

Gene Expression and Epigenetics Analysis of Pediatric High-Grade Gliomas

----- Xinyi Zhao

Introduction

Pediatric high-grade gliomas (pHGGs) are a leading cause of pediatric cancer death. Unlike the adult HGGs, there are no effective therapeutic intervention or treatment exists for this sort of cancers. Although the adult and pediatric HGGs are histologically similar, they are clinically, biologically, and molecularly distinct. And Diffuse intrinsic pontine glioma (DIPG), the untreatable, heterogeneous pHGG of the brainstem, is a highly aggressive cancer affects mostly young children and is uniformly fatal, which needs specific treatment approaches.

An important discovery on the molecular principles of pHGGs reveals that more than a third of patients harbor somatic mutations in H3F3A, the gene encoding the histone variant H3.3. There are two amino acid substitutions in H3.3 exiting in this gene 1) Mutation of Lysine 27 to methionine; 2) mutation of glycine 34 to arginine(G34R) or valine(G34V). The H3.3K27M mutant has been investigated more extensively because 1) it is more frequent in pHGGs patients; 2) the tri-methylation of K27 has a critical role in repressing gene transcription and compacting chromatin. Many studies indicated that cell lines with H3.3K27M mutant show hyperproliferation compared with their isogenic WT cell lines. In addition, H3F3A mutations were shown to result in global loss of trimethylation of lysine 27 of histone 3 in almost all cells of a tumor tissue, which will lead to entire gene expression changes. In the meanwhile, mutations on H3.1 and H3.2 were also uncovered in pHGGs although these kinds of mutation take only a small proportion of this sort of cancers. So, identification of the epigenetic features and gene expressional profiles of different subgroups of pHGGs by the mutation types of Histone H3 will provide important information on targeted therapy.

Data Resource

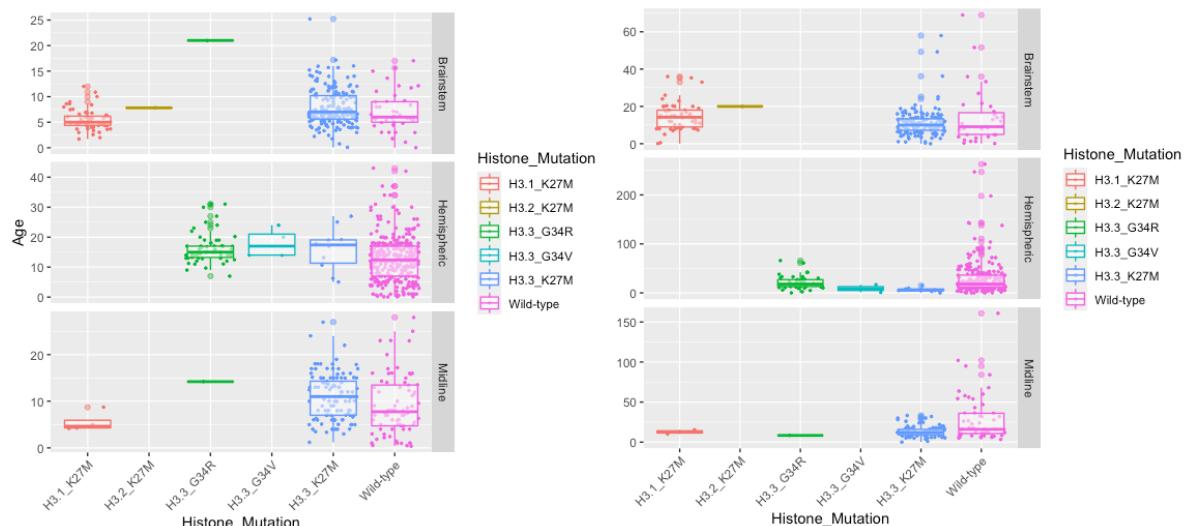
Most of the datasets were obtained from TCGA, PedCBioPortal, Xena, NCBI GEO (Gene Expression Omnibus) and some supplemental data from PubMed.

Main datasets: GSE26576, GSE50021, GSE19578, GSE34824, GSE36245, GSE50022, GSE36278, GSE55712
E-TABM-857, E-TABM-1107

Results

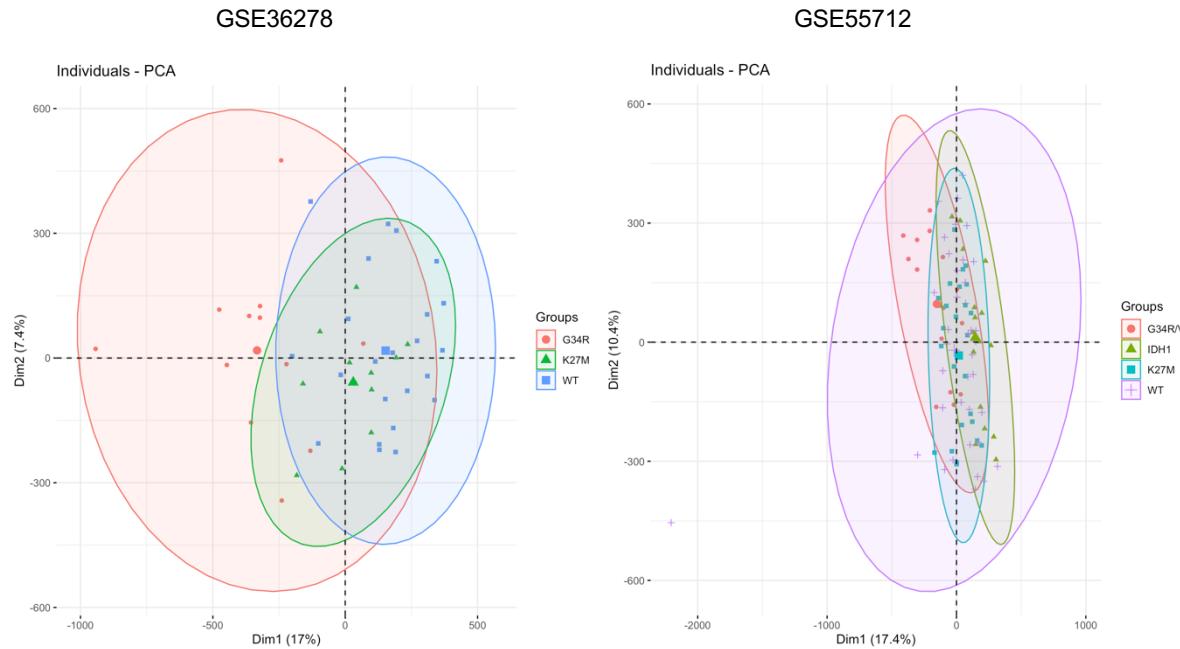
Molecular subgroups of pHGGs and the anatomical location, age at diagnosis and overall survival

Based on the boxplot of different types of Histone H3 mutants, wildtype Histone H3 tumors were distributed mainly in hemisphere and a small population were in brain stem and midline; H3.3G34R/V tumors were almost located to the cerebral hemisphere; and H3.1/H3.3K27M were predominantly restricted in brainstem and midline. The average age at diagnosis of H3.1/H3.3K27M mutant are obviously lower than H3.3G34R/V mutant, and represents the younger age group. However, there were not significant difference of overall survival among these H3 mutation subgroups.

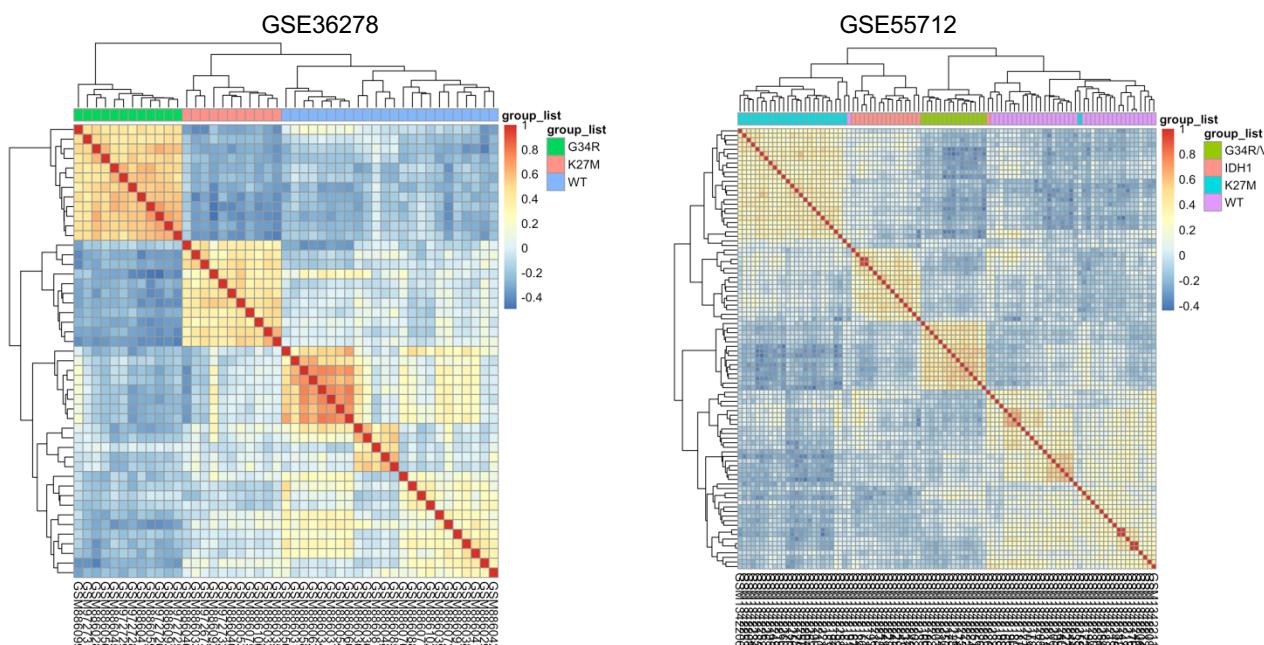


Methylation-based subgroup-Specific Alterations of pHGGs

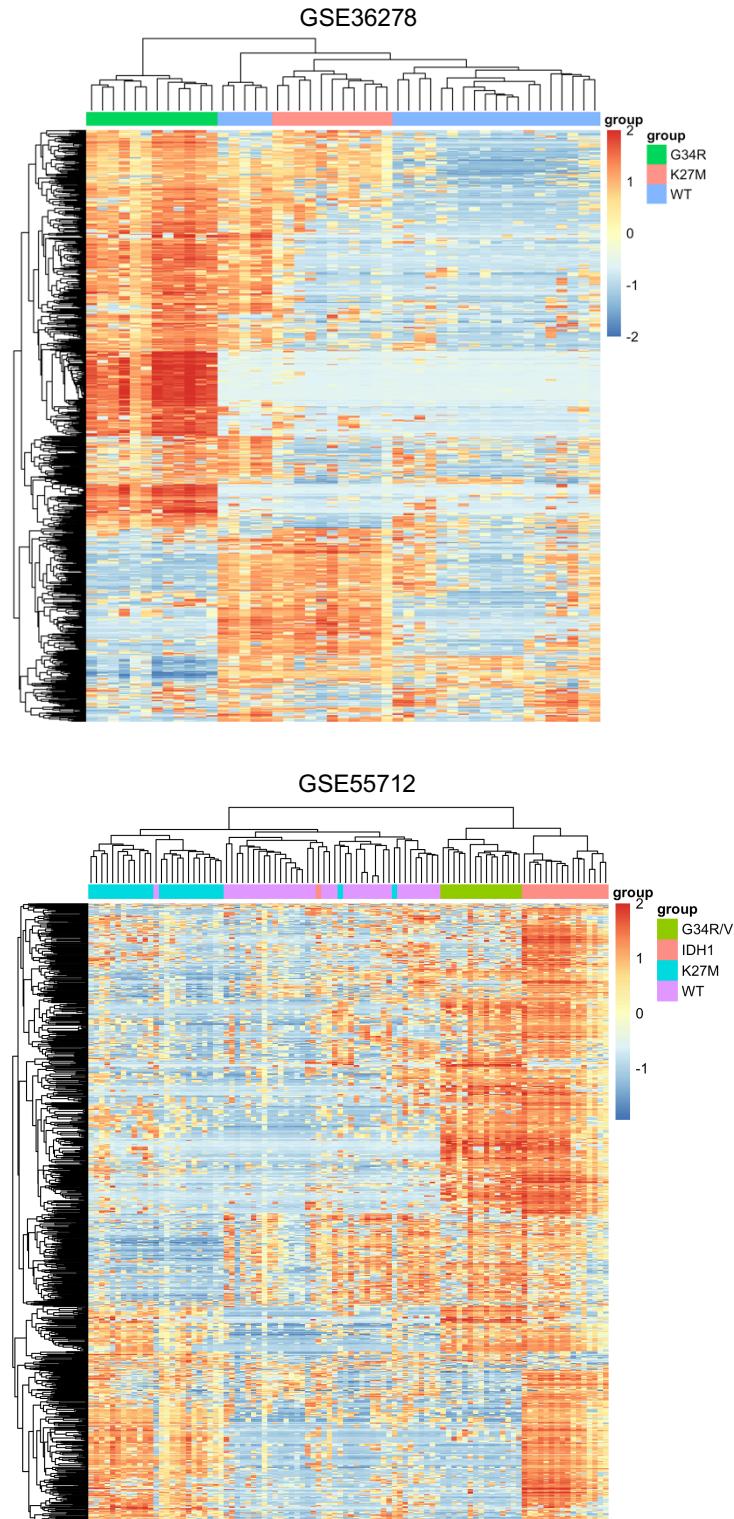
The main datasets come from GSE36278 and GSE55712 (Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)), both the datasets include subclass information of Histone H3 mutants. First, all the adult samples were removed from these datasets and the correlation of the Histone H3 mutant subgroups were analyzed by PCA and correlation plots (the samples of normal tissues are removed from all the datasets in this study to make the pure comparison among the subgroups of tumor tissues):



Although the subgroups by PCA plot are not separated clearly, the correlation analysis shows that samples with different Histone H3 mutant status are separated very well. It seems that GSE36278 datasets provided more explicit group information because 1) the patient samples of GSE36278 were all collected from GBM which made the datasets more consistent, but the GSE55712 patient samples include different types of glioma (cHGA-HGG, GLIOMA, mHGA), which makes the classification more complex; 2) GSE55712 included another mutant type, IDH1, that led to the grouping different from GSE36278. Thus, filtering the data into small consistent groups will be done for the further study.

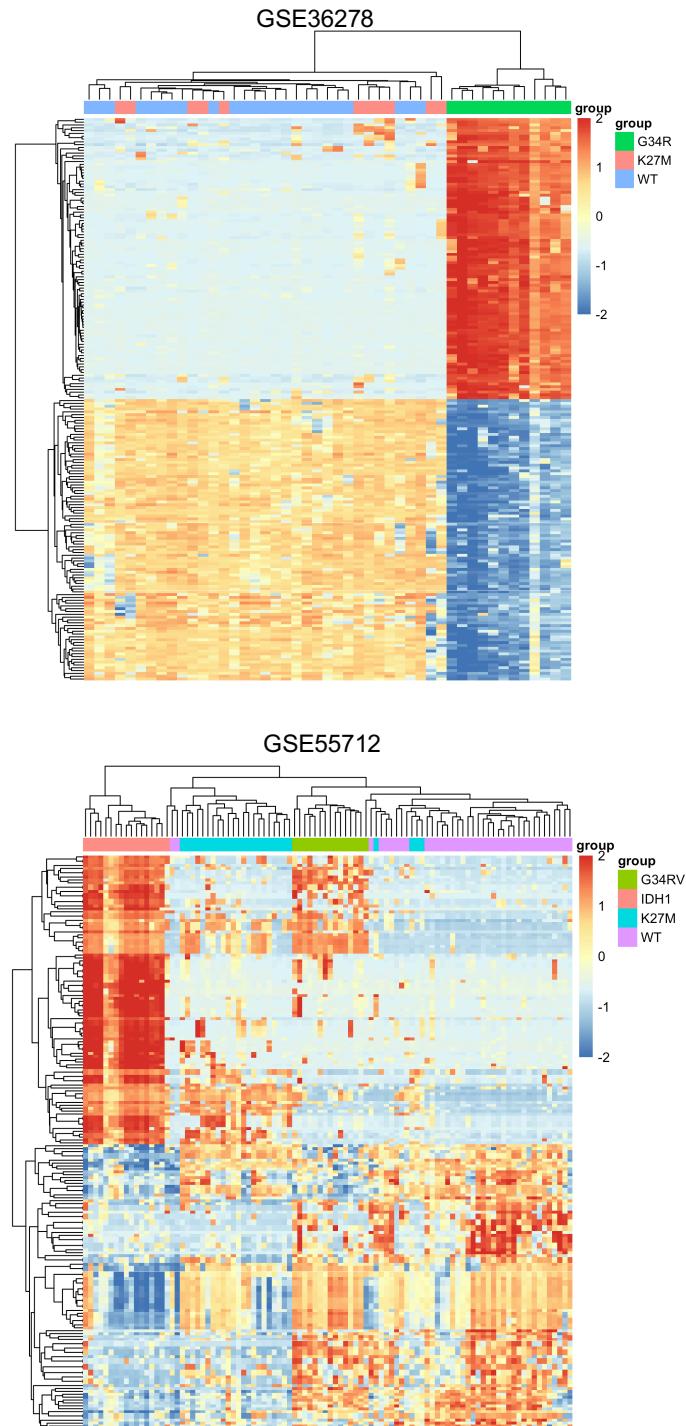


The heatmaps with the raw methylation values confirmed that the classifications are clear, and the preliminary differences among these subgroups are consistent between the two datasets. The G34R/V group shows higher global methylation level than the K27M and WT groups, and the IDH1 displays more obvious methylation than other groups.

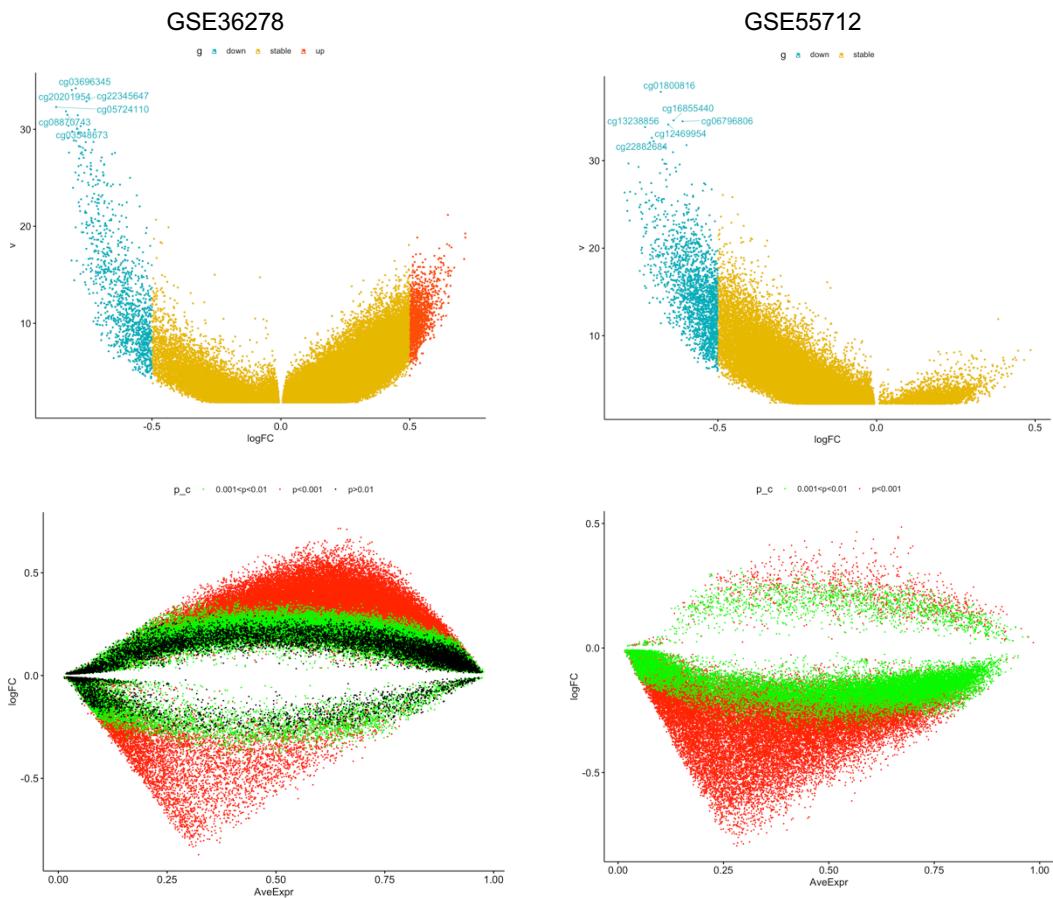


Differences analysis and screening the differentially expressed probes:

The algorithm used in this analysis is based on Bioconductor, “champ” package. The scale heatmap results are consistent with the distribution of raw data heatmaps, and the scale heatmaps indicate two distinct clusters of all these probes, the methylation increased, and methylation decreased clusters based on the methylation levels of G34R/V group or IDH1 group, these groups represent different methylation conditions from other groups. Compared with G34R/V or IDH1 groups, the H3.3K27M group and WT group show close methylation status. Because Up to 90% of cases of DIPG harbor a lysine to methionine (K to M) point mutation in histone 3 (H3) at amino acid 27 (H3K27M), it is necessary to figure out the global methylation difference between H3.3K27M and WT groups in more details. This part of work will be described in the next part.

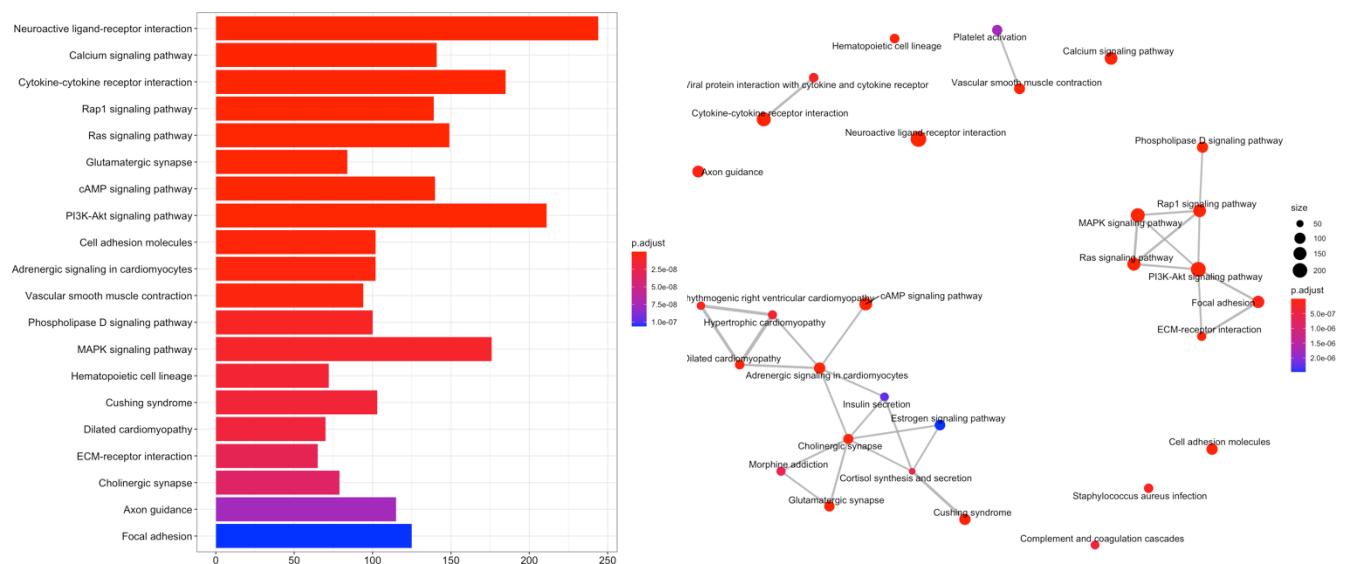


The increased and decreased probes:



The volcano and MA maps indicate the population of the probes with increased or decreased methylation values, which reveals that the population of methylation level decreased probes are much more than the methylation increased probes.

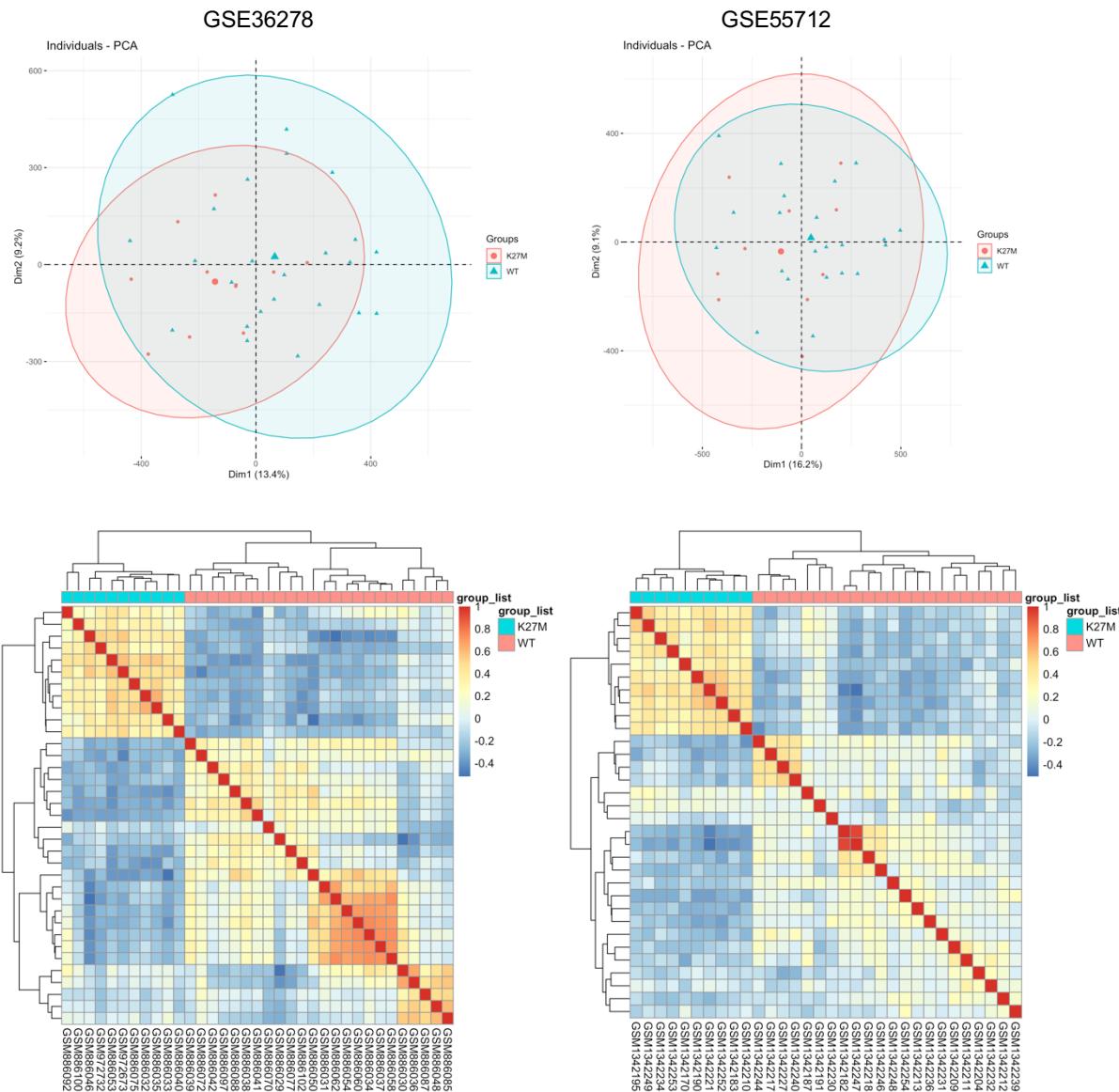
Differentially expressed genes and the corresponding signaling pathways:



By KEGG and GO analysis, the signaling pathways that show significant differences among groups include neuroactive ligand-receptor interaction, PI3K-Akt, Cytokine-cytokine receptor interaction, Ras signaling pathway, cAMP signaling pathway, calcium signaling pathway, and Rap1 signaling pathway, etc., which are highly involved in cancer development.

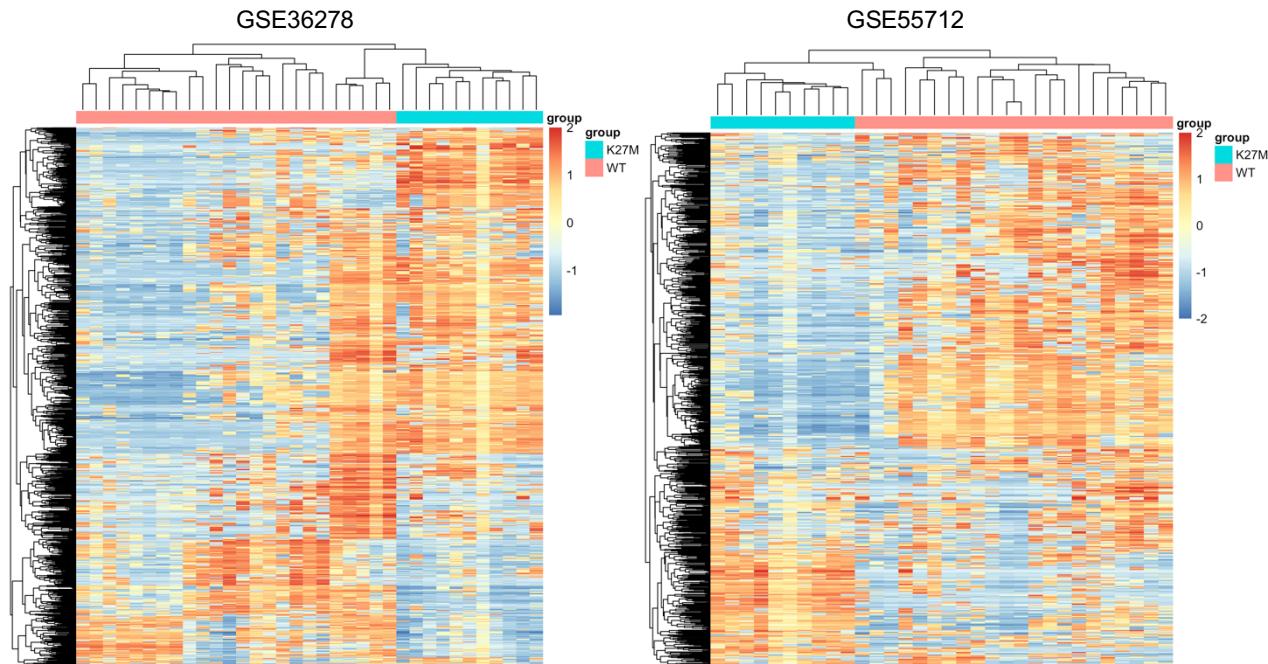
Methylation-based Alterations in H3.3K27M mutant of DIPG

In DIPG patients H3.3K27M mutations occur in 65% of patients and tumors localize throughout the midline and pons, so it is necessary to specifically analyze the DNA methylation status in H3.3K27M mutations. The main datasets come from GSE36278 and GSE55712, and only the types of H3.3K27M and WT are kept in this study.
PCA and correlation analysis:



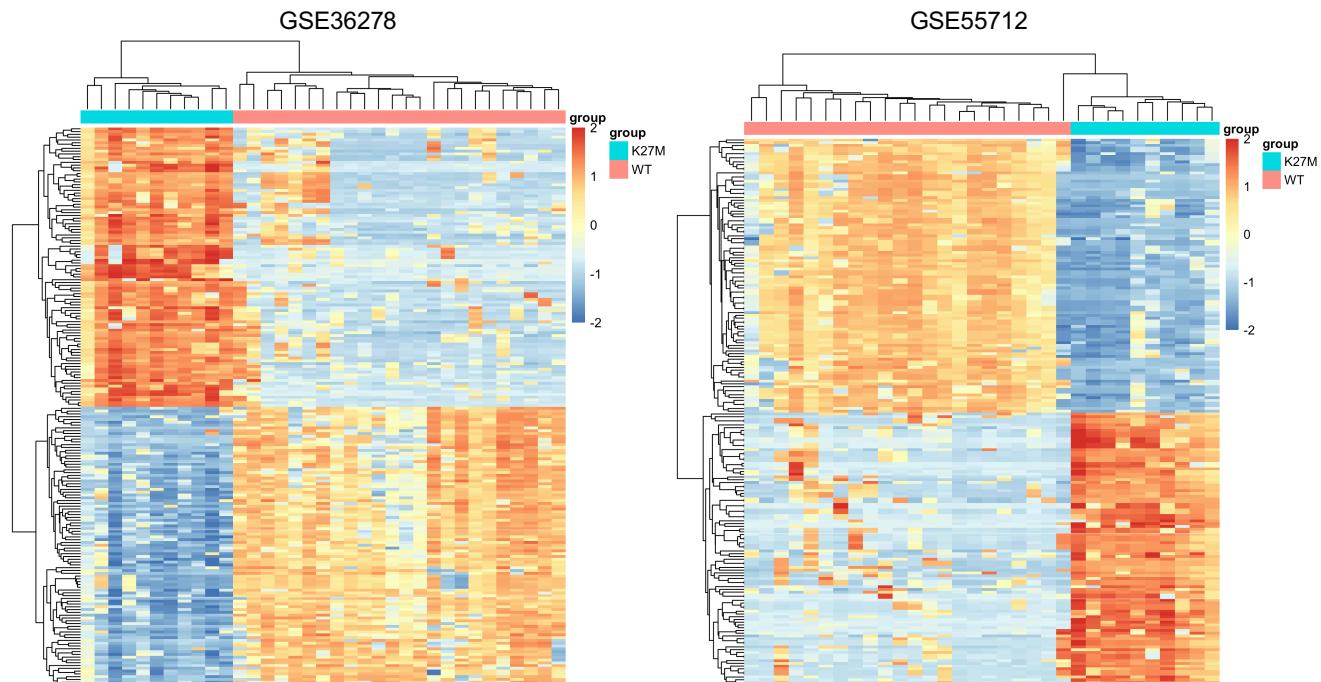
The PCA plot still cannot clearly distinguish these two groups, but the correlation maps show kind of better result and the K3.3K27M mutant can be separated from WT samples, although this kind of difference is not so dramatic as the difference between G34R/V to K27M or WT.

The heatmaps with raw methylation data also indicate obvious difference between H3.3K27M and WT groups but we still cannot figure out which group has higher entire DNA methylation level.

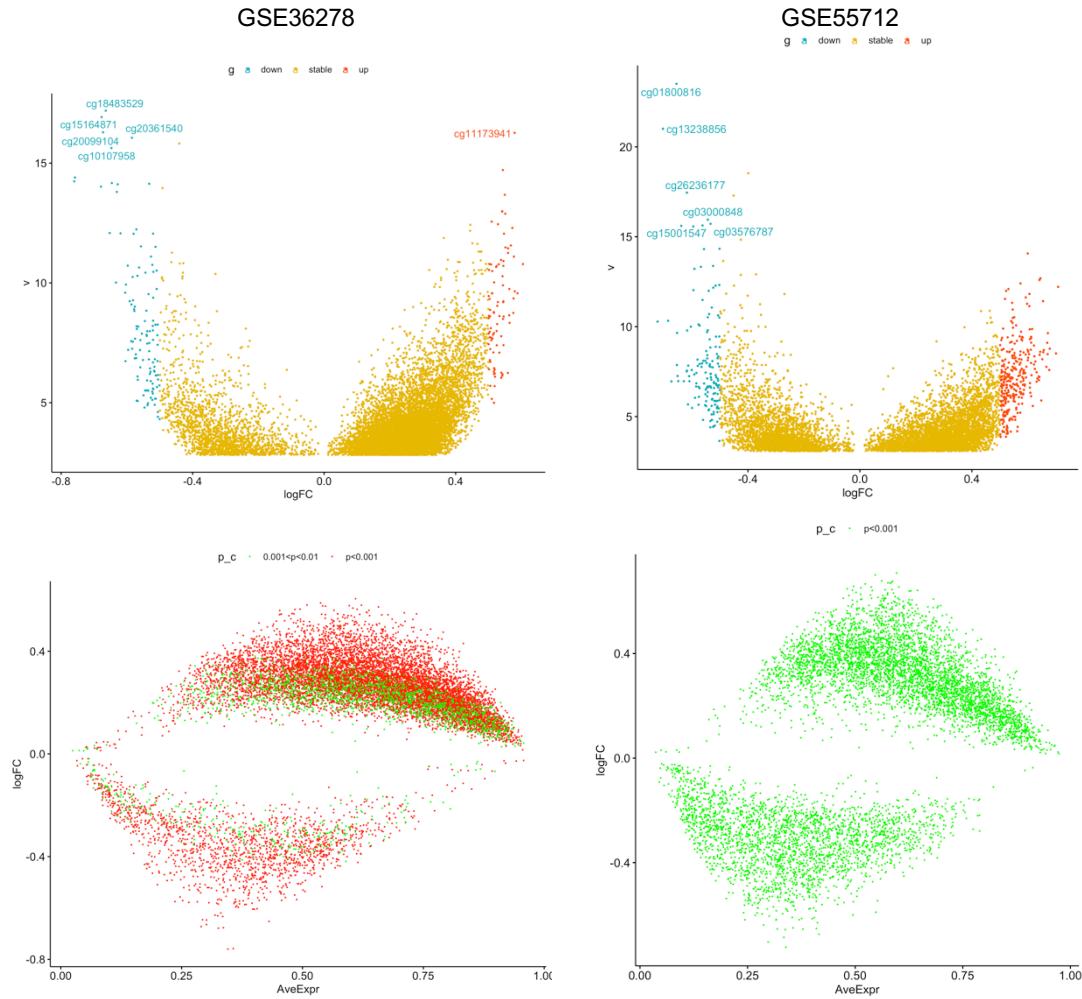


Difference analysis and screening the differentially expressed probes:

The methylation status difference analysis has been done by “champ” method on the two groups, and these two datasets reveal quite consistent results. The scale heatmaps show that the probes can be separated into two clusters for each group, methylation increased, and methylation decreased clusters. The H3.3K27M and WT groups have quite different methylation levels for each cluster. And generally, the entire methylation levels in H3.3K27M groups are higher than the WT group.

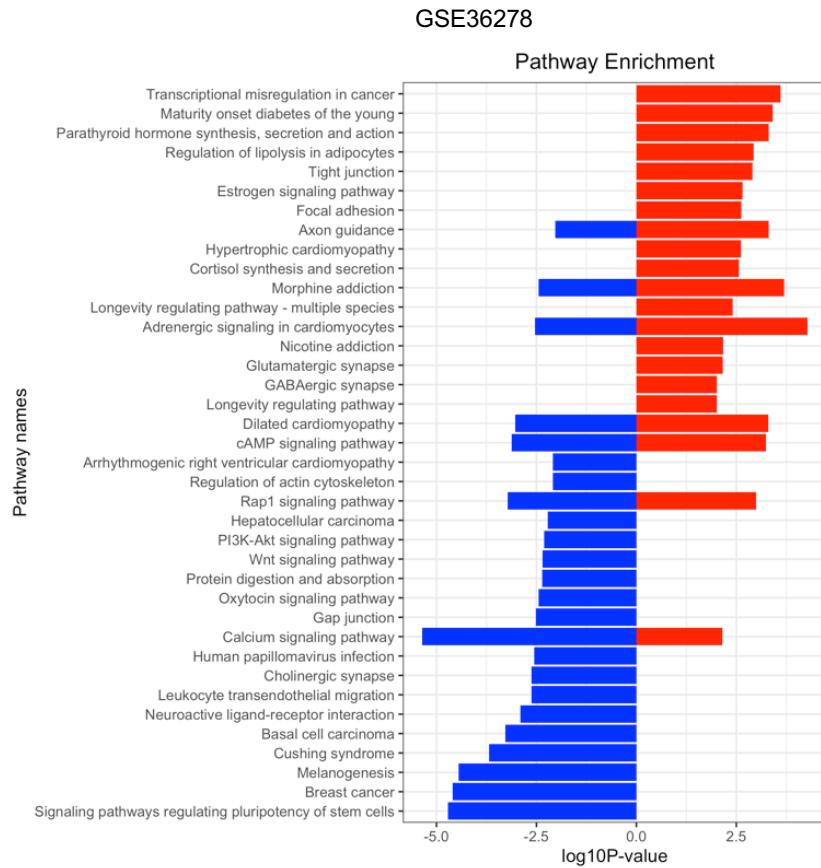


The volcano maps and the MA plots further confirm these results. Although there are both methylation increased genes and methylation decreased genes in H3.3K27M group compared with the WT group, there are more methylation increased genes in K27M group. This difference will directly influence the gene expression profile. This conclusion is kind of different with some previous report, but the related signaling pathway analysis will provide more detailed information for clinical research.



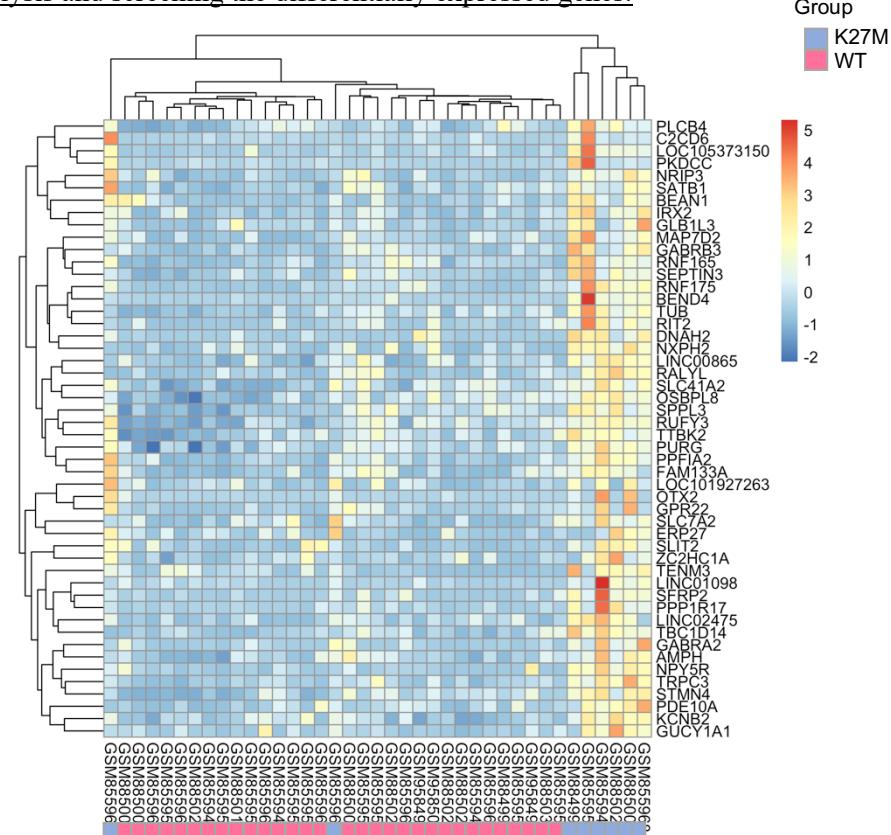
Differentially expressed genes and the corresponding signaling pathways:

By KEGG and GO analysis, the signaling pathways that show significant differences in H3.3K27M samples include: Transcriptional misregulation in cancer, cAMP signaling pathway, Rap1 signaling pathway, PI3K-Akt signaling, Wnt signaling, Neuroactive ligand-receptor interaction, etc., which have high similarity with the changed signaling by the analysis among the entire Histone mutation groups. How these signaling influences brain glioma development needs further investigation.

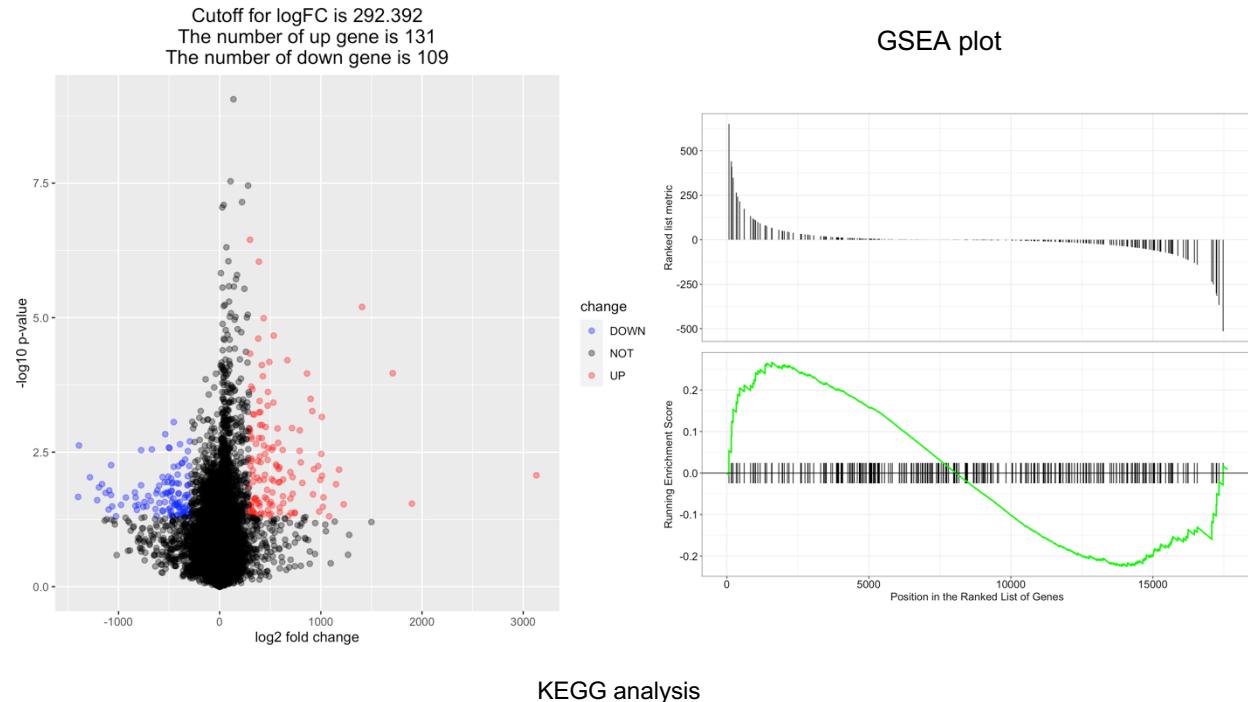


Gene expressional alterations in pHGGs (mRNA Microarray: GSE34824 & GSE36245)

Difference analysis and screening the differentially expressed genes:



The heatmap represents the top 50 differentially expressed gene between H3.3K27M and WT samples and the volcano map indicate the expressional differences of those genes. Although the fold change values of the differentially expressed genes are not quite dramatic between the two groups by volcano map and the Enrichment Scores are not high in GSEA plot, the K27M and WT groups can definitely be separated by heatmap. It means that their expressional profiles have essential differences.



KEGG analysis

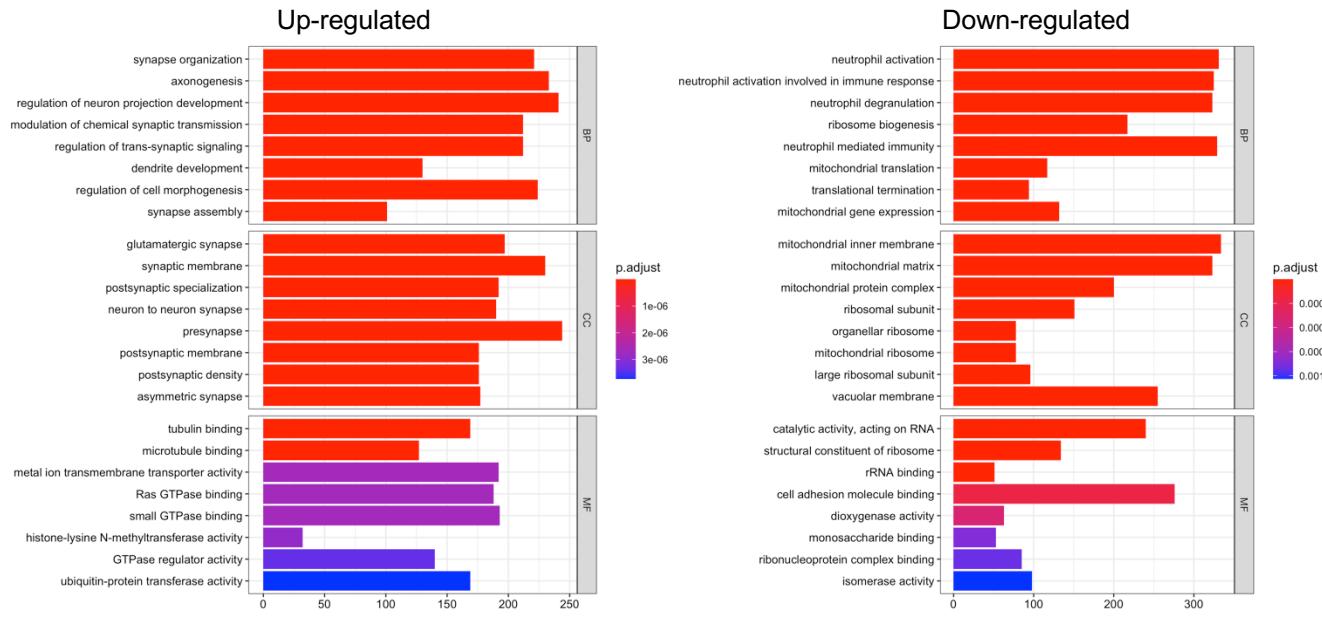
ID	Description	GeneRatio	BgRatio	pvalue	p.adjust
hsa05132	Salmonella infection	205/6167	213/8048	1.005458e-15	3.247631e13
hsa04360	Axon guidance	172/6167	181/8048	1.031520e-11	1.531817e-09
hsa05165	Human papillomavirus infection	300/6167	331/8048	1.422740e-11	1.531817e-09
hsa04140	Autophagy - animal	132/6167	137/8048	1.256636e-10	1.014733e-08
hsa04510.	Focal adhesion	187/6167	201/8048	3.942910e-10	2.547120e-08
hsa05014	Amyotrophic lateral sclerosis	324/6167	364/8048	5.662534e-10	3.048331e-08

Differentially expressed genes and the corresponding signaling pathways:

GO analysis explored the upregulated and downregulated genes in H3.3K27M mutant samples. There are multiple upregulated signaling pathways involved in neuron development and synapse transmission, including synapse organization, axonogenesis, regulation of neuron projection development, modulation of chemical synaptic transmission, regulation of trans-synaptic signaling, synapse assembly, etc. In addition, the small GTPases and Ras GTPase subfamily signaling are also overexpressed in K27M mutant (eg. Ras GTPase binding, small GTPase binding, GTPase regulator activity), which turn on genes involved in cell growth, differentiation and survival. Ras inhibitors are being studied as a treatment for cancer and other diseases with Ras overexpression.

The down-regulated signaling pathways mainly include neutrophil related factors, such as neutrophil activation involved in immune response, neutrophil degranulation and neutrophil mediated immunity, as well as mitochondria related signal such as mitochondrial protein complex, mitochondrial matrix and mitochondrial gene expression. The Neutrophils (also known as neutrocytes or heterophils) are the most abundant type of granulocytes and make up 40% to 70% of all white blood cells in humans. Lacking neutrophils related gene expression will influence the immune response and increase the susceptibility to cancer. Mitochondria generate most of the cell's supply of adenosine triphosphate (ATP), used as a source of chemical energy. Downregulation of mitochondria related genes will lead to the cell energy supply defect.

The heatplot was created based on the KEGG and enrichment analysis, it reveals the signaling pathways that differently expressed in K27M mutant samples and provide additional information for understanding the molecular mechanism of this cancer. Some of the signaling pathways are highly involved in cancer development such as Wnt signaling, Rap1 signaling pathway, MAPK signaling pathway, ErbB signaling pathway, cell cycle, cAMP signaling pathway, etc.

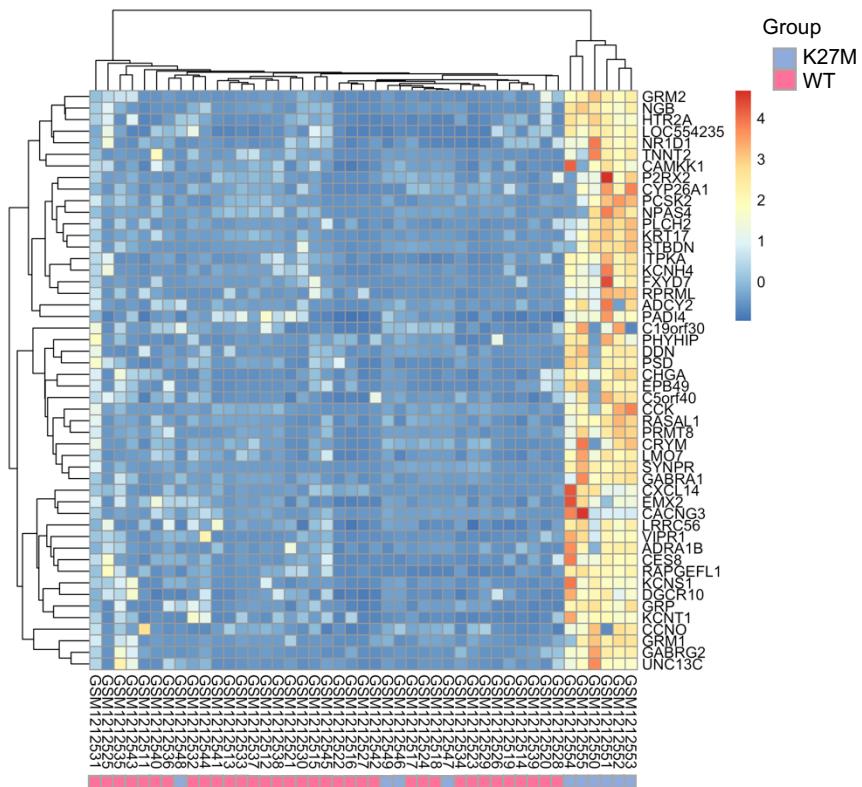


The mechanisms of these pathways on pHGGs tumorigenesis need to be further investigated to explore the effective treatment strategies.

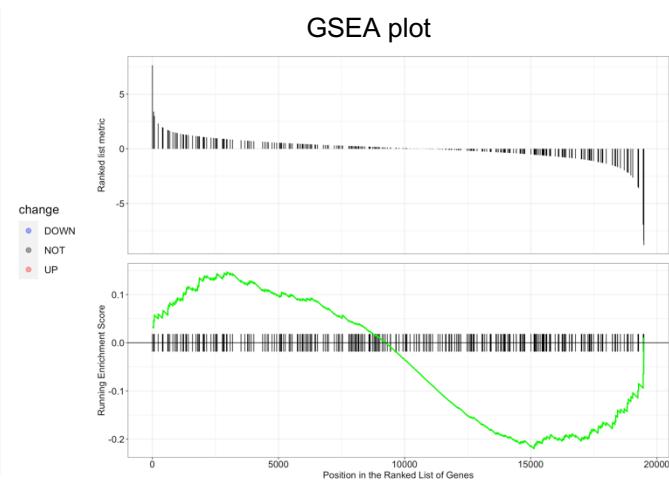
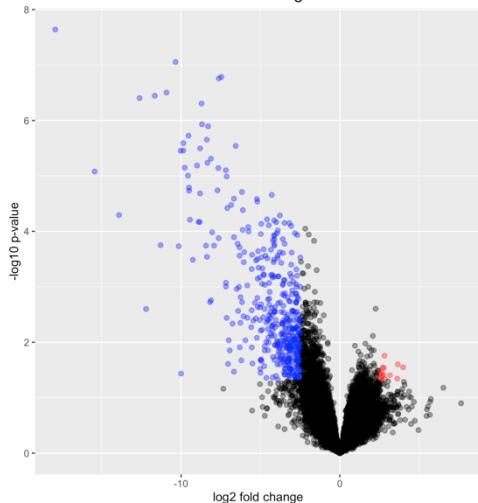
Gene expressional alterations in DIPG (mRNA Microarray of GSE50021)

Difference analysis and screening the differentially expressed genes:

The heatmap represents the top 50 differentially expressed genes between H3.3K27M and WT samples of DIPG patients. The volcano map shows the expressional differences of those genes. Different from the heatmap of pHGGs result, the heatmap for DIPG patients shows clearer separation between the two groups and the H3.3K27M group has obvious gene downregulation in most cases. This result has been confirmed by the volcano map because the number of down genes is much more than the number of up genes. Although the GSEA plot has not good enrichment score, the volcano map indicates higher logFC value and good separation of the up and down gene populations.



Cutoff for logFC is 2.461
The number of up gene is 15
The number of down gene is 431

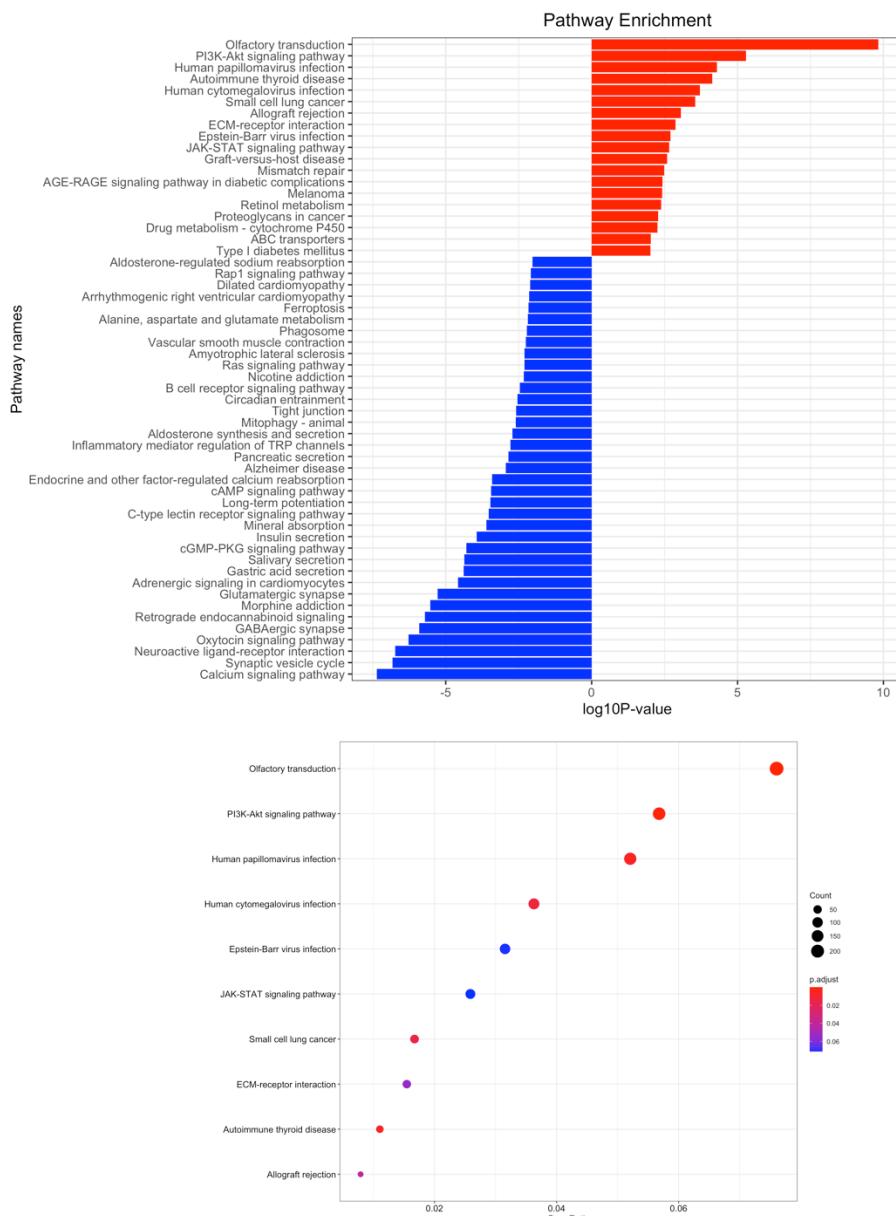


KEGG analysis

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust
hsa04080	Neuroactive ligand-receptor interaction	335/7258	341/8048	1.193150e-09	3.853875e-07
hsa05165	Human papillomavirus infection	325/7258	331/8048	2.927864e-09	4.728500e-07
hsa04151	PI3K-Akt signaling pathway	346/7258	354/8048	1.083442e-08	1.166506e-06
hsa05152	Tuberculosis	179/7258	180/8048	1.416158e-07	1.074193e-05
hsa04261	Adrenergic signaling in cardiomyocytes	149/7258	149/8048	1.772578e-07	1.074193e-05
hsa04020	Calcium signaling pathway	199/7258	201/8048	1.995405e-07	1.074193e-05

Differentially expressed genes and the corresponding signaling pathways:

The KEGG enrichment analysis discovered series of upregulated and down-regulated genes in H3.3K27M mutant, such as PI3K-Akt signaling pathway, JAK-STAT signaling pathway, Rap1 signaling pathway, ECM-receptor interaction, cAMP signaling pathway, etc. They are partially overlap with the differentially expressed genes in other kind of pHGGs, but there are many signaling pathways that are DIPG H27M mutant specific.

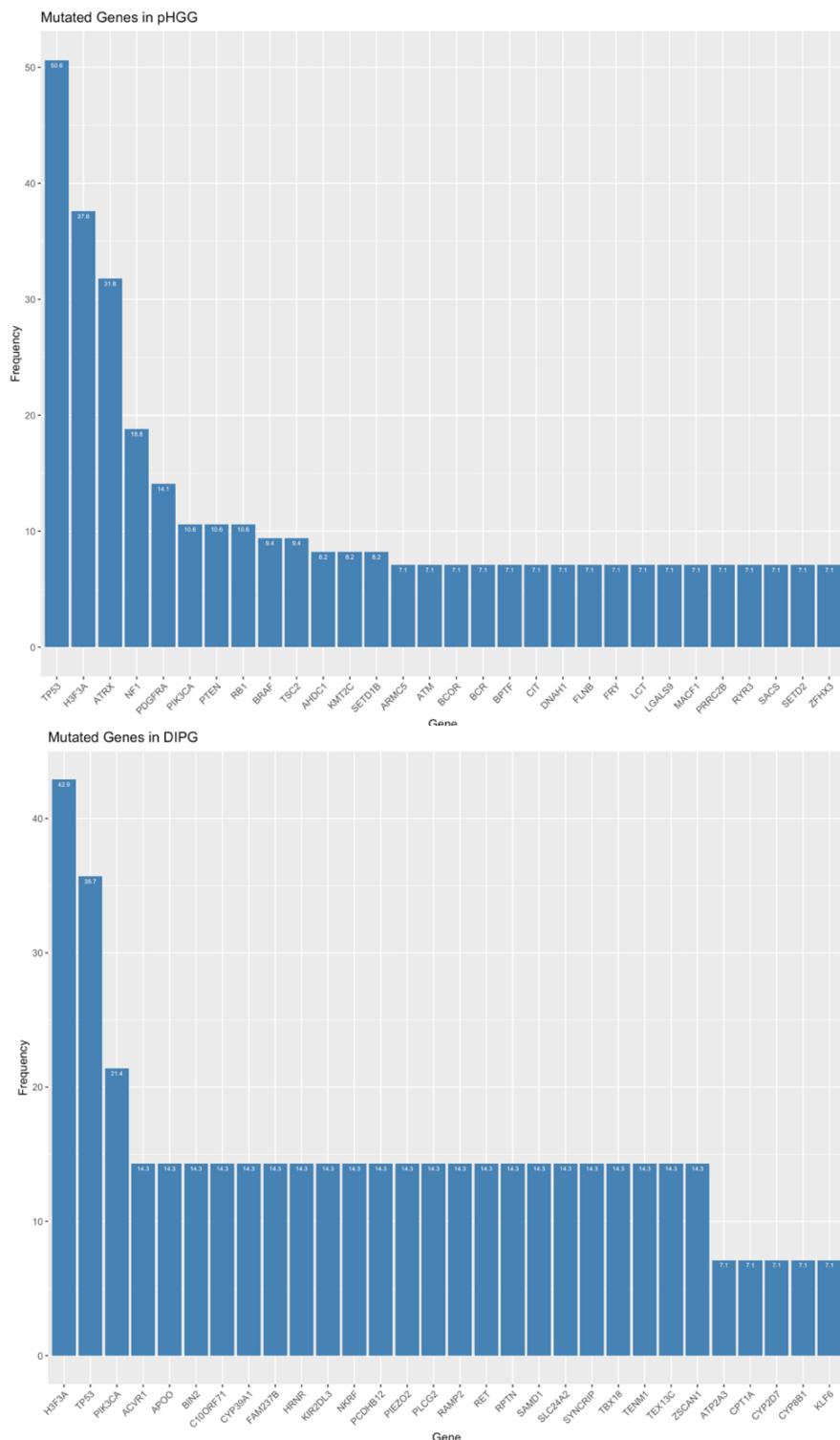


It seems that the gene expressional profiles in DIPG K27M mutant is kind of different from other kind of pHGGs. It's specific features suggest distinct treatment and the creation of new medication.

Gene mutation analysis in pHGGs/DIPG

Data resource: PedCBioPortal

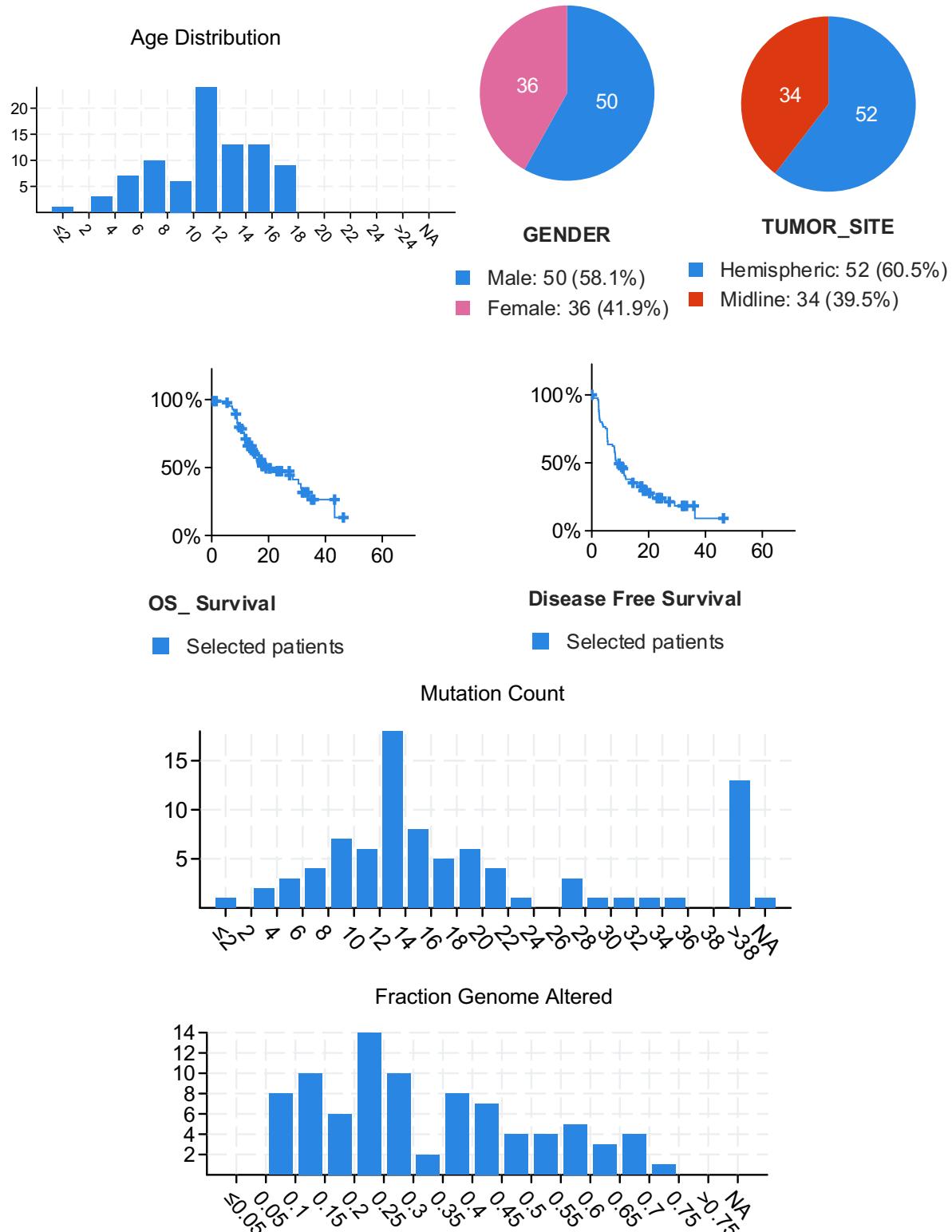
Based on the Paediatric High Grade Glioma datasets from “HERBY Clinical Trial, Cancer Cell 2018”, “ICR London, Cancer Cell 2017” and “CBTTC, Provisional”, as well as DIPG dataset from “Diffuse Intrinsic Pontine Glioma (CBTTC, Provisional)”, there are totally 10091 mutated genes in pHGGs and 197 mutated genes specifically in DIPG. The top 30 mutated genes based on the frequency (%) are listed below:



There are a certain group of genes that are common in these two datasets, such as H3F3A, TP53, PI3K3CA, etc., which are highly related to carcinogenesis. And there are also many DIPG specific mutations such as ACVR1, which has been reported by some study.

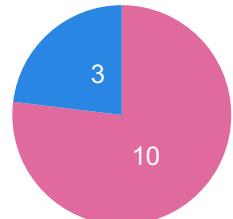
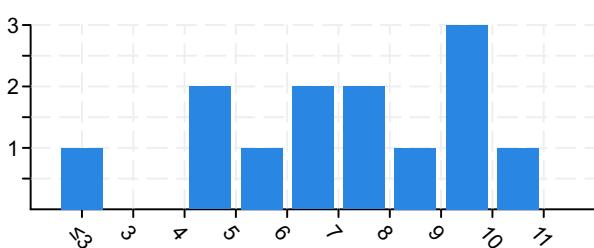
Supplemental information analyzed by PedcBioPortal

Database: pHGG (Paediatric High Grade Glioma datasets from “HERBY Clinical Trial, Cancer Cell 2018”, “ICR London, Cancer Cell 2017” and “CBTTC, Provisional”)



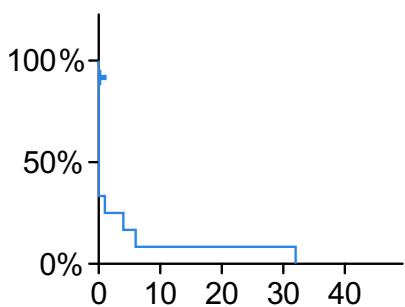
Database: DIPG (Diffuse Intrinsic Pontine Glioma (CBTTC, Provisional)")

Age Distribution



GENDER

- Female: 10 (76.9%)
- Male: 3 (23.1%)

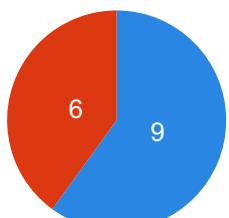


OS_Survival

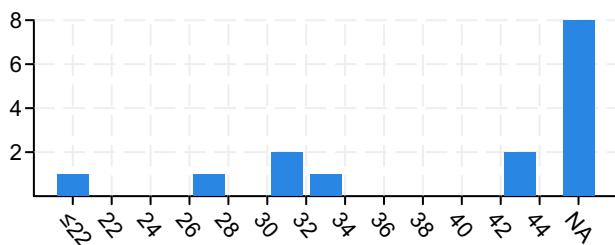
All patients

- Brain Stem- Pons: 12 (92.3%)
- Brain Stem- Midbrain/Tectum: 1 (7.7%)

TUMOR_SITE



Mutation Count



TUMOR_TYPE

- Progressive: 9 (60.0%)
- Initial CNS Tumor: 6 (40.0%)

TUMOR_TYPE