Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease

Paugh et al

Supplementary Figure Legends

Supplementary Table Legends

Figure S1: *Heat map of copy number imbalances in pediatric HGG grouped by tumor grade*

Heat map showing segmentation analysis of normalized data from 500K SNP arrays to identify copy number gains (red) and losses (blue) in 78 pediatric HGG. Chromosome positions are indicated along the y axis and separated by a dotted line. Tumors are grouped by WHO tumor grade. Sample identifiers are listed across the top. GBM: glioblastoma, GBV: glioblastoma variant, AA: anaplastic astrocytoma, OT: anaplastic oligodendroglioma or anaplastic oligoastrocytoma, RV: Rare variant tumor types. Scale bar shows color gradient to indicate copy number. Diffuse intrinsic pontine gliomas are marked with triangles. Tumors showing relatively stable genomes are indicated with asterisks.

Figure S2: *PDGFRA amplifications*

(A) Chromosome 4 plots for tumors with focal *PDGFRA* amplification, for which gene expression data was also available. The $\log_2$ ratio for each SNP (y axis) is plotted according to their chromosomal position (x axis) for each tumor. (B) Representative FISH images to verify amplification of *PDGFRA* (red) compared to a chromosome 4 centromeric control (green).

Figure S3: *CDKN2A/B are the targets of the most common deletions* (A) Chromosome 9 plots for tumors with focal *CDKN2A/B* deletion, for which gene expression data was also available. The $\log_2$ ratio for each SNP (y axis) is plotted according to their chromosomal position (x axis) for each tumor. (B) Shortest region of overlap. The extent of deletion for each tumor is shown. The minimal common region of deletion for this group of tumors is highlighted with a gray bar. The location of individual genes within the minimal common region and 500KB upstream and downstream of this region are also shown. (C) Expression of all probes within the maximal region of deletion defined above is shown. The minimal common region of deletion is again shown with a gray bar. (D) Representative FISH images to verify homozygous deletion of *CDKN2A* (green) compared to a chromosome 9 centromeric control (red).

Figure S4: *Gain of chromosome 1q was associated with reduced survival in pediatric glioblastomas.* Survival analysis of pediatric glioblastomas with or without chromosome 1q gain. n=44, p=0.04. IR-induced glioblastomas were excluded from this analysis.

Figure S5: *A subset of tumors with stable genomes show evidence of tumor purity including IDH1 biallelic mutation, focal homozygous deletions, or LOH.*

(A). Exon 4 of *IDH1* was sequenced in 78 pediatric HGG and 11 LGG. Only 2 somatic missense mutations were found in a single tumor, HGG153. The representative electropherograms for both identified mutations, R49C and G97D are shown. The top panels show the wild type sequence of *IDH1* from the matched normal DNA and the bottom panels show the mutations present in the tumor sample. Subcloning and sequencing of PCR products from exon 4 revealed that the two mutations were present in trans, therefore representing biallelic mutation.

(B). Whole genome plots from HGG099 and HGG067 showing focal changes in *CDKN2A* in the context of an overall stable genome. Eight other samples also showed focal homozygous deletions which would not be detectable in samples with substantial normal tissue contamination. However, these deletions did not target known glioma suppressor genes, and we did not have matched normal DNA for these samples, so we cannot rule out the possibility of rare copy number variants in these additional cases. (C). View of paired LOH analysis using dChipSNP for 4 tumors lacking large-scale genomic imbalances for which matched normal DNA was available. Chromosome numbers are indicated along the y axis on the left. Left: Yellow indicates retention and blue indicates LOH. Right: The distribution of LOH probability scores ranging from 0 to 1. The red line indicates the threshold value of 0.71. HGG046 shows multiple areas of copy neutral LOH, HGG153 shows only one very focal area of loss also detected by LOH. HGG048 and HGG069 did not show detectable LOH.

Fig. S6: *Unsupervised hierarchical clustering (UHC) of pediatric HGG shows subgroups similar to those previously identified in adult HGG.* A) Dendrogram of the UHC using the top 1000 most variable probe sets selected using the median absolute deviation (MAD) scores, and heatmap featuring the top 350 signature probe sets of each subgroup. Three main subgroups were identified comprising 51%, 24.5% and 24.5% of tumors. (B). GSEA using adult HGG subgroup signature markers showed that the pediatric subgroups identified by UHC are highly similar to the subgroups previously identified in adult HGG. Plots of the running enrichment scores showed highly significant enrichment of the Prolif markers in HC1 (nominal p = 0.00196, FDR = 0.00467), PN markers in HC2 (nominal p = 0.00426, FDR = 0.002), and Mes markers in HC3 (nominal p = 0, FDR=0.00167). 33,928 genes were analyzed. Heatmaps for each GSEA for the

specific subgroup compared to the other two subgroups are shown below the running enrichment scores.

Table S1: *Clinicopathological data of sample cohort*

IR-induced indicates that the patient previously received cranial irradiation for a different cancer.

Table S2: *Real-time PCR primers and probes used to validate SNP copy number analysis*

Table S3: *Focal amplifications and focal deletions*

Table S4: *Global comparison of large-scale copy number gains and losses in pediatric HGG compared with adult glioblastoma data (downloaded from TCGA) for each chromosome arm.* p values are Fisher's exact test.

Sheet 1: Large scale gains

Sheet 2: Large scale losses

Table S5: *Up-regulated genes in each of the 3 subgroups from unsupervised hierarchical cluster analysis and over-represented GO terms among using p value cutoff of 0.01.* The top 350 up-regulated probe sets displayed in the heatmap are colored yellow.

Sheet 1: Up-regulated genes in HC1

Sheet 2: Up-regulated genes in HC2

Sheet 3: Up-regulated genes in HC3

Sheet 4: Over-represented biological process GO terms for the genes up-regulated in HC1

Sheet 5: Over-represented cellular process GO terms for the genes up-regulated in HC1 **Sheet 6**: Over-represented molecular function GO terms for the genes up-regulated in HC1

Sheet 7: Over-represented KEGG pathway for the genes up-regulated in HC1

Sheet 8: Over-represented biological process GO terms for the genes up-regulated in HC2

Sheet 9: Over-represented cellular process GO terms for the genes up-regulated in HC2

Sheet 10: Over-represented molecular function GO terms for the genes up-regulated in HC2

Sheet 11: Over-represented KEGG pathway for the genes up-regulated in HC2

Sheet 12: Over-represented biological process GO terms for the genes up-regulated in HC3

Sheet 13: Over-represented cellular process GO terms for the genes up-regulated in HC3

Sheet 14: Over-represented molecular function GO terms for the genes up-regulated in HC3

Sheet 15: Over-represented KEGG pathway for the genes up-regulated in HC3.

Table S6: *Differentially expressed genes between anaplastic astrocytomas (Grade III) and glioblastomas (grade IV).*

Sheet 1: Differentially expressed genes with $p<0.01$

Sheet 2: GO analysis of over-represented biological processes for genes upregulated in pediatric glioblastoma compared to anaplastic astrocytoma

Table S7: *Differentially expressed genes between tumors with stable genomes compared to all other pediatric HGG*

Sheet 1: Differentially expressed genes with p<0.01

Sheet 2: GO analysis of over-represented biological processes for genes downregulated in tumors with stable genomes

Sheet 3: GO analysis of over-represented cellular processes for genes downregulated in tumors with stable genome

Table S8: *Differentially expressed genes between tumors from infants (<3 years of age) compared to pediatric HGG from children >3 years of age*

Sheet 1: Genes overexpressed in tumors from infants

Sheet 2: GO analysis of over-represented biological processes for genes upregulated in tumors from infants

Sheet 3: GO analysis of over-represented molecular functions for genes upregulated in tumors from infants

Sheet 4: Over-represented KEGG pathway for genes upregulated in tumors from infants

Sheet 5: Genes downregulated in tumors from infants

Sheet 6: GO analysis of over-represented biological processes for genes downregulated in tumors from infants

Sheet 7: GO analysis of over-represented molecular function for genes downregulated in tumors from infants

Table S9: *Differentially expressed genes between IR-induced tumors and all other pediatric HGG*

Sheet 1: Gene list of differentially expressed genes in IR-induced tumors compared with other pediatric HGG

Sheet 2: GO analysis of over-represented biological processes for genes upregulated in IR-induced tumors

Sheet 3: GO analysis of over-represented molecular functions for genes upregulated in IR-induced tumors

Table S10: *Differentially expressed genes between pediatric and adult glioblastoma*

Sheet 1: Differentially expressed genes in pediatric glioblastoma compared to adult glioblastoma

Sheet 2: Over-represented GeneGO maps among the differentially expressed genes listed in Sheet 1

Sheet 3: Differentially expressed genes between PDGFRA amplified tumors and EGFR amplified tumors among TCGA data

Sheet 4*:* Ranking and signal-to-noise ratio of genes in TCGA PDGFRA set in pediatric glioblastoma. The leading edge genes identified by GSEA are marked "YES" under the Core Enrichment column.

Sheet 5: Top GeneGO networks built from the leading-edge genes of the TCGA PDGFRA gene set in pediatric glioblastoma.