Somatic POLE mutations cause an ultramutated giant cell high-grade glioma subtype with better prognosis

METHODS

Exome Sequencing and Analysis of sequencing results:

We performed whole-exome capture and next generation sequencing of 136 adult gliomas. Ninety-one of these samples had matching blood samples; 53 of which were primary gliomas.

Exome Capture and Sequencing: Nimblegen/Roche human solution-capture exome array (Roche Nimblegen, Inc.) was used to capture the exomes of blood and tumor samples according to the manufacturer's protocol with modifications[1]. Sequencing of the library was performed on Illumina HiSeq instruments using 74 base pairs paired-end reads by multiplexing two tumor samples or three blood samples per lane.

We have performed a deeper coverage of tumors as compared to matching blood samples to achieve a better resolution with subclonal mutations (average target coverage was 194.3 and 121.3, respectively). The average percentage of reads with at least 20x coverage was 91.0% and 88.4% for tumor and blood, respectively (Supplementary Table S1).

We have performed a quality control step on the raw reads before alignment for filtering out low quality reads and adapter contamination as detailed previously[2].

The alignment is performed to the human genome reference sequence (version GRCh37) with BWA (version 0.5.9-r16)[3] followed by Stampy (version 1.0.16)[4]. PCR duplicates were excluded from further analysis as previously described[5] using MarkDuplicates algorithm from Picard (version 1.47, http://picard.sourceforge.net/).

We performed multi-sequence local realignment around putative and known insertion/deletion sites. This was followed by the base quality score recalibration using the Genome Analysis Toolkit (GATK, version 2.5-20)[5].

Germline and Somatic Variant Calling: For germline mutations (in blood) we used Unified Genotyper implemented in Genome Analysis Toolkit (GATK, version 2.5-2) and called variants using additional 250 exomes from individuals of European descent. For the tumor-blood matched samples, we determined the somatic mutations using Haplotyper caller implemented in Genome Analysis Toolkit (GATK, version 2.5)[5] and calculated a somatic score according to the method described by Li[6].

Variant annotation was performed after variant calling using Ensembl database (version 69) with the help of Variant Effect Predictor (VEP, v2.7) tool (http://useast.ensembl.org/info/docs/variation/vep/vep_script.html). Missense variants are annotated to be deleterious, if either SIFT [7] or Polyphen2 [8] predicts it to be deleterious or damaging. We selected the most-deleterious consequence out of all annotated transcripts for each variant site based on the consequence ordering suggested by VEP.

For variant quality filtering of the germline exome data, we eliminated variants that are within 10 base pairs from a putative insertion/deletion. We also filtered out variants using variant quality score recalibration lod score (log odds ratio of being a true vs. false variant) to achieve a required sensitivity of 99% and 95% for SNPs and INDELs respectively.

For the somatic variant quality control, we filtered out the variants according to the classes of genotype changes in tumor with respect to the normal and kept the following types of alterations (i) somatic, where the normal has a homozygous genotype for the

reference allele and the tumor has a heterozygous genotype or has a homozygous genotype for the variant allele. (ii) loss of heterozygosity, where the normal is heterozygous for the reference allele and the tumor has a homozygous genotype for the variant allele. We also used various quality metrics to filter out variants: (i) with a somatic score less than 20, (ii) overlapping a RepeatMasker or segmental duplication annotated region, (iii) with low quality (<30) and low quality-by-depth values (<1), (iv) with high mapping quality zero reads, (v) with strand bias, (vi) in a mutation cluster of size >2 (vii) with homopolymer runs of length >= 10 base pairs within +/- 5 base pairs around the mutation or from the right of the mutation or (viii) with ClippingRankSum (calculated by GATK) < -3.0 or > 3.0.

In addition, we excluded any variant with a frequency of greater than or equal to 1% in the NHLBI Exome Variant Server Database (http://evs.gs.washington.edu/EVS/) and 1000Genome Database[9]. We also used our internal database of 2,216 exomes (excluding the common SNPs) to compare the variant allele frequency for each gene and excluded the variants in genes that have greater than 150 variant alleles in this database.

Mutation Signature Analysis:

We have calculated the mutation signature of individual tumors' somatic mutations by considering 6 major mutation classes, i.e., G:C > T:A, G:C > A:T, G:C > C:G, A:T > G:C, A:T > C:G, A:T > T:A (Fig. S1). In the discovery cohort of 53 primary adult HGGs, we observed a significant enrichment on the C>T transitions with the following mean values (Fig. S1-S2, Table S3): C>T 64.78% (range = [37.50%-100%]), C>A 9.18% ([range = 0%-19.61%]), C>G 6.14% (range = [0%-16.67%]), T>A 5.26% (range = [0%-12.50]%), T>G 3.84% (range = [0%-10.55%]) and T>C 10.78% (range = [0%-22.86%]).

GBM-10468, an adult HGG with *POLE* mutation had an increased C > A transversion ratio. Out of all somatic mutations, 45.84% were C>T transitions, 20% were C>A transversions, 0.20% were C>G transversions, 0.9% were T>A transversions, 16.13% were T>G transversions and 16.79% were T>C transitions.

GBM-60001 had a similar pattern with 61% C>T transitions, 19.21% C>A transversions, 0.02% C>G transversions, 1% T>A transversions, 8.4% T>G transversions and 9.8% T>C transversions.

Overall, ultramutated samples did not show a significant difference compared to the rest of the HGGs for C>T transitions, T>G transversions and T>C transitions.

For C > A transversions, ultramutated samples had (avg=20%) a significant increase compared to the rest of the HGGs (avg= 9.4%)(P = 3.8e-06). For C>G transversions ultramutated samples had (mean= 0.2%) a significant decrease in this class of mutations compared to the rest of the HGGs (mean= 6.3%)(P = 4.15e-16). Similarly for T>A transversions, ultramutated samples had (mean= 0.9%) a significant decrease compared to the rest (mean= 5.2%, P = 1.91e-14) (Fig. S2).

Two-sided t-test was used to calculate the significance of signature distributions among the ultramutated and non-ultramutated samples.

Copy number variation (CNV) and admixture rate calculation from exome data

The log ratio of depth of coverage between tumor and blood was calculated using GATK-*Depth Of Coverage* tool. CNV segments were then called from the log ratio of depth of coverage using ExomeCNV R package[10]. False positive CNV events were corrected by calculating minor allele frequencies (BAF) in each CNV segment. (Fig. S3)

We estimated the admixture rate based on CNV analysis of paired tumor and blood samples. Copy number loss regions were extracted and for those regions the BAF of each tumor snp that was heterozygous in blood were calculated. Finally the admixture rate was estimated from the degree of deviation from homozygosity using qpure R package[11].

CNV and mutation status of all HGGs in Yale cohort (53 adult, 2 pediatric) for genes that are frequently altered in GBMs are listed in Supplementary Table S4.

Clonality Analysis:

Clonality rate is defined to be the percent of tumor cells harboring the identified somatic mutation and correlates with the temporal evolution of the tumor[12,13].

We estimated the percent of cells that harbor the heterozygous somatic mutations based on the observed variant allele frequency, ploidy at the site of variant and the admixture rate similarly as previously described[14]. We observed that *POLE* exonculease domain mutations were clonal, i.e > 80% of tumor cells harbored these mutations in all cases. Moreover, we analyzed the distribution of variant allele frequencies of all somatic mutations (coding and non-coding) for all ultra-mutated cases and observed that majority of somatic mutations had a lower variant allele frequency than *POLE* exonuclease mutations, suggesting that the increased mutation burden occurred after the *POLE* mutations.

Sanger Confirmation for the segregation of *MSH6* mutation

Coding regions and exon-intron boundaries of *MSH6* were evaluated by Sanger sequencing using standard protocols. Amplicons were cycle sequenced on ABI 9800 Fast

Thermo cyclers, and post cycle sequencing clean-up was carried out with CleanSEQ System (Beckman Coulter Genomics). The amplicons were analyzed on 3730×L DNA Analyzer (Applied Biosystems Inc.)

The mutation identified via whole-exome sequencing was confirmed as being homozygous versus heterozygous, in the two affected siblings versus their parents, respectively (Fig. S4).

Clinical and histological features

Clinical features for all tumors, such as age at diagnosis, progression-free survival time were analyzed, when available (Supplementary Table 3). Histological features of the samples were analyzed by two independent neuropathologists using H&E, GFAP and nuclear p53 stainings, when available.

Treatment information for POLE mutant cases in Yale cohort:

All adult cases and the older of the two pediatric cases (age =8) have been treated according to standard protocols, including maximum surgical resection, followed by radiation and temolozomide therapy. The younger sibling had maximum surgical resection, followed by ciplatin and etoposide therapy. Given the age of the patient, radiation therapy was not recommended.

TCGA data access and analysis:

The clinical information for 567 GBM samples are downloaded from https://tcga-data.nci.nih.gov/tcga/. Mutation information for the 2 ultramutated GBM samples from the TCGA database is provided by the Center for Molecular Oncology and the Computational Biology Center at Memorial Sloan-Kettering Cancer Center through

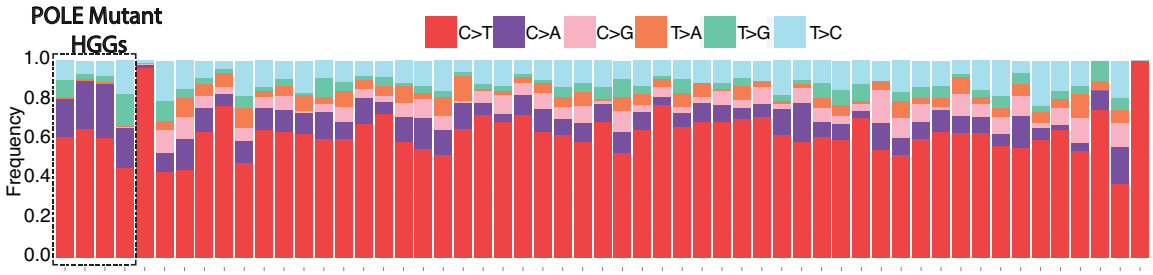
cbiportal (http://www.cbioportal.org/)/. The data included the coding somatic mutations for the 2 cases with genomic positions, including reference and alternate alleles, variant impact class, variant type, transcrtip change. We have used the protein altering mutation count to assess the ultramutation phenotype. We have also used the reference and alternate allele information to compare the mutation signature of these samples to the previously identified *POLE* mutated samples' signature.

Survival Analysis

We have used the survival R package[15] to perform the Kaplan-Meier analysis on the time to recurrence metric between the *POLE* mutated ultra-mutated samples and the rest. We used the same package to calculate the logrank p value for the significance of difference in time to recurrence in two datasets, *POLE* mutant vs. *POLE* wildtype HGGs.

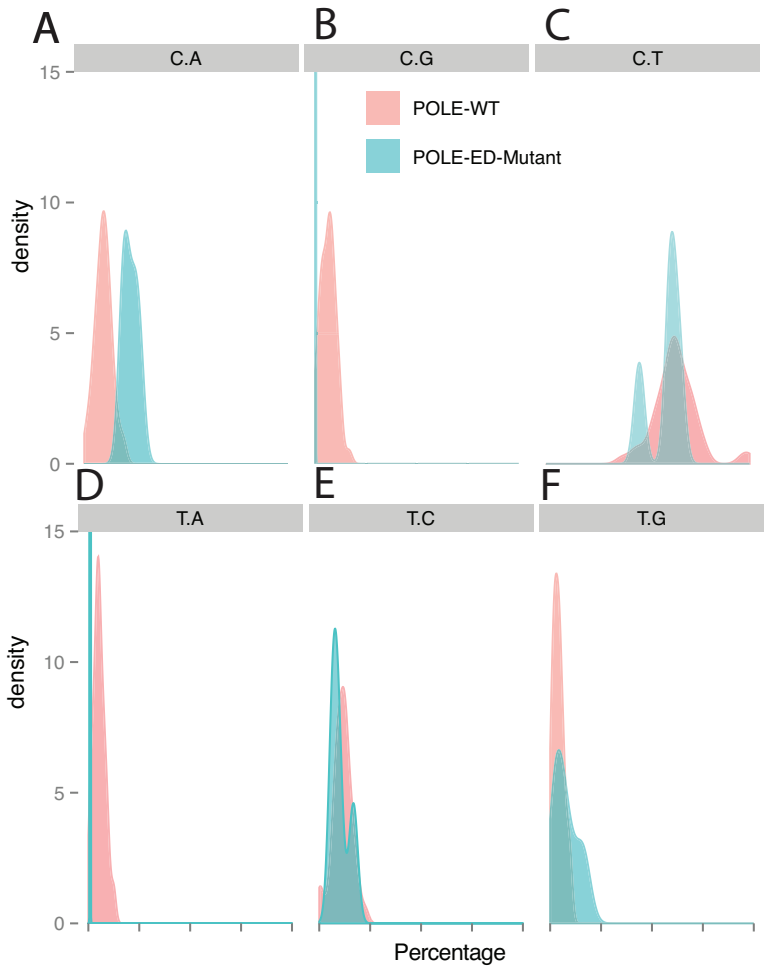
In order to assess the clinical differences among the ultramutated adult samples and remaining HGGs, we compared the age at diagnosis. The average age in 4 adult HGGs with *POLE* mutation was 35.5, whereas the non-ultramutated HGGs in Yale cohort was 58 (range =[23-42] vs. [22-83]). The two-tailed t-test was used to calculate the p-value of 0.005. The median age for anaplastic astrocytomas and glioblastomas is reported as 54 and 64, respectively[16].

FIGURES:



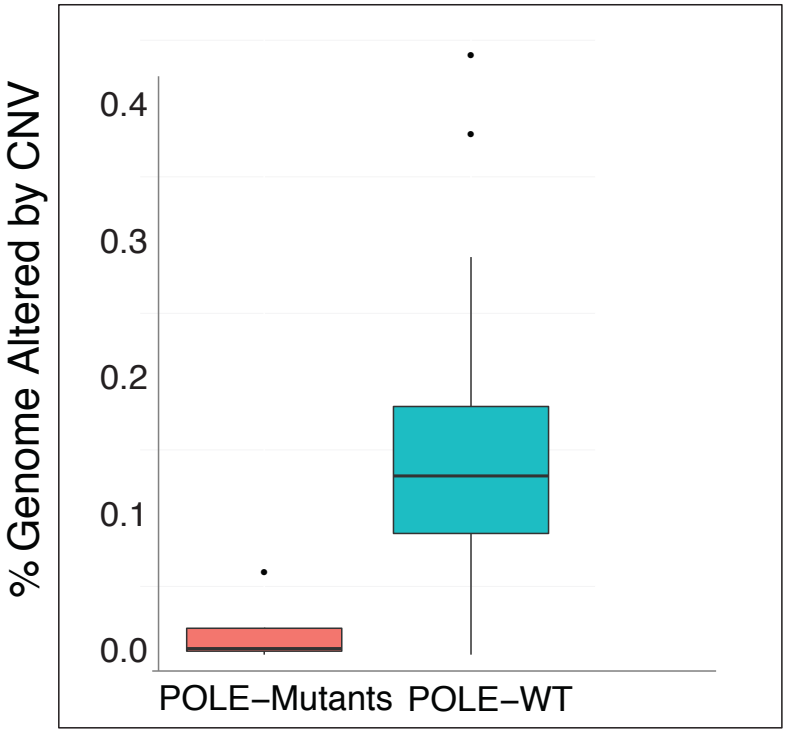
Supplementary Figure S1: Mutation Spectrum for 55 primary HGGs.

Each of the 55 tumor samples are plotted along the horizontal axis. Analysis of mutation signatures, as determined by the relative percentage of all 6 possible types of single nucleotide mutation types. The ultramutated samples, which cluster to the left side of the panel and marked by the box, are characterized by an increased percentage of C>A mutations and decreased C>G and T>G as compared to the other samples.



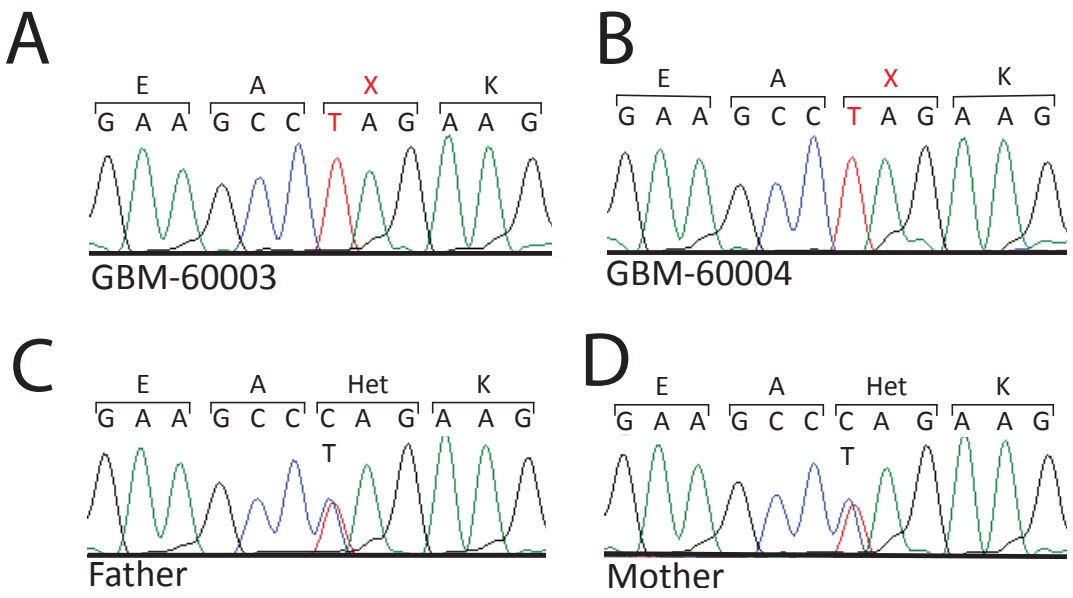
Supplementary Figure S2: Mutation signature distribution in 55 primary HGGs

Density distribution of the 6 classes of mutation in 2 groups of samples, *POLE* mutant ultramutated HGGs (POLE-ED-Mutant) and the rest (POLE-WT). Ultramutated HGGs have an increased C > A transversion ratio (A). Interestingly, *POLE* mutant samples show a significantly low ratios of C > G (B) and T > A transversions (D). C > T (C), T > C transitions (E) and T > G (F) transversions show a similar distribution for both groups.



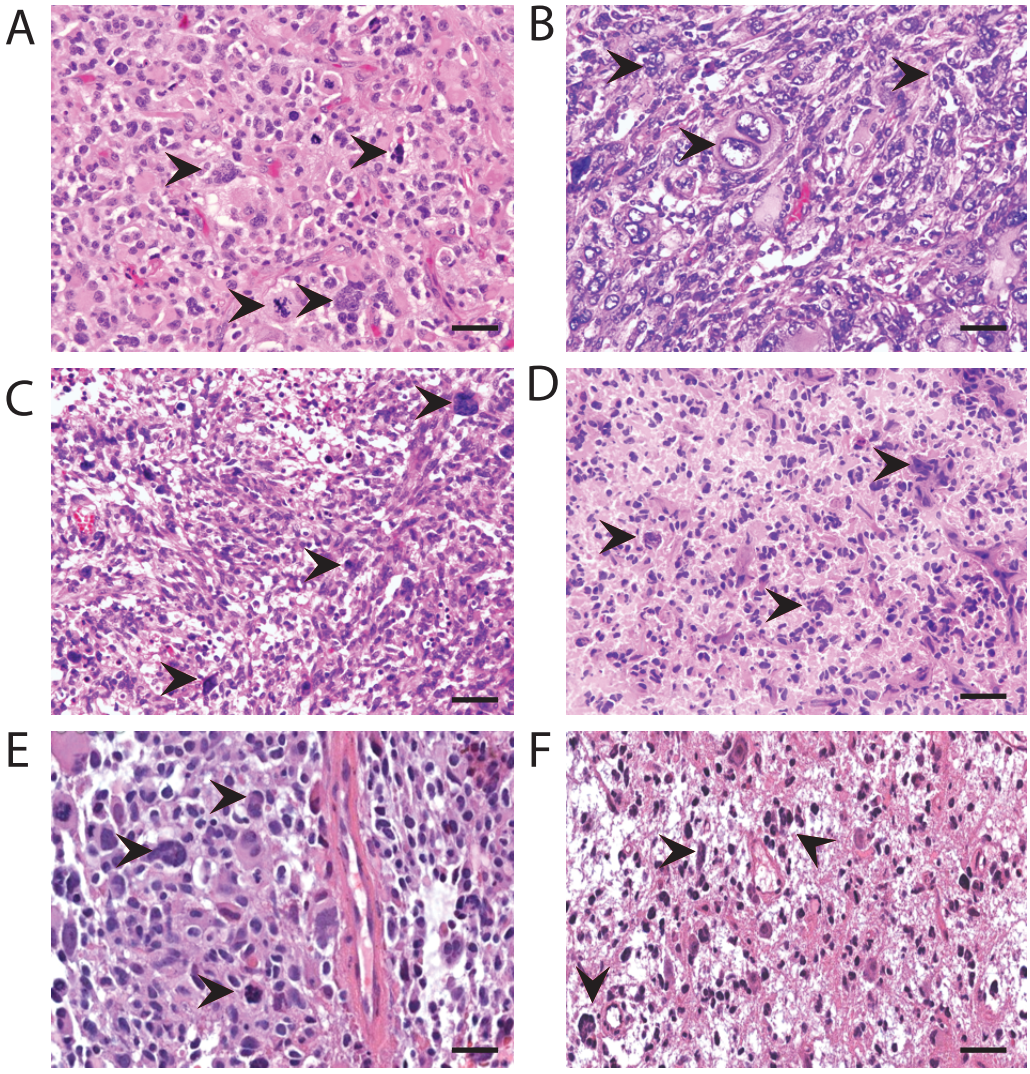
Supplementary Figure S3: Distribution of percentage of genome alteration by CNV events.

POLE-Mutant samples include 2 adult and 2 pediatric samples and the *POLE*-WT samples are the remaining 51 adult primary HGGs in Yale cohort. (P = 9.575e-05, two-sided t-test).



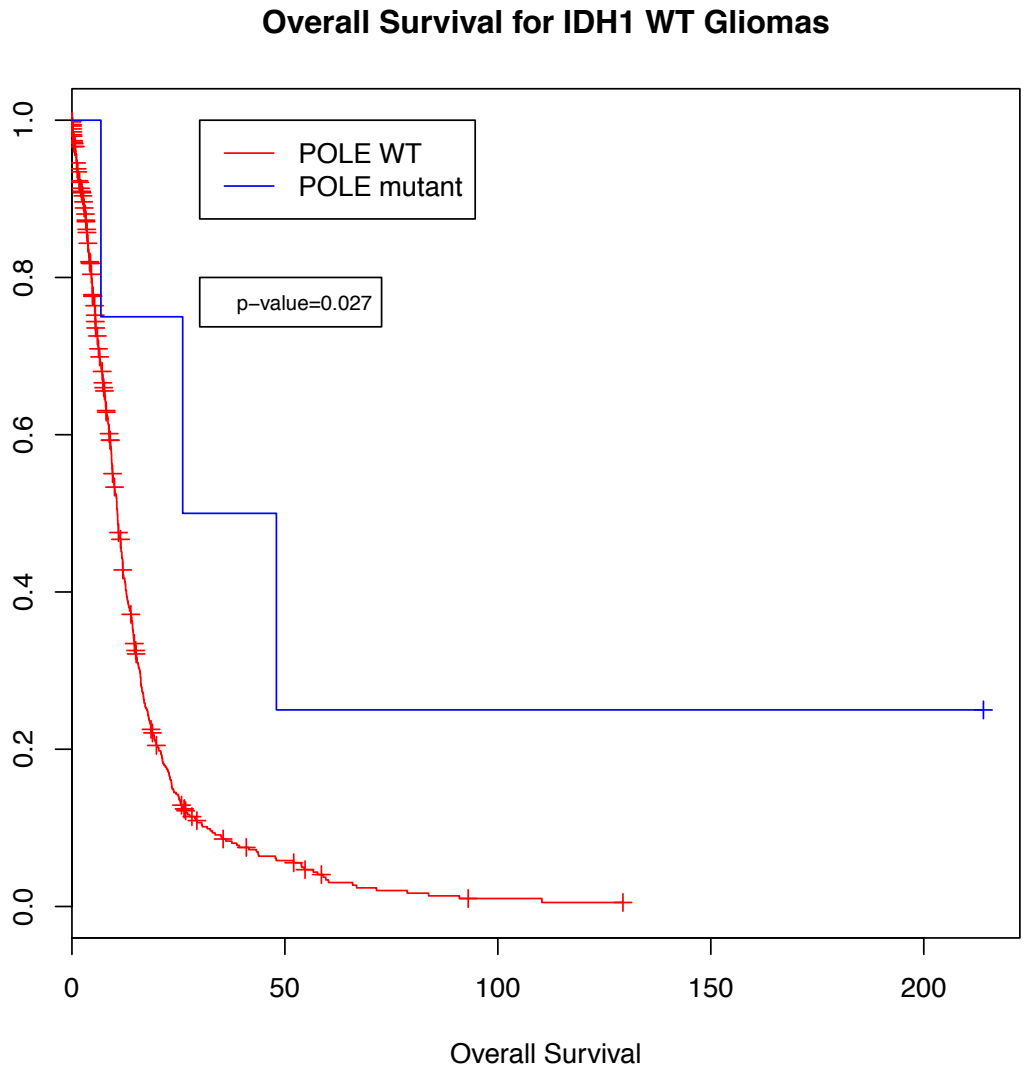
Supplementary Figure S4: Sanger confirmation for the germline homozygous *MSH6* in pediatric GBM cases

Sanger confirmation for the germline homozygous *MSH6* mutation in pediatric GBM cases (A-B) and their parents (C-D) are shown. The mutation identified via whole-exome sequencing was confirmed as being homozygous in the two affected siblings and heterozygous in their parents. The bases outlined in red in the sequence indicate the mutated base pairs.



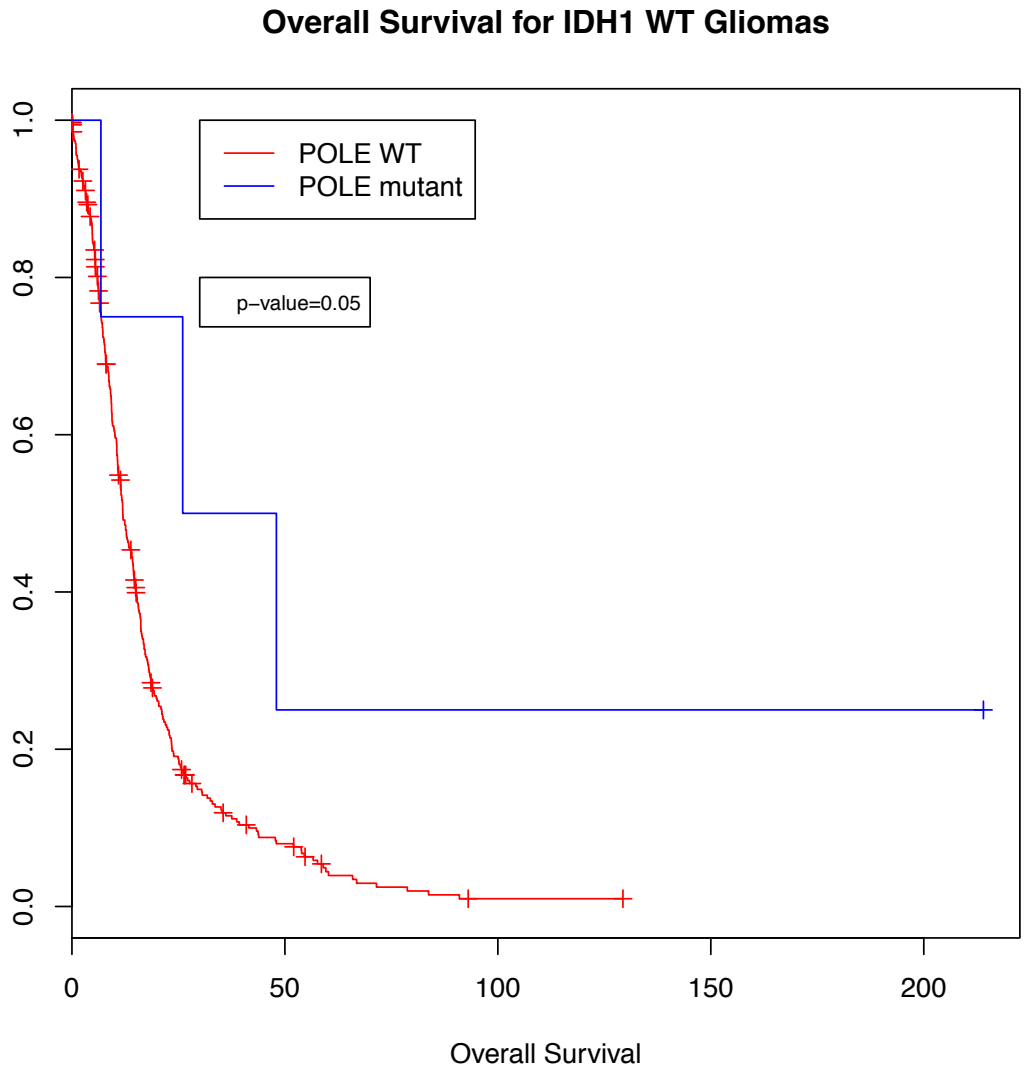
Supplementary Figure S5: H&E stained sections of 6 *POLE* mutant cases

H&E stained sections show numerous large multinucleated giant cells with clumped nuclei, and cells with many smaller, eccentrically placed nuclei (black arrowheads) for (A) GBM-60001, (B) GBM-60004, (C) GBM-60003, (D) GBM-10468, (E)-(F) TCGA cases. Scale bars = 50μm.

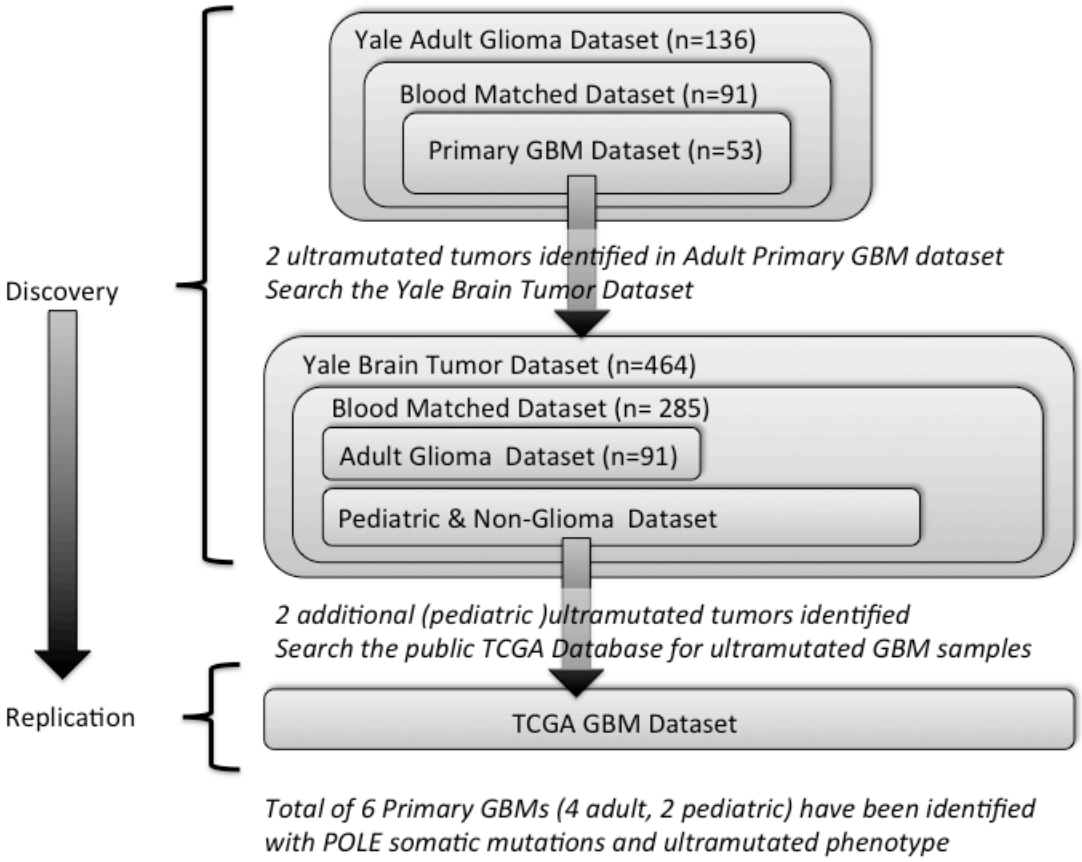


Supplementary Figure S6: Time to recurrence analysis with the TCGA GBM dataset (n=569)

Time to recurrence analysis using the public TCGA data (n=567) and the *POLE* mutant Yale adult samples (n=2), with total samples size of 569. The p-value (p = 0.026) calculated to assess the significance of difference in the time of recurrence between *POLE* mutant and wildtype samples is the same as reported with the Yale discovery cohort (n=53).



Supplementary Figure S7: Time to recurrence analysis with the TCGA GBM dataset with samples younger than 64 (n=365)



Supplementary Figure S8: Flow chart for the identification of POLE mutated ultramutated gliomas.

References:

1. Bilguvar K, Ozturk AK, Louvi A, et al. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature.* 2010; 467(7312):207-210.
2. Clark VE, Erson-Omay EZ, Serin A, et al. Genomic Analysis of Non-NF2 Meningiomas Reveals Mutations in TRAF7, KLF4, AKT1, and SMO. *Science.* 2013; 339(6123):1077-1080.
3. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009; 25(14):1754-1760.
4. Lunter G, Goodson L. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research.* 2011.
5. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011; 43(5):491-498.
6. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011; 27(21):2987-2993.
7. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols.* 2009; 4(7):1073-1081.
8. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010; 7(4):248-249.
9. Consortium TGP. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012; 491(7422):56-65.
10. Sathirapongsasuti JF, Lee H, Horst BAJ, et al. Exome Sequencing-Based Copy-Number Variation and Loss of Heterozygosity Detection: ExomeCNV. *Bioinformatics.* 2011.
11. Song S, Nones K, Miller D, et al. qpure: A Tool to Estimate Tumor Cellularity from Genome-Wide Single-Nucleotide Polymorphism Profiles. *PLoS ONE.* 2012; 7(9):e45835.
12. Nik-Zainal S, Van†Loo P, Wedge DC, et al. The Life History of 21 Breast Cancers. *Cell.* 2012; 149(5):994-1007.
13. Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet.* 2012; 13(11):795-806.
14. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature.* 2012; 486(7403):400-404.
15. Therneau TMG, Patricia M. *Modeling Survival Data: Extending the Cox Model*: New York; 2000.
16. Ostrom QT, Gittleman H, Farah P, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2006-2010. *Neuro-Oncology.* 2013; 15(suppl 2):ii1-ii56.

TABLES

Table S1. Sequencing metrics for 55 primary HGG cases from the Yale cohort. Sequencing metrics for 55 primary (i.e. wildtype for *IDH1-R132*) HGG cases from the Yale cohort are listed. *POLE* mutant ultramutated cases are marked with red fonts

Sample ID	% Mismatches	Mean Target Coverage	% Target Bases at 20x
GBM-60001	0.21%	233.69	94.91%
GBM-60003	0.30%	153.08	93.58%
GBM-60004	0.25%	197.51	93.68%
GBM-10468	0.34%	281.31	94.59%
GBM-20032	0.28%	227.65	94.10%
GBM-10457	0.13%	153.54	92.59%
GBM-30239	0.15%	196.62	93.67%
GBM-30056	0.27%	280.11	94.53%
GBM-30021	0.33%	209.87	92.06%
GBM-10352	0.61%	181.85	89.44%
GBM-39035_1	0.26%	209.51	91.35%
GBM-30092	0.24%	296.38	94.21%
GBM-10448	0.35%	245.30	94.22%
GBM-30099	0.24%	182.19	92.13%
GBM-20016	0.32%	212.30	91.34%
GBM-20045	0.22%	258.04	94.76%
GBM-30059	0.27%	251.67	93.37%
GBM-20030	0.38%	100.19	89.36%
GBM-10449	0.29%	237.74	93.98%
GBM-20034	0.36%	255.76	93.45%
GBM-20010	0.28%	232.23	92.85%
GBM-20012	0.32%	208.95	91.72%
GBM-20017	0.30%	298.13	94.31%
GBM-30143	0.25%	319.48	95.41%
GBM-10265	0.28%	233.96	93.32%
GBM-20048	0.25%	201.42	93.32%

GBM-10450	0.27%	141.69	87.84%
GBM-30026	0.28%	246.06	92.19%
GBM-10365	0.22%	175.76	90.21%
GBM-20044	0.22%	263.81	94.08%
GBM-20031	0.29%	193.65	90.70%
GBM-30109	0.59%	213.27	94.00%
GBM-10132	0.28%	254.76	94.01%
GBM-20050	0.21%	271.86	94.91%
GBM-30031	0.33%	184.70	90.99%
GBM-39003	0.15%	126.96	89.44%
GBM-30107	0.56%	199.94	93.27%
GBM-10269	0.32%	185.05	91.27%
GBM-10474	0.21%	289.94	95.24%
GBM-10355	0.21%	180.47	90.02%
GBM-20006	0.33%	228.37	91.80%
GBM-10461	0.32%	236.99	94.14%
GBM-30118	0.57%	247.09	94.52%
GBM-30142	0.25%	341.20	95.54%
GBM-30028	0.24%	241.55	93.03%
GBM-10400	0.19%	139.06	87.13%
GBM-30029	0.26%	289.97	93.98%
GBM-20013	0.20%	202.30	91.32%
GBM-39005	0.17%	152.01	90.43%
GBM-20028	0.56%	72.44	81.53%
GBM-30111	0.93%	176.66	92.49%
GBM-20043	0.38%	124.13	90.36%
GBM-20041	0.19%	134.63	87.83%
GBM-10333	0.35%	155.10	92.15%
GBM-20051	0.21%	285.80	95.04%

Table S2. Mutation signature data

Mutation Signature data for 55 primary HGG (i.e. wildtype for *IDH1-R132*) cases from Yale and 2 *POLE* mutant cases from TCGA are listed. *POLE* mutant ultramutated cases are marked with red fonts.

Sample ID	C>T	C>A	C>G	T>A	T>G	T>C
GBM-60004	61.16%	27.07%	0.10%	0.81%	3.31%	7.56%
TCGA-06-5416	61.87%	25.11%	0.23%	0.68%	4.35%	7.76%
TCGA-DU-6392	59.09%	25.06%	0.34%	1.47%	3.27%	10.77%
GBM-60003	65.58%	24.18%	0.11%	0.82%	2.90%	6.41%
GBM-10468	45.83%	20.12%	0.21%	0.91%	16.13%	16.80%
GBM-10474	58.82%	19.61%	7.84%	1.96%	1.96%	9.80%
GBM-60001	61.32%	19.21%	0.22%	1.01%	8.43%	9.80%
GBM-10333	37.50%	18.75%	12.50%	6.25%	6.25%	18.75%
GBM-30239	44.57%	16.00%	10.86%	9.71%	4.57%	14.29%
GBM-20013	56.00%	16.00%	10.00%	6.00%	6.00%	6.00%
GBM-10449	55.34%	15.53%	9.71%	2.91%	3.88%	12.62%
GBM-20010	65.17%	13.48%	1.12%	12.36%	2.25%	5.62%
GBM-30059	67.91%	13.43%	4.48%	4.48%	2.24%	7.46%
GBM-30099	60.40%	13.42%	4.03%	4.03%	9.40%	8.72%
GBM-10461	55.00%	13.33%	16.67%	5.00%	0.00%	10.00%
GBM-30107	62.32%	13.04%	2.90%	1.45%	2.90%	17.39%
GBM-20034	52.11%	12.68%	7.04%	9.86%	2.82%	15.49%
GBM-20030	58.75%	12.50%	7.50%	8.75%	1.25%	11.25%
GBM-10265	63.74%	12.09%	7.69%	5.49%	1.10%	9.89%
GBM-30056	64.05%	11.76%	6.54%	5.88%	3.27%	8.50%
GBM-39035_1	64.76%	11.43%	5.71%	2.86%	1.90%	13.33%
GBM-30092	64.00%	11.00%	7.00%	5.00%	4.00%	9.00%
GBM-10352	48.35%	10.99%	6.59%	9.89%	6.59%	17.58%
GBM-30118	64.06%	10.94%	1.56%	4.69%	6.25%	12.50%
GBM-20044	53.03%	10.61%	10.61%	7.58%	9.09%	9.09%
GBM-10448	62.86%	10.48%	0.95%	8.57%	0.95%	16.19%
GBM-30143	72.41%	10.34%	5.75%	3.45%	1.15%	6.90%
GBM-10450	58.57%	10.00%	8.57%	7.14%	4.29%	11.43%

GBM-20041	75.00%	10.00%	0.00%	5.00%	10.00%	0.00%
GBM-10132	69.01%	9.86%	2.82%	7.04%	0.00%	11.27%
GBM-30026	68.67%	9.64%	1.20%	1.20%	6.02%	13.25%
GBM-10457	43.67%	9.31%	12.03%	4.22%	10.55%	20.22%
GBM-10365	64.94%	9.09%	3.90%	5.19%	3.90%	12.99%
GBM-30028	60.44%	8.79%	8.79%	3.30%	5.49%	13.19%
GBM-30142	52.17%	8.70%	10.14%	8.70%	4.35%	15.94%
GBM-20016	60.58%	8.65%	7.69%	7.69%	4.81%	10.58%
GBM-20048	62.20%	8.54%	6.10%	4.88%	4.88%	13.41%
GBM-20050	69.01%	8.45%	7.04%	1.41%	2.82%	11.27%
GBM-10400	63.33%	8.33%	6.67%	3.33%	3.33%	15.00%
GBM-39005	63.33%	8.33%	11.67%	8.33%	1.67%	6.67%
GBM-10355	60.00%	8.00%	4.00%	5.33%	8.00%	14.67%
GBM-10269	61.25%	7.50%	7.50%	5.00%	7.50%	11.25%
GBM-30109	66.28%	6.98%	3.49%	6.98%	6.98%	9.30%
GBM-30031	71.67%	6.67%	8.33%	3.33%	0.00%	10.00%
GBM-30021	77.17%	6.30%	3.15%	7.09%	2.36%	3.94%
GBM-20045	73.00%	6.00%	3.00%	5.00%	5.00%	8.00%
GBM-30029	56.72%	5.97%	8.96%	4.48%	5.97%	17.91%
GBM-20012	72.62%	5.95%	5.95%	1.19%	2.38%	11.90%
GBM-20028	60.00%	5.71%	2.86%	5.71%	2.86%	22.86%
GBM-39003	70.42%	5.63%	4.23%	7.04%	4.23%	8.45%
GBM-20017	68.82%	4.30%	9.68%	2.15%	2.15%	12.90%
GBM-20043	54.17%	4.17%	12.50%	12.50%	4.17%	12.50%
GBM-20031	77.63%	3.95%	5.26%	3.95%	1.32%	7.89%
GBM-20006	71.19%	3.39%	3.39%	1.69%	8.47%	11.86%
GBM-30111	65.22%	2.17%	8.70%	4.35%	4.35%	15.22%
GBM-20032	96.65%	1.28%	0.49%	0.30%	0.06%	1.22%
GBM-20051	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table S3. Clinical information for 55 primary HGG (i.e. wildtype for *IDH1-R132*) cases from the Yale cohort.
POLE mutant ultramutated cases are marked with red font. Status for deleterious germline MSH6 and somatic POLE mutations are reported. Hom. = Homozygous mutation, WT= mutation not found.

Sample ID	Age at DX	Diagnosis	Grade	DFS Status	DFS	OS Status	OS Months	Count of Protein alt. Mutations	Count of Syn. Mutations	% Genome alteration by CNV Events	Germline MSH6 Mutation	Somatic POLE Mutation
GBM-60001	42	GBM	IV	Yes	48	Deceased	48	7527	2493	6.09%	V379A, Hom.	V411L
GBM-60003	8	GBM	IV	Yes	24	Deceased	30	4861	57	0.55%	Q160*, Hom	S297F
GBM-60004	2	GBM	IV	No	6	Deceased	6	4780	33	0.00%	Q160*, Hom	S459F
GBM-10468	42	GBM	IV	Yes	26	Alive	28	4652	1522	0.34%	WT	P286R
GBM-20032	NA	GBM	IV	Yes	3	Alive	15	659	364	16.37%	WT	WT
GBM-10457	51	anaplastic oligo.	III	NA	NA	Alive	10	121	31	13.99%	WT	WT
GBM-30239	59	GBM	IV	NA	NA	Hospice	4	77	31	27.15%	WT	WT
GBM-30056	54	GBM	IV	NA	7	NA	13	75	23	38.10%	WT	WT
GBM-30021	76	GBM	IV	Yes	15	Hospice	15	74	18	14.22%	WT	WT
GBM-10352	47	GBM	IV	Yes	11	Deceased	20	67	26	17.63%	WT	WT
GBM-[illegible]	52	NA	IV	NA	NA	NA	NA	66	16	17.90%	WT	WT

GBM-30092	52	GBM	IV	NA	NA	NA	11	64	22	19.52%	WT	WT
GBM-10448	68	GBM	IV	NA	NA	Deceased	8	61	22	14.64%	WT	WT
GBM-30099	72	GBM	IV	NA	NA	NA	25	57	33	19.52%	WT	WT
GBM-20016	59	GBM	IV	Yes	7	Alive	17	55	16	16.22%	WT	WT
GBM-20045	NA	GBM	IV	Yes	2	Alive	17	55	11	9.12%	WT	WT
GBM-30059	73	GBM	IV	NA	NA	NA	NA	55	26	8.51%	WT	WT
GBM-20030	59	GBM	IV	Yes	16	Alive	20	53	27	8.64%	WT	WT
GBM-10449	51	GBM	IV	Yes	6	Alive	10	52	21	20.62%	WT	WT
GBM-20034	53	GBM	IV	Yes	8	Deceased	22	49	22	13.58%	WT	WT
GBM-20010	49	GBM	IV	Yes	6	Alive	21	49	14	13.05%	WT	WT
GBM-20012	55	GBM	IV	Yes	3	Alive	18	45	17	14.22%	WT	WT
GBM-20017	NA	GBM	IV	Yes	NA	NA	NA	45	16	12.82%	WT	WT
GBM-30143	63	GBM	IV	No	6	Alive	6	44	18	8.10%	WT	WT
GBM-10265	49	GBM	IV	Yes	5	Deceased	28	42	14	21.53%	WT	WT
GBM-20048	NA	GBM	IV	Yes	8	Alive	12	42	13	14.40%	WT	WT

GBM-10450	68	GBM	IV	Yes	3	Deceased	10	39	8	43.89%	WT	WT
GBM-30026	55	GBM	IV	No	15	NA	15	39	23	18.46%	WT	WT
GBM-10365	66	GBM, oligo.	IV	Yes	7	Deceased	19	37	18	13.56%	WT	WT
GBM-20044	NA	GBM	IV	Yes	3	Alive	5	37	10	29.12%	WT	WT
GBM-20031	64	GBM	IV	Yes	8	Deceased	8	37	10	12.78%	WT	WT
GBM-30109	56	GBM	IV	Yes	11	Hospice	11	36	14	13.90%	WT	WT
GBM-10132	60	GBM	IV	Yes	3	Alive	49	35	19	26.80%	WT	WT
GBM-20050	NA	GBM	IV	Yes	4	Alive	10	35	20	12.37%	WT	WT
GBM-30031	68	Anaplastic oligo-astrocytoma	III	NA	NA	NA	12	35	9	8.57%	WT	WT
GBM-39003	83	GBM	IV	NA	NA	NA	2	35	16	7.70%	WT	WT
GBM-30107	76	GBM	IV	Yes	7	Alive	18	34	15	0.00%	WT	WT
GBM-10269	49	GBM, oligo.	IV	No	24	Alive	24	32	17	28.46%	WT	WT
GBM-10474	47	GBM	IV	Yes	14	Alive	17	32	11	26.23%	WT	WT
GBM-10355	54	GBM	IV	Yes	3	Deceased	17	31	16	9.57%	WT	WT

GBM-20006	64	GBM	IV	Yes	1	Deceased	26	31	8	0.93%	WT	WT
GBM-10461	63	anaplastic oligo.	III	No	10	Alive	10	30	7	25.11%	WT	WT
GBM-30118	42	GBM	IV	Yes	10	NA	14	30	17	10.11%	WT	WT
GBM-30142	70	GBM	IV	NA	NA	NA	4	30	15	12.08%	WT	WT
GBM-30028	69	GBM	IV	NA	NA	Hospice	4	30	25	12.98%	WT	WT
GBM-10400	61	GBM	IV	NA	NA	Deceased	15	26	14	11.50%	WT	WT
GBM-30029	62	GBM	IV	NA	NA	NA	4	26	12	5.08%	WT	WT
GBM-20013	60	GBM	IV	Yes	12	Deceased	12	26	10	5.78%	WT	WT
GBM-39005	48	GBM	IV	NA	NA	NA	NA	26	13	11.89%	WT	WT
GBM-20028	63	GBM	IV	Yes	6	Deceased	28	23	5	12.71%	WT	WT
GBM-30111	65	GBM	IV	No	17	Alive	17	18	14	13.09%	WT	WT
GBM-20043	22	GBM	IV	Yes	6	Deceased	26	16	3	7.11%	WT	WT
GBM-20041	62	GBM	IV	Yes	10	Alive	21	8	3	0.80%	WT	WT
GBM-10333	50	GBM	IV	Yes	NA	Alive	28	7	3	4.87%	WT	WT
GBM-20051	NA	GBM	IV	Yes	2	Alive	11	1	0	0.07%	WT	WT

Table S4. Mutational profile for all HGGs in Yale cohort (53 adults and 2 pediatric cases).
The CNVs and protein altering SNP/INDEL status for the genes that are frequently altered in GBMs are listed. (MUT= SNPs and INDELs with the amino acid change, WT= Wildtype, AMPLIFICATION, DELETION =CNV events.)

	Frequently Altered Genes in GBMs									
ID	TP53	NF1	EGFR	CDKN2A	PDGFRA	PTEN	MDM2	CDK4	PIK3R1	PIK3CA
GBM-60004	MUT:p.306R/*, MUT:p.273R/C	MUT:p.192R/*	WT	WT	MUT:p.987A/V	WT	WT	WT	WT	MUT:p.957T/I
GBM-60003	MUT:p.283R/C	MUT:p.119H/P,MUT:p.2151L/P,MUT:p.2214E/D	WT	WT	WT	MUT:p.166V/L	WT	WT	WT	WT
GBM-10457	WT	DELETION	AMPLIFICATION	WT	WT	DELETION	WT	WT	WT	WT
GBM-10461	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	DELETION	WT	WT	WT
GBM-10468	MUT:p.342R/*	MUT:p.,MUT:p.2718F/C	MUT:p.675R/Q	MUT:p.70F/C	MUT:p.398N/I,MUT:p.1042S/L	MUT:p.68Y/H	WT	WT	MUT:p.119E/K,MUT:p.429S/Y	WT
GBM-10333	WT	DELETION	AMPLIFICATION	AMPLIFICATION	WT	WT	WT	WT	WT	WT
GBM-10269	WT	AMPLIFICATION	AMPLIFICATION	DELETION	AMPLIFICATION	DELETION	AMPLIFICATION	AMPLIFICATION	WT	MUT:p.1043M/T
GBM-10474	WT	WT	AMPLIFICATION	DELETION	WT	WT	WT	WT	WT	WT
GBM-10352	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	WT	WT	WT	WT
GBM-60001	MUT:p.273R/C, MUT:p.213R/Q,MUT:p.124C/R	MUT:p.1362R/*,MUT:p.1676A/T	WT	WT	MUT:p.561V/A,MUT:p.937P/Q	WT	WT	MUT:p.267S/L,MUT:p.135R/C	MUT:p.412S/Y	MUT:p.88R/Q

GBM-20010	WT	WT	AMPLIFICATION	WT	WT	DELETION	AMPLIFICATION	AMPLIFICATION	WT	WT
GBM-20012	MUT:p.282R/W	WT	AMPLIFICATION	DELETION	WT	DELETION	WT	AMPLIFICATION	WT	MUT:p.1M/V
GBM-20013	WT	WT	AMPLIFICATION	DELETION	WT	WT	WT	WT	WT	WT
GBM-20016	WT	WT	AMPLIFICATION	DELETION	WT	WT	WT	WT	WT	WT
GBM-20017	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	WT	WT	WT	WT
GBM-10365	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	WT	WT	WT	WT
GBM-20028	DELETION	DELETION	WT	DELETION	DELETION	MUT:p.240-241	DELETION	DELETION	WT	WT
GBM-10449	MUT:p.267R/W,MUT:p.124C/R	AMPLIFICATION	AMPLIFICATION	WT	AMPLIFICATION	WT	AMPLIFICATION	WT	WT	WT
GBM-20030	WT	DELETION	WT	DELETION	DELETION	MUT:p.22D/E	DELETION	DELETION	WT	WT
GBM-20031	WT	DELETION	AMPLIFICATION	DELETION	WT	WT	WT	WT	WT	WT
GBM-20032	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	WT	WT	MUT:p.448-449KL/K	WT
GBM-20034	MUT:p.190P/L	WT	AMPLIFICATION	WT	WT	DELETION	WT	AMPLIFICATION	WT	WT
GBM-10265	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	WT	AMPLIFICATION	DELETION	DELETION
GBM-10132	AMPLIFICATION	DELETION	AMPLIFICATION	WT	WT	DELETION	WT	AMPLIFICATION	WT	DELETION
GBM-10355	WT	WT	AMPLIFICATION	DELETION	AMPLIFICATION	DELETION	WT	WT	WT	WT
GBM-10450	WT	DELETION	AMPLIFICATION	DELETION	DELETION	DELETION	DELETION	DELETION	DELETION	DELETION
GBM-20041	WT	WT	AMPLIFICATION	WT	WT	WT	WT	WT	WT	WT

GBM-20043	WT	DELETION	WT	DELETION	WT	WT	DELETION	AMPLIFICATION	WT	AMPLIFICATION
GBM-20044	AMPLIFICATION	AMPLIFICATION	AMPLIFICATION	DELETION	WT	WT	WT	AMPLIFICATION	WT	WT
GBM-20045	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	AMPLIFICATION	WT	WT	WT
GBM-20048	WT	WT	AMPLIFICATION	DELETION	WT	WT	DELETION	WT	MUT:p.457 Q/QEKSRE YDRLYEE	WT
GBM-10448	MUT:p.111L /P	WT	AMPLIFICATION	DELETION	WT	DELETION	WT	WT	WT	WT
GBM-20050	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	WT	WT	MUT:p.468E /-	WT
GBM-20051	MUT:p.151P /S	WT	AMPLIFICATION	WT	WT	WT	WT	WT	WT	WT
GBM-20006	MUT:p.	WT	WT	WT	WT	WT	WT	WT	WT	MUT:p.345 N/S
GBM-30107	WT	WT	WT	DELETION	WT	WT	WT	WT	WT	WT
GBM-30109	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	WT	AMPLIFICATION	WT	WT
GBM-30111	DELETION	WT	WT	DELETION	WT	DELETION	WT	WT	WT	WT
GBM-30118	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	AMPLIFICATION	AMPLIFICATION	WT	WT
GBM-30142	WT	WT	WT	DELETION	MUT:p.1041 S/N	DELETION	WT	WT	MUT:p.567 K/E	WT
GBM-30143	WT	WT	WT	DELETION	WT	DELETION	WT	WT	MUT:p.556-559YREI/Y	WT
GBM-30021	WT	WT	AMPLIFICATION	DELETION	AMPLIFICATION	DELETION	AMPLIFICATION	AMPLIFICATION	WT	WT
GBM-30239	WT	WT	AMPLIFICATION	WT	WT	DELETION	AMPLIFICATION	AMPLIFICATION	WT	AMPLIFICATION
GBM-30026	WT	WT	AMPLIFICATION	DELETION	AMPLIFICATION	DELETION	AMPLIFICATION	WT	WT	AMPLIFICATION
GBM-30028	AMPLIFICATION	AMPLIFICATION	AMPLIFICATION	DELETION	WT	DELETION	WT	WT	WT	WT

GBM-30029	WT	DELETION	WT	DELETION	WT	DELETION	WT	WT	WT	WT
GBM-30031	WT	WT	AMPLIFICATION	WT	WT	WT	AMPLIFICATION	AMPLIFICATION	MUT:p.404-405LI/L	WT
GBM-30056	MUT:p.248R/Q	AMPLIFICATION	AMPLIFICATION	DELETION	WT	DELETION	WT	AMPLIFICATION	AMPLIFICATION	WT
GBM-30059	WT	WT	AMPLIFICATION	DELETION	WT	DELETION	WT	WT	WT	WT
GBM-30092	AMPLIFICATION	WT	WT	WT	WT	DELETION	WT	WT	WT	WT
GBM-30099	MUT:p.216V/M	MUT:p.2539	WT	WT	WT	WT	WT	WT	WT	WT
GBM-10400	WT	WT	AMPLIFICATION	WT	WT	DELETION	AMPLIFICATION	AMPLIFICATION	WT	WT