	1	421	2.2	1
101	mut_1	LRP2		2	169691864	169927368	1144	8	3	133	1718	7	1.1	139	47	4.4	0
918	del_mut	LZTR1		22	19666557	19683326	1419	8	1.7	83	86	3	22.4	4	30	4.4	1
60	amp_mut	MDM4		1	202752133	202793870	4	2	17.3	5	2001	8	0.6	39	192	0.7	0
923	del_mut	MXRA5	X		3236607	3274684	290	5	6.6	6	65	3	15.1	1	220	2.9	0
84	amp_mut	MYCN		2	15998133	16004580	25	3	4.9	2	1773	8	2.1	60	905	0.7	0
808	del_mut	NF1		17	26446070	26728820	2063	10	1.5	43	84	3	8.3	1	1536	2.2	1
625	mut_1	NOS1		12	116135361	116283965	725	7	4.3	12	1004	6	4.9	18	36	3.7	0
982	del_mut	ODZ1	X		123337436	123925347	755	7	6.8	10	75	3	14.5	1	650	2.2	0
966	del_mut	PCDH11X	X		90920915	91764878	177	5	4.9	1	1	1	35	1	50	2.9	0
648	del_mut	PCNX		14	70443874	70651852	2298	12	1.1	28	228	4	24.5	1	952	2.2	1
188	amp_mut	PDGFRA		4	54790020	54859169	2	2	14.3	2	1430	7	2.6	19	314	2.9	0
60	amp_mut	PIK3C2B		1	202658380	202726097	8	2	17.1	5	2051	9	0.6	1	729	0.7	0
160	amp_mut	PIK3CA		3	180349004	180435191	51	3	9.6	4	737	5	1.5	2	13	5.9	1
239	del_mut	PIK3R1		5	67558217	67633405	442	6	4.3	12	67	3	3.8	1	16	5.1	0
919	del_mut	PIWIL3		22	23445000	23500683	2001	10	1.5	21	165	3	24.5	9	554	2.2	1

921	del_mut	PKDREJ		22	45030223	45037883	2362	12	1.3	97	170	3	28.6	8	596	2.9	0
291	del_mut	PKHD1		6	51588103	52060382	1465	9	1.3	256	103	3	7.7	1	34	4.4	0
105	del_mut	PLCL1		2	198377670	198722851	1691	9	2.8	53	203	3	1.5	1	722	2.2	1
152	mut_1	PPP2R3A		3	137167256	137349423	1317	8	4.7	176	1409	7	4.9	3	1270	2.9	1
747	del_mut	PRKCB		16	23754800	24139431	1741	9	2.3	219	191	3	9.2	1	146	2.2	0
504	del_mut	PTEN		10	89613174	89718512	1915	10	0.4	239	2	1	82.7	1	18	8.8	4
898	del_mut	PTPRT		20	40134805	41251971	249	5	27.7	100	226	4	2.6	1	93	2.9	1
633	del_mut	RB1		13	47775883	47954027	2540	14	0.4	111	14	2	36	1	636	3.7	4
454	mut_1	RIMS2		8	104582151	105334627	852	7	5.3	63	1437	7	3.2	1	406	4.4	3
75	amp_mut	SDCCAG8		1	241485942	241730016	30	3	10.9	3	1055	6	1.3	4	190	0.7	0
254	mut_1	SLIT3		5	168025648	168660711	1277	8	3.4	34	983	5	4.5	2	698	2.9	0
999	del_mut	SLITRK4	X		142543607	142550685	619	7	7	1092	138	3	14.7	1	329	2.2	0
353	del_mut	SNX9		6	158164281	158286097	1098	8	2.1	7	200	3	19.4	3	211	2.2	0
50	mut_1	SPTA1		1	156847119	156923130	464	6	9	30	1700	7	0.4	113	23	6.6	2
351	del_mut	SYNE1		6	152484514	153000227	1347	8	1.9	58	30	2	19	2	46	3.7	0
460	del_mut	TEK		9	27099146	27220172	304	6	2.8	1	5	2	43.5	1	86	2.2	0
920	del_mut	TMPRSS6		22	35791424	35829639	2156	11	1.3	119	163	3	26.9	4	94	2.2	1
796	del_mut	TP53		17	7512444	7531588	449	6	0.6	26	4	2	10.2	1	1	25	2
217	del_mut	TRIML1		4	189297591	189305643	330	6	1.7	3	87	3	7	1	35	2.9	0
453	mut_1	TRPA1		8	73096039	73150373	591	6	4.7	107	786	5	4.1	3	29	3.7	1
901	amp_mut_1	TSHZ2		20	51022283	51537372	246	5	27.3	8	1267	6	2.1	1	98	2.2	0

Supplementary Table 5. Somatic point mutations identified in 18 novel GBM genes from an independent dataset of 83 GBM samples and matched normal DNA. Each mutation reported in this list has been confirmed as present in the corresponding tumor DNA and absent in the matched normal DNA by Sanger sequencing.

SAMPLE	CHR	POS	REF	ALT	N DEPTH	N FREQ	T DEPTH	T FREQ	SomaticSc	MBH:Gene	MBF:Mutation Typ	MBF:Protein Mutation
D-3469-T	chr1	6186732	G	C	660	0.3%	433	33.5%	225	CHD5	Missense	p.H1326Q
D-10-0001-T	chr1	6202556	C	T	493	0.0%	413	29.5%	225	CHD5	Missense	p.G718D
CG-3572	chr1	215844367	A	T	574	0.3%	604	32.6%	225	USH2A	Missense	p.S4694T
CG-3560	chr1	216052409	C	T	254	0.4%	271	17.3%	69	USH2A	Missense	p.G2752E
CG-3609	chr1	216173785	G	A	530	0.0%	347	17.0%	117	USH2A	Missense	p.P2149S
CG-3609	chr1	216243499	C	T	546	0.5%	356	16.3%	76	USH2A	Missense	p.R1998H
CG-3553	chr1	216373115	G	A	537	0.0%	446	17.7%	157	USH2A	Missense///Misser	p.A1222V///p.A1222V
D-04-0069-T	chr1	216373156	C	A	359	0.0%	679	42.3%	225	USH2A	Missense///Misser	p.W1208C///p.W1208(
D-08-0611-T	chr10	32562198	C	A	431	0.2%	286	29.4%	225	EPC1	Nonsense	p.E586*
D-09-0634-T	chr12	117703231	C	A	1007	0.0%	350	33.7%	225	NOS1	Missense	p.G676C
CG-3589	chr12	117723064	G	A	412	0.0%	569	45.9%	225	NOS1	Missense	p.T455I
D-987-1E-T	chr14	42355932	C	T	597	0.2%	487	16.4%	110	LRFN5	Missense	p.T35I
CG-3560	chr14	71445051	C	T	205	0.5%	175	52.0%	225	PCNX	Missense	p.P666L
CG-3609	chr15	28387512	C	T	446	0.0%	207	37.7%	225	HERC2	Missense	p.E3858K
CG-3609	chr15	28443550	G	A	495	0.0%	287	29.6%	225	HERC2	Missense	p.T2662I
CG-3609	chr17	3627816	T	C	340	0.3%	115	27.8%	225	GSG2	Missense	p.V196A
D-09-0439-T	chr2	141739739	C	G	674	0.0%	496	38.9%	225	LRP1B	Missense	p.M959I
CG-3560	chr2	141773291	C	T	241	0.0%	326	22.7%	173	LRP1B	Missense	p.V722I
D-06-0427-T	chr2	170044680	C	T	507	0.0%	432	38.7%	225	LRP2	Missense	p.R3043H

CG-3609	chr2	170077027	G	T	495	0.4%	310	34.8%	225	LRP2	Missense	p.A1862D
D-987-1E-T	chr2	170148856	G	C	496	0.0%	443	15.8%	55	LRP2	Missense	p.Q226E
D-1225-T	chr2	170175279	T	A	229	0.0%	500	33.8%	225	LRP2	Missense	p.Q101H
CG-3589	chr20	40739031	C	T	398	0.3%	925	27.0%	225	PTPRT	Missense///Misser	p.V1066I///p.V1085I
D-01-0141-T	chr20	40743859	G	A	378	0.0%	629	44.0%	225	PTPRT	Missense///Misser	p.R1027C///p.R1046C
D-09-0362-T	chr22	21346566	G	T	460	0.0%	226	39.8%	225	LZTR1	Nonsense	p.E353*
CG-3560	chr22	25155856	C	T	352	0.9%	306	19.0%	134	PIWIL3	Missense	p.G68E
CG-3567	chr4	189068521	G	A	432	0.7%	354	43.8%	225	TRIML1	Missense	p.V468I
CG-3589	chr5	45262787	C	T	437	0.5%	509	48.7%	225	HCN1	Missense	p.A637T
D-09-0386-T	chr5	45696092	g	T	122	0.0%	100	51.0%	225	HCN1	Missense	p.A35D
D-09-0362-T	chr5	129241099	C	T	368	0.3%	284	35.6%	225	CHSY3	Missense	p.R193W
CG-3596	chr5	150943071	G	A	531	0.0%	272	28.7%	225	FAT2	Missense	p.A1130V
D-09-0527-T	chr5	161116713	G	A	129	0.0%	275	39.3%	225	GABRA6	Missense	p.E201K
D-10-0171-T	chr5	161119130	G	C	943	0.0%	435	39.8%	225	GABRA6	Missense	p.R337T
CG-3572	chr6	46847711	C	A	719	0.1%	392	40.1%	225	GPR116	Nonsense///Nonse	p.E294*///p.E294*
D-09-0634-T	chr6	51768831	G	A	32	0.0%	432	43.1%	106	PKHD1	Missense///Misser	p.T2273M///p.T2273M
CG-3553	chr6	51920425	G	A	583	0.2%	410	17.1%	130	PKHD1	Missense///Misser	p.P599L///p.P599L
CG-3567	chr8	25181445	G	A	592	0.0%	759	40.4%	225	DOCK5	Missense	p.R566K
CG-3602	chr8	25261204	C	T	494	0.0%	294	43.2%	225	DOCK5	Missense	p.S1686L
D-06-0427-T	chr8	104933973	G	A	586	0.0%	425	21.9%	193	RIMS2	Missense///Misser	p.M719I///p.M527I
CG-3612	chr8	105001535	G	A	555	0.4%	450	28.7%	225	RIMS2	Missense///Misser	p.R977Q///p.R769Q
D-10-0171-T	chr8	105025741	G	A	912	0.2%	440	41.8%	225	RIMS2	Missense///Misser	p.R1039Q///p.R853Q
CG-3609	chr9	16437492	C	T	361	0.3%	153	35.3%	225	BNC2	Missense	p.A234T
CG-3609	chrX	39914722	C	T	511	0.2%	292	36.0%	225	BCOR	Missense///Misser	p.R1513Q///p.R1495Q
CG-3609	chrX	39923014	G	A	479	0.4%	240	17.9%	56	BCOR	Nonsense///Nonse	p.Q1198*///p.Q1180*
D-08-0624-T	chr2	141526899	G	A	9216.77	697	287	410	c.5641C>T	LRP1B	Missense	p.L1881F
D-01-0141-T	chr20	40743859	G	A	9129.77	912	532	380	c.3079C>T	PTPRT	Missense///Misser	p.R1027C///p.R1046C
CG-3609	chr14	71500751	G	A	118.77	512	450	62	c.3772G>A	PCNX	Missense	p.V1258I
CG-3609	chr12	57553626	G	A	155.77	404	350	50	c.1817G>A	LRP1	Missense	p.G606D
CG-3469	chrX	39930363	C	T	448	0.002232	341	0.369501	225	BCOR	Missense///Misser	p.G1034D///p.G1016D
CG-3469	chr5	11111107	C	T	814	0.001229	298	0.325503	225	CTNND2	Missense	p.A776T
MB54	chr5	11199650	T	G	290	0	180	0.272222	225	CTNND2	Missense	p.K629Q

CG-3346	chr12	57910759	C	T	267	0	370	0.278378	225	DDIT3	Missense	p.A115T
CG-3469	chr12	57910825	C	T	707	0	321	0.274143	225	DDIT3	Missense	p.E93K
CG-3442	chr5	150920231	T	C	708	0.001412	460	0.430435	225	FAT2	Missense	p.K2979R
MB67	chr5	150933942	C	G	249	0	201	0.358209	225	FAT2	Missense	p.G1309A
CG-3469	chr5	150947868	G	A	742	0	325	0.344615	225	FAT2	Missense	p.L209F
CG-3512	chr6	46826711	C	T	528	0	609	0.241379	225	GPR116	Missense///Misser	p.E977K///p.E977K
CG-3507	chr5	45262184	G	A	423	0.002364	372	0.336022	225	HCN1	Missense	p.R838C
CG-3375	chr5	45262126	G	T	619	0.003231	492	0.321138	225	HCN1	Missense	p.P857H
MB54	chr15	28424116	C	T	250	0	171	0.175439	105	HERC2	Missense	p.G3027D
MB54	chr2	141660555	T	C	251	0.003984	207	0.173913	94	LRP1B	Missense	p.T1234A
CG-3346	chr2	170072888	C	T	290	0	410	0.309756	225	LRP2	Missense	p.A1901T
CG-3469	chr2	170115595	C	T	965	0.001036	497	0.203219	177	LRP2	Missense	p.R818K
CG-3346	chr22	21345988	C	T	247	0	289	0.273356	225	LZTR1	Missense	p.T288I
CG-3466	chr12	117710315	C	T	232	0	254	0.429134	225	NOS1	Missense	p.V572M
CG-3346	chr12	117724006	G	A	278	0	434	0.276498	225	NOS1	Missense	p.T398I
CG-3346	chr12	117703242	C	T	270	0	353	0.212465	122	NOS1	Missense	p.R672H
CG-3465	chrX	91133620	T	C	544	0	175	0.251429	224	PCDH11X	Missense///Misser	p.V794A///p.V794A///
CG-3346	chrX	91518135	C	T	167	0	234	0.192308	75	PCDH11X	Missense	p.A1046V
MB54	chrX	91642930	C	A	104	0	72	0.138889	66	PCDH11X	Missense///Misser	p.S1114Y///p.S1104Y
MB54	chrX	91873904	G	A	190	0	160	0.13125	49	PCDH11X	Missense///Misser	p.G1337S///p.G1327S
CG-3469	chr14	71492853	C	T	851	0.00235	451	0.290466	225	PCNX	Missense	p.A1068V
MB54	chr14	71434960	G	C	241	0	249	0.236948	225	PCNX	Missense	p.G172R
MB63	chr6	51613077	A	G	185	0.005405	180	0.5	225	PKHD1	Missense///Misser	p.S3113P///p.S3113P
CG-3346	chr6	51936946	G	A	365	0	494	0.263158	225	PKHD1	Missense///Misser	p.P190L///p.P190L
MB54	chr6	51735365	C	A	255	0	217	0.175115	92	PKHD1	Missense///Misser	p.V2475F///p.V2475F
MB54	chr6	51695675	G	C	300	0.003333	207	0.15942	61	PKHD1	Missense///Misser	p.D2762E///p.D2762E
MB54	chr6	51890038	G	T	288	0	136	0.147059	80	PKHD1	Missense///Misser	p.L1524I///p.L1524I
CG-3469	chr3	135809422	G	A	839	0.001192	383	0.334204	225	PPP2R3A	Missense///Misser	p.G947E///p.G326E
CG-3469	chr8	104922404	G	A	900	0.003333	375	0.309333	225	RIMS2	Missense///Misser	p.S556N///p.S364N
CG-3477	chr8	105263324	A	T	738	0	607	0.163097	112	RIMS2	Missense///Misser	p.K1255I///p.K1069I
CG-3477	chr9	27158002	C	A	813	0	531	0.250471	225	TEK	Missense	p.Q76K
CG-3469	chr20	51872033	G	A	648	0.001543	295	0.352542	225	TSHZ2	Missense	p.G679E

Supplementary Table 6. Enrichment of copy number variants and single nucleotide variants in MutComFocal genes. For each gene, statistical significance for enrichment of amplified genes for SNVs as well as deleted genes for SNVs was assessed using Fisher's exact test. Annotation information in each column is described below:

gene: MutComFocal gene
mut: Number of samples with mutations (SNVs)
amp+mut: Number of samples with amplifications and mutations
amp: Number of samples with amplifications
p-value (amp+mut): Fisher's exact p-value for enrichment of amplifications and mutations
del+mut: Number of samples with deletions and mutations
del: Number of samples with deletions
p-value (del+mut): Fisher's exact p-value for enrichment of deletions and mutations

Gene	Mutations	Amplification/Mutation	Amplification	Fisher exact (amp/mut)	Deletion/Mutation	Deletion	Fisher exact (del/mut)
LZTR1	5	0	2	1	4	27	0.007005
GPR116	4	0	1	1	1	8	0.22998
PCDH11X	3	0	13	1	1	19	0.38504
PCNX	4	0	1	1	1	23	0.55191
BCOR	5	1	14	0.44512	0	23	1
BCORL1	1	0	15	1	0	24	1
CTNND2	3	0	1	1	0	3	1
DDIT3	1	1	20	0.15625	0	5	1
FAT2	6	0	1	1	0	3	1
HCN1	5	0	1	1	0	3	1
HERC2	4	0	5	1	0	13	1
LRP1B	3	0	1	1	0	0	1
LRP2	6	0	0	1	0	0	1
NLRP2	1	0	21	1	0	8	1
NOS1	5	0	2	1	0	2	1
NOS1AP	1	0	3	1	0	0	1

PCNXL3	1	0	1	1	0	1	1
PKHD1	5	0	0	1	0	6	1
PKHD1L1	1	0	3	1	0	1	1
PPP2R3A	4	0	3	1	0	5	1
RIMS2	6	0	3	1	0	1	1
TEK	3	0	2	1	0	56	1
TEKT4	1	0	1	1	0	1	1
TSHZ2	3	0	27	1	0	0	1

Supplementary Table 7. Correlation between copy number and expression in MutComFocal gene candidates. Pearson's correlation is computed along with the corresponding p-value calculated by the paired Student's t-test.

Gene	Corr	P-val
DDIT3	0.650937	<10-6
LZTR1	0.356576	<10-6
PCNX	0.407483	<10-6
PPP2R3A	0.230559	0.000001
HERC2	0.203723	0.000016
RIMS2	0.152049	0.001362
TEK	0.134765	0.004583
CTNND2	0.101266	0.033501
NOS1	0.094272	0.047872

Supplementary Table 8. Allele-specific expression of MutComFocal gene candidates. For each mutation, we calculate the number of reads calling the reference (R) and variant (V) allele. We calculate a score for allele-specific expression of the variant V/(V+R), as well as the corresponding pooled p-value per gene. Finally, we include the copy number and expression log ratios between tumor and normal per gene. Annotation information in each column is described below:

Gene: MutComFocal gene
sample: TCGA sample
ref: Reference allele
var: Variant allele
CNV: Copy number log2 ratio
Exp: Expression log2 ratio
total_depth: Total depth of reads
ref_depth R: Depth of reads calling reference allele
var_depth V: Depth of reads calling variant allele
V/(V+R): Monoallelic expression of variant allele
pval_per_gene: Pooled p-value of monoallelic expression of a gene

Gene	sample	ref	var	CNV	Exp	total_ depth	ref_de pth R	var_de pth V	V/(V+R)	pval_per _gene
LZTR1	TCGA-06-2562.chr22:21340179-21340179.T.C.vcf	T	C	-0.7697	-1.365875	50	15	35	0.7	0.000704
LZTR1	TCGA-06-5413.chr22:21344765-21344765.G.A.vcf	G	A	-0.4326	-1.723125	40	15	24	0.615384	0.000704
BCOR	TCGA-06-2559.chrX:39923055-39923055.C.T.vcf	C	T	0.1682	-0.491733	17	2	15	0.882352	0.002079
BCOR	TCGA-06-0171.chrX:39911528-39911528.A.C.vcf	A	C	0.3669	-1.192733	30	30	0	0	0.002079
HCN1	TCGA-14-1034.chr5:45262901-45262901.A.G.vcf	A	G	0.0476	-0.539077	3	3	0	0	0.065751
HCN1	TCGA-06-0747.chr5:45262309-45262309.T.A.vcf	T	A	0.0326	1.483769	3	3	0	0	0.065751
HCN1	TCGA-14-1034.chr5:45262901-45262901.A.G.vcf	A	G	0.0476	-0.539077	0	0	0	0	0.065751
HCN1	TCGA-26-5135.chr5:45262136-45262136.G.A.vcf	G	A	-0.0083	-1.142545	0	0	0	0	0.065751
HCN1	TCGA-26-5135.chr5:45353296-45353296.C.T.vcf	C	T	-0.0083	-1.142545	0	0	0	0	0.065751
FAT2	TCGA-14-1034.chr5:150901466-150901466.G.A.vcf	G	A	0.0381	0.150333	0	0	0	0	0.088672
FAT2	TCGA-06-5413.chr5:150947262-150947262.G.A.vcf	G	A	0.0222	-0.505333	1	1	0	0	0.088672
FAT2	TCGA-14-1034.chr5:150901466-150901466.G.A.vcf	G	A	0.0381	0.150333	0	0	0	0	0.088672
FAT2	TCGA-19-2629.chr5:150934173-150934173.A.G.vcf	A	G	0.0516	0.254	0	0	0	0	0.088672
PKHD1	TCGA-06-0152.chr6:51889738-51889738.G.A.vcf	G	A	0.0667	0.073818	0	0	0	0	0.088672

PKHD1	TCGA-06-0171.chr6:51799064-51799064.G.A.vcf	G	A	0.1631	-0.172182	0	0	0	0	0.088672
PKHD1	TCGA-06-1804.chr6:51523897-51523897.C.T.vcf	C	T	0.0226	-0.585455	0	0	0	0	0.088672
PKHD1	TCGA-26-5134.chr6:51484145-51484145.G.C.vcf	G	C	0.0235	0.033444	0	0	0	0	0.088672
PCNX	TCGA-06-0238.chr14:71495452-71495452.A.C.vcf	A	C	0.0267	-0.849	4	1	3	0.749998	0.089058
PCNX	TCGA-26-5134.chr14:71540387-71540387.C.T.vcf	C	T	-0.8095	-1.2234	3	3	0	0	0.089058
GPR116	TCGA-06-2565.chr6:46826170-46826170.G.A.vcf	G	A	-0.0145	3.228714	118	118	0	0	0.121353
GPR116	TCGA-12-0619.chr6:46856078-46856078.T.C.vcf	T	C	0.1473	1.792357	12	12	0	0	0.121353
GPR116	TCGA-27-1835.chr6:46826114-46826114.C.T.vcf	C	T	-0.8081	3.319786	34	34	0	0	0.121353
HERC2	TCGA-02-2483.chr15:28501298-28501298.A.C.vcf	A	C	-0.0668	-0.0888	1	1	0	0	0.170074
HERC2	TCGA-06-2559.chr15:28501298-28501298.A.C.vcf	A	C	-0.1225	-0.5756	0	0	0	0	0.170074
LRP2	TCGA-06-2563.chr2:170090092-170090092.G.A.vcf	G	A	-0.0111	-0.147	0	0	0	0	0.170074
LRP2	TCGA-06-2559.chr2:170009381-170009381.A.G.vcf	A	G	0.0164	-0.086091	2	2	0	0	0.170074
PCDH11X	TCGA-06-2559.chrX:91132792-91132792.G.A.vcf	G	A	0.183	0.043333	0	0	0	0	0.170074
PCDH11X	TCGA-19-2619.chrX:91132985-91132985.C.A.vcf	C	A	NaN	0.315	0	0	0	0	0.170074
RIMS2	TCGA-06-2563.chr8:104922392-104922392.C.T.vcf	C	T	-0.0144	0.645727	3	3	0	0	0.170074
RIMS2	TCGA-06-0744.chr8:105001597-105001597.C.T.vcf	C	T	0.0634	0.358455	1	1	0	0	0.170074
DDIT3	TCGA-06-0238.chr12:57911096-57911096.C.A.vcf	C	A	2.4725	2.5065	1433	1433	0	0	0.25
LRP1B	TCGA-06-2557.chr2:141986959-141986959.C.A.vcf	C	A	-0.0711	4.093	2	2	0	0	0.25
NOS1	TCGA-06-5415.chr12:117710315-117710315.C.T.vcf	C	T	0.3943	0.134125	0	0	0	0	0.25
NOS1AP	TCGA-19-2629.chr1:162257211-162257213.GAA.-.vcf	GAA	-	0.0673	-0.4058	14	14	0	0	0.25
PCNXL3	TCGA-12-0619.chr11:65383841-65383841.G.-.vcf	G	-	-0.2392	NaN	8	8	0	0	0.25
PKHD1L1	TCGA-06-0645.chr8:110476724-110476724.G.A.vcf	G	A	0.0133	-0.128545	0	0	0	0	0.25
PPP2R3A	TCGA-06-2570.chr3:135721982-135721982.G.A.vcf	G	A	0.0168	-0.081375	6	6	0	0	0.25
TEKT4	TCGA-06-0238.chr2:95537569-95537569.G.A.vcf	G	A	-0.0742	0.156	0	0	0	0	0.25
TEK	TCGA-06-0184.chr9:27229172-27229172.C.T.vcf	C	T	0.0973	0.685	18	18	0	0	0.25
TSHZ2	TCGA-12-0618.chr20:51870661-51870661.G.A.vcf	G	A	-0.0659	2.6396	12	5	6	0.545454	0.290527
CTNND2	TCGA-06-2562.chr5:11565132-11565132.C.T.vcf	C	T	-0.0899	3.7124	64	35	29	0.453125	0.970506
CTNND2	TCGA-06-0686.chr5:11082954-11082954.G.A.vcf	G	A	-0.1561	4.0373	48	29	11	0.275	0.970506
CTNND2	TCGA-06-0878.chr5:11022883-11022883.A.C.vcf	A	C	0.0531	4.3736	214	106	105	0.49763	0.970506

Supplementary Table 9. EST-based expression in the human brain of the novel GBM genes from NCBI.

UniGene	Start	brain (1100989)
314543	CTNND2	295
162757	LRP1	217
148909	CHL1	122
90791	GABRA6	115
78788	LZTR1	79
434890	HERC2	57
522898	CHD5	56
655271	RIMS2	53
136893	LRFN5	53
657729	LRP2	46
89640	TEK	45
167805	EPC1	41
656461	LRP1B	39
446559	PCNX	38
526879	PTPRT	29
362806	GPR116	27
505777	DDIT3	25
533468	HAUS6	23
195403	DOCK5	21

591255	FAT2	18
518155	PPP2R3A	12
655673	PCDH11X	10
656581	BNC2	9
132599	DOCK8	8
353176	HCN1	8
405659	PIWIL1	7
208388	FANCD2	7
473117	TSHZ2	6
534059	GSG2	4
377488	TPTE2	1
655974	USH2A	1
593292	ADCY5	1
348618	TRIML1	1
662050	PKHD1	0
611691	CHSY3	0
186424	BCOR	0
628475	TRIML2	0
448343	PIWIL3	0
735734	NOS1	0

Supplementary Table 10. Gene fusions identified through RNA sequencing. This item is presented separately as an Excel file.

Annotation information in each column is described below:

sample: Name of TCGA or private sample.
#chrom5p: 5' chromosome
#start5p: 5' genomic start coordinate
#end5p: 5' genomic end coordinate
#chrom3p: 3' chromosome
#start3p: 3' genomic start coordinate
#end3p: 3' genomic end coordinate
strand5p: 5' strand
strand3p: 3' strand
genes5p: 5' gene
genes3p: 3' gene
total_frags (split inserts+split reads): Total number of split inserts and split reads
spanning_frags (split reads): Number of split reads
GeneBreakpoint5p: The genomic coordinate of the breakpoint in the 5' gene
GeneBreakpoint3p: The genomic coordinate of the breakpoint in the 3' gene
FrameType: Reading frame of gene fusions. Values include in-frame, frameshift, or null (no transcript information was found in the Ensembl Homo_sapiens.GRCh37.60.gtf file).
FusedSequence: Reconstructed sequence of the fusion RNA transcript
ProteinStart5p: The start coordinate of the 5' protein segment
ProteinStop5p: The stop coordinate (breakpoint) of the 5' protein segment
ProteinStart3p: The start coordinate (breakpoint) of the 3' protein segment
ProteinStop3p: The stop coordinate of the 3' protein segment
ProteinSequence: Reconstructed sequence of the fusion protein
ExonBreak5p: The last exon of the 5' gene before the breakpoint
ExonBreak3p: The first exon of the 3' gene after the breakpoint

Supplementary Table 11. Relative expression of EGFR fusion and wild-type transcripts. Expression is estimated using the depth of reads covering the fusion breakpoint or wild-type exon junctions excluded from the fusion transcript. These wild-type exons include exons 25-26, 26-27, and 27-28.

EGFR-SEPT14				
Sample	FusionBp	Exon25-26	Exon26-27	Exon27-28
TCGA-28-2513	1464	21	12	25
TCGA-27-1837	796	6	5	6
TCGA-06-0750	414	69	61	101
TCGA-32-5222	495	256	190	348
TCGA-28-1747	142	426	300	502
TCGA-06-2557	13	1031	657	1254

EGFR-PSPH				
Sample	FusionBp	Exon25-26	Exon26-27	Exon27-28
TCGA-28-5209	5648	216	122	232
TCGA-06-5408	37	232	200	292
TCGA-28-5215	28	29	26	44

Supplementary Table 12. Genomic breakpoints of gene fusions detected through whole-exome DNA sequencing. This item is presented separately as an Excel file. Annotation information in each column is described below:

sample: Name of TCGA sample
split reads: Total number of split reads
gene5p: 5' gene
chr5p: 5' chromosome
sense5p: 5' sense
start5p: 5' genomic start coordinate
end5p: 5' genomic end coordinate
breakpoint5p: 5' genomic coordinate of breakpoint
exonBeforeBreakpoint5p: Exon number of 5' gene before the breakpoint
gene3p: 3' gene
chr3p: 3' chromosome
sense3p: 3' sense
start3p: 3' genomic start coordinate
end3p: 3' genomic end coordinate
breakpoint3p: 3' genomic coordinate of breakpoint
exonAfterBreakpoint3p: Exon number of 3' gene after the breakpoint
split inserts: Total number of split inserts
posA5p: Coordinate of split insert read closest to 5' end in 5' gene
posB5p: Coordinate of split insert read closest to 3' end in 5' gene
readDir5p: Read direction of split insert reads in 5' gene
posA3p: Coordinate of split insert read closest to 5' end in 3' gene
posB3p: Coordinate of split insert read closest to 3' end in 3' gene
readDir3p: Read direction of split insert reads in 3' gene.

Supplementary Table 13. Analysis of the incidence of EGFR-SEPT14 and EGFR-PSPH gene fusions in GBM harboring or not the EGFRvIII rearrangement.

Isoform	EGFR-SEPT14	EGFR-PSPH	Non-Fusion	Total
EGFRvIII	1	1	14	16
No EGFRvIII	5	2	64	71
Total	6	3	78	87

Supplementary Table 14. Enrichment of classical/mesenchymal subtype among samples with EGFR-SEPT14 or EGFR-PSPH.

	Classical	Mesenchymal	Proneural	Neural
EGFR Fusion	3	5	1	0
No Fusion	37	47	38	0
Total	40	52	39	28

Fisher's p-value = 0.0500 for the enrichment of classical/mesenchymal subtype.

Supplementary Table 15. Antibodies and concentrations used in immunofluorescence staining.

B-III Tubulin	**Mouse**	**1:400**	**Promega**
δ-Catenin	**Guinea Pig**	**1:500**	**Acris**
Fibronectin	**Mouse**	**1:1000**	**BD-Pharmingen**
Col5A1	**Rabbit**	**1:200**	**Santa Cruz Biotech**
PSD-95	**Rabbit**	**1:500**	**Invitrogen**
Smooth muscle actin	**Mouse**	**1:200**	**Sigma**

Supplementary Table 16. Antibodies and concentrations used for Western blots and immunopreciptation assays.

Anti-Vinculin	Mouse	1:400	SIGMA
Anti-N-Cadherin	Mouse	1:200	BD-Pharmingen
Cyclin A	Rabbit	1:500	Santa Cruz Biotech
p27	Mouse	1:250	BD Transduction
B-III Tubulin	Mouse	1:400	Promega
δ-Catenin	Guinea Pig	1:500	Acris
Fibronectin	Mouse	1:1000	BD-Pharmingen
p107	Rabbit	1:1000	Santa Cruz Biotech
Nestin	Mouse	1:500	BD-Pharmingen
CD133	Rabbit	1:200	Abcam
Sox2	Rabbit	1:500	Cell Signaling
EGFR	Mouse	1:1000	Millipore
AKT	Rabbit	1:1000	Cell Signaling
pAKT-S473	Rabbit	1:1000	Cell Signaling
ERK1/2	Rabbit	1:1000	Cell Signaling
pERK1/2	Rabbit	1:1000	Cell Signaling
STAT3	Rabbit	1:1000	Santa Cruz Biotech
pSTAT3-Y705	Rabbit	1:1000	Cell Signaling
LZTR1	Rabbit	1:1000	Abcam
Cul3	Rabbit	1:1000	Bethyl

Supplementary Table 17. Primers used for screening gene fusions from cDNA.

hEGFR-RT-FW1	**5'- GGGTGACTGTTTGGGAGTTGATG -3';**
hSEP14-RT-REV1	**5'- TGTTTGTCTTTCTTTGTATCGGTGC-3';**
hEGFR-RT-FW1	**5'-AGAGGTGACCACCAATCAGC-3';**
hPSPH-RT-REV1	**5'-CGTGTCCCACACAGAGACAG-3';**
hNFASC-RT- FW1	**5'- AGTTCCGTGTCATTGCCATCAAC-3';**
hNTRK1-RT-REV1	**5'- TGTTTCGTCCTTCTTCTCCACCG-3';**
hCAND1-RT- FW1	**5'- GGAAAAAATGACATCCAGCGAC-3'**
hEGFR-RT-REV1	**5'- TGGGTGTAAGAGGCTCCACAAG-3'**

Supplementary Table 18. Primers used for genomic detection of gene fusions.

genomic EGFR-FW1	**5'- GGATGATAGACGCAGATAGTCGCC-3'**
genomic SEPT14-REV1	**5'- TCCAGTTGTTTTTTCTCTTCCTCG-3'**
genomic NFASC-FW1	**5'- TCCGAGTCCAGGCTGAAAATG-3'**
genomic NTRK1-REV1	**5'- CTACTTCCTATCTCACCCCAAAAGG-3'**
genomic CAND1-FW1	**5'- GCAATAGCAAAACAGGAAGATGTC-3'**
genomic EGFR-REV1	**5'- GAACACTTACCCATTCGTTGG-3'**

Supplementary Table 19. Primers used for semiquantitative RT-PCR to detect exogenous Myc-LZTR1 WT and mutant LZTR1-R801W.

LZTR1 FW	**5'- TCCCACATCTCAGACAAGCA-3'**
His-Tag REV	**5'- TCAATGGTGATGGTGATGATG-3'**
GAPDH FW	**5'- GAAGGTGAAGGTCGGAGTCAAC-3'**
GAPDH REV	**5'- CAGAGTTAAAAGCAGCCCTGGT-3'**

References

1 **Srivastava, M. *et al.* The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 466, 720-726 (2010).**

2 **Stogios, P. J., Downs, G. S., Jauhal, J. J., Nandra, S. K. & Prive, G. G. Sequence and structural analysis of BTB domain proteins. *Genome Biol* 6, R82 (2005).**

3 **Soding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21, 951-960 (2005).**