The Landscape of Structural Variants Within Pediatric High Grade Glioma

Table of Contents

Abstract

Pediatric high grade glioma is class of aggressive pediatric brain tumors with no current effective means of treatment. Unlike other pediatric tumors that often are characterized by specific, high frequency driver alterations in an otherwise stable genomic background, previous genomic studies have shown pediatric high grade glioma to have a complex and heterogeneous genomic profile, with high rates of structural rearrangement and somatic point mutations, in addition to alteration in chromatin modifying genes that suggest profound importance of somatic epigenetic changes. To further elucidate the spectrum of somatic events that lead to this devastating disease, we present a complete genomic characterization of the largest cohort of pediatric high grade gliomas, including the largest cohort of Diffuse Intrinsic Pontine Gliomas (DIPG), subjected to Whole genome sequencing (WGS) to date. In addition to confirming previous findings, our numerical power and deep sequencing approach has allows us to present the first comprehensive landscape of structural rearrangements in this disease. This includes elucidating new significantly mutated genes, as well as highlights new recurrent structural rearrangements involving the MYC and MYCN loci, as well as ID2, which is a key downstream effector of MYC signaling.

Acknowledgements

I want to thank Dr. Bandopadhayay and Dr. Shapira for allowing me to be involved in this project. Specifically, I'd like to thank Dr. Shapira and Noah Greenwald, who did the majority of the work on this project to date and I'm excited to continue to follow up on our initial exciting findings with them moving forward. In addition, Galen Gao, Ashton Berger, and Dimitry Livitz were instrumental in helping this project move forward, so I would like to thank them for all their help. Finally, an important acknowledgement to Dr. Mark Kieran, without whose foresight in pushing for and conducting the Molecularly determined treatment of DIPG trial (DIPG-BATs trial: NCT01182350) this project would possible.

I also want to acknowledge Dr. Beroukhim and Dr. Van Allen, my previous scientific mentors, who have helped me grow so much, both as a scientist, and as a person. I would especially like to thank Dr. Van Allen for being so flexible in allowing me to write about my new project with Dr. Bandopadhayay rather than my originally planned thesis project on germ-cell tumors. The later was a fantastic project and it was so much fun to work on it with Dr. Van Allen, but I was really excited to take the opportunity this thesis provides to discuss new results, so I thank Dr. Van Allen (and the committee) for allowing me that flexibility. I would also like to thank the committee for their dedication to medical education and scientific inquiry in taking the time to review and discuss this work with me.

Finally I would like to thank my partner Kate, for helping me through every step of the way.

Introduction

Pediatric High Grade Gliomas

Pediatric high grade gliomas (pHGG) are the most common malignant tumor of childhood(*1*) and their prognosis remains remarkably poor, with median overall survival just 12-15 months(*2*). Central nervous system tumors comprise the majority of solid tumors in pediatric populations 0-14, with an incidence of 5.5/100,000 in this population(*3*). However, the majority of these cases are classified as Non-Malignant, generally characterized by slow growth and lack of metastatic potential(*3*). Within malignant CNS neoplasms of childhood, 80% are classified as HGG (including glioblastoma)(*3*) which account for approximately 450 reported new cases a year in the United States(*3*). Unlike non-malignant classes of pediatric CNS tumors, pHGGs are extremely aggressive with 5 year survivals ranging from 20-33% depending on grade and surgical accessibility(*4*) and are the primary contributor to CNS malignancies being the most common cause of malignancy-related death in childhood within the United States(*3*). Pediatric HGG is a heterogenous class of tumors that is phenotypically distinct from adult glioma(*4*). Glioma is a term for tumors appearing to derive from glial cell lineages (Astrocyte, oligodendrocyte or ependymal cells), yet traditionally is used when describing astrocytomas or oligodendrogliomas(*4*). Gliomas histologically graded as Grade I and II are considered LGG, whereas Grade III or Grade IV are considered HGG. Within HGG, the most common histopathological categorizations for Grade III are anaplastic astrocytoma and anaplastic oligodendroglioma and for Grade IV are diffuse intrinsic pontine gliomas (DIPG) and glioblastoma. Like in other tumors, these histopathological categories have some predictive power on prognosis, with higher grades associated with poorer prognosis(*5*). While gliomas account for 56-70% of pediatric CNS tumors(*5*), the majority of these are classified as low-grade glioma (LGG) and unlike in adult glioma, pediatric LGG have very low rates of transformation to high grade (7%

as compared to up to 50% in adults)(*4*, *6*). And whereas the vast majority of adult gliomas occur supratentorially, pHGG can be localized throughout the CNS (*5*), with infratentorial tumors more common in younger children than in adolescents and adults(*5*). In particular, the vast majority of infratentorial pHGG are found within the brainstem(*7*), with 80% of these characterized as DIPG, which have a particularly dismal prognosis due to their aggressive nature and a location and distribution not amenable to surgical resection(*7*). While clinical management for pHGG outside the brainstem generally involves surgical resection with negative margins, the location and diffuse nature of DIPG makes limiting disease through surgical management untenable(*4*) and without any effective radiation or pharmaceutical options available, median survival for DIPGs is less than one year from time of diagnosis. Given that surgery is not a management option and the vast majority of tumors located within the brainstem are eventually diagnosed as DIPG, there has, in the past, been considerable controversy and vacillation about the need for invasive biopsies to confirm diagnosis for tumors in this location, with consensus, up until recently, to instead rely solely on imaging for diagnosis(*8*–*11*). Due to the lack of primary tumor samples for analysis, the molecular characterization, and therefore our understanding of development of these tumors has lagged behind other malignancies(*11*). However, due to recent advances in precision based therapies in other tumor classes through a better understanding of the genetic alterations that drive tumor development, a recent change in consensus has emerged, with biopsy for molecular characterization now increasingly recommended on a patient-by-patient basis, both for clinical trial enrollment as well as to improve understanding of the disease(*12*). Our group, led by Dr. Kieran, led the first multicenter clinical trial to incorporate upfront biopsy into treatment stratification, which has created an invaluable resource of pretreatment tumor samples.

Molecular characterization of pediatric glioma

Predictably, the phenotypic differences recognized between histopathologically and geographically distinct pediatric gliomas are driven by underlying differences in genetic alterations. Due to the scarcity in samples, our biological understanding of pediatric glioma has historically lagged behind the adult counterpart. However, over the last few years, increased rates of biopsies combined with the dramatic reduction of cost in next generation sequencing has rapidly advanced our characterization of the genetic (and epigenetic) changes frequently present in these diseases. Unlike in LGG, which tend to be characterized by few overall somatic alterations and characteristic alterations within the MAPK pathway (*13*, *14*), the pHGG genome tends to be relatively complex with significant differences between alterations found in DIPG versus non-DIPG pHGG, both of which are grossly distinct from their adult counterpart. In 2012, Wu et al(*15*) discovered frequent H3 histone mutations through Whole genome sequencing (WGS) of 7 pHGG and verified mutations in the gene *H3F3A* in 78% of DIPG and 22% of non-brainstem HGG. At around the same time, exome sequencing on 43 pediatric GBM uncovered these same mutations(*16*). Further studies have demonstrated that the H3.3 Lys27Met mutation is most characteristic midline locations (brainstem and thalamus) whereas the H3.3 Gly34Arg/Val mutation is more common in hemispheric tumors(*13*) and follow-up experiments(*17*) showed that these mutations lead to separate and distinct methylation profiles. The change in methylation associated with H3K27me3 is primarily characterized by global reduction of the repressive H3K27me3 mutation through inhibition of polycomb repressive complex 2 (PRC2)(*18*), though paradoxically, some regions of the genome seem to have increased levels of H3K27m3(*19*). Conversely, the Gly34Arg/Val mutation produces hypomethylation in subtelomeric regions through currently unclear mechanisms(*17*). The importance of epigenetic and genetic changes in subtelemoric regions in these tumors was further solidified when it was shown that 44% of pGBM contain mutations in the H3.3-*ATRX*-*DAXX* pathway(*20*). ATRX and DAXX are members of a transcription/chromatin remodeling complex that incorporates H3.3 into heterochromatin at telomeres and several transcription factor binding sites(*21*, *22*). These *H3F3A* G34R

mutations seem to drive *MYCN* expression(*23*), a gene that is also found focally amplified in some DIPG tumors. Mutations that are frequent in aHGG, including IDH1/2 mutations and inactivating mutations in p53, occur in pHGG but appear to be more rare(*20*). Importantly, *IDH1/2* mutations and Histone mutations are mutually exclusive across samples, with these rare *IDH1/2* mutations predominantly occurring in adolescents(*4*). Separately, Taylor et al(*24*) showed that activating mutations in *ACVR1* seem to occur specifically in DIPG, mutations that are also found in the germline of individuals with a congenital syndrome fibrodysplasia ossificans progressive, the pathophysiology of which is thought to be derived from constitutive activity of TGF-B(*25*). Most recently, WGS with concomitant RNA sequencing for fusion transcripts, on 53 pGBM found cases of rearrangements involving known cancer genes, including *FGFR3*, *NTRK2*, and *PIK3R2*(*26*). Significantly, they found 6 samples with fusions involving the gene *MET*, which encodes an oncogenic tyrosine kinase involved in the MAPK signaling cascade. However, other than *MET*, it was difficult to assess the functional significance of other rearrangements due to limited sample size from this study.

Chromosomal abnormalities, DNA copy number changes, and patterns of rearrangments

Adult Glioblastoma (aGBM) tumors are highly aneuploid and have long been characterized by frequent gain in chromosome 7 and loss of 10q(*27*), a pattern not commonly seen in pHGG(*28*, *29*). And while most pHGG contain a significant amount of aneuploidy, as many as 14% of patients do not seem to have any copy number abnormalities, at least as measured by array-based methods(*4*). However, there are a few known recurrent focal copy number changes and common, though not pervasive, patterns of chromosomal abnormalities have been described, including gains of 1q, and losses of 16q and 4q(*4*). Whereas the most frequent focal copy number amplification in aGBM is *EGFR*(*30*), *EGFR* amplifications occur relatively infrequently in pHGG. In fact, regions that are significantly recurrently altered in pHGG, including amplifications of *MYC*, *MYCN*, copy changes in cell cycle regulators (amplification of *CDK6*,

CCND2, loss of *CDKN2C*), and *KRAS* all occur at low frequency suggesting there may be many separate pathways towards tumorigenesis in this malignancy(*28*, *31*, *32*). One region that is subject to more focal amplification in pHGG than aHGG is a region on 4q12 containing known oncogenes *PDGFRA*, *KIT*, and *VEGFR2*(*4*). Due to its consistent overexpression in gene expression studies, many hypothesize *PDGFRA* is the predominant target of this amplicon, but that has yet to be shown conclusively, to some degree due to limited resolution in previous datasets. Finally, chromothripsis, which is a recently discovered phenomena involving the mechanistic "shattering" and reassembly of an entire chromosome during mitosis(*33*–*37*), seems to be a frequent occurrence in these tumors(*26*), though importantly, true chromothripsis can be difficult to identify(*38*) and these tumors have not yet been subjected to a rigorous, unbiased algorithm to infer these events.

Comparing DIPG with other pHGG, gains of chromosomes 2,8q and 9q and losses of 16q, 17p and 20p are more common in DIPG, while fewer DIPG tend to exhibit these "stable" genomes(*4*). Similarly, DIPG are characterized by increased rates of focal events, including amplification of the *PDGFRA* locus and increased copy number changes affecting the PI3K and Rb pathways(*39*), pathways also frequently altered in aHGG. This may suggest that the genesis of DIPG may share some similarities with aHGG when compared to other pHGG.

Patterns of genomic alteration as a window on the mechanistic causes of tumorigenesis

With the decreasing cost of WGS, large datasets both within and across many different tumor types are being generated at increasing rates(*40*). Sometimes derided as more expensive than exome sequencing, while providing no more information about driver gene status, by providing a complete picture of the somatic genetic alterations with the tumor cell, WGS is less confounded by selective pressures and can better elucidate the mechanistic processes that govern tumorigenesis(*41*–*43*). As these mechanistic processes can often be therapeutic targets(*44*–*46*), understanding exactly which processes are at play

during tumor evolution and how they shape the genome can be important to our understanding of cancer vulnerabilities, as well as potential mechanisms of tumor resistance to treatment. Recently, work by the International Cancer Genome Consortium has highlighted the patterns of somatic single nucleotide variation (SSNV) and structural rearrangements (SV) across large panels of, primarily adult, tumors(*43*, *47*). These papers find unique signatures of mutational processes, often associated with different specific deficits in DNA repair, environmental exposures, or malignancy classes(*42*, *43*). Of particular interest to us, while primary tumors of the Brain and CNS are the 8th most common malignancy among adults over 40 (incidence of ~40/100,000), they are the most common class of malignancy in children age 0-14 (~5.5/100,000 vs ~5.0/100,000 for leukemias). This suggests a fundamental process that makes the pediatric CNS more susceptible to tumorigenesis than other tissues in the body. Of interest, single cell sequencing of neural tissue of post-mortem, otherwise normal human embryos was recently shown to have an increased frequency of somatic mutations (*48*). The authors hypothesize that the mosaicism caused by this increased rate of somatic mutations, which seems specific to early neural development, may be important in normal neuronal differentiation by altering neuronal development. This provides an interesting hypothesis for the genesis of pediatric CNS tumors that has yet to be tested. Similarly, the processes and patterns of SV, which has been known to be an important driver of malignancy ever since the Philadelphia chromosome was described(*49*), is now being characterized in more detail than ever before, providing fascinating glimpses into the determinants of DNA fragility and reconstruction of the cancer genome. These covariants that predict DNA breakpoints include DNA elements such as Short and Long Interspersed nuclear elements (SINE and LINE), fragile sites, gene expression, replication timing, and histone post-translational modification(*43*). This new insight allows us to better determine regions of DNA breakpoints and SV across the genome that are significant due to an evolutionary advantage, as opposed to being mechanistically favored. Because of the importance of SV and fusion proteins in many pediatric tumors(*50*, *51*), including the

recent description of *MET* fusions in the largest, to date, collection of whole genome sequenced pGBM(*26*), suggests importance of SV in pHGG development.

Results

We performed a combined genomic analysis of 251 pediatric high-grade gliomas, including 158 Diffuse Intrinsic Pontine Gliomas (DIPG), 103 GBM (GBM). This included whole genome sequencing (WGS) on 178 tumors (93 DIPG, 70 GBM) and quantitative RNA-seq on 105 samples (45 DIPG, 47 C-GBM, 12 M-GBM), including a total of 61 unpublished tumors.

Purity/Ploidy

Using WGS to determine purity and ploidy of tumor cells within each sample, we found these primary tumors to be of much high purity and much lower rates of aneuploidy than samples previously derived from adult GBM. We applied the ABSOLUTE algorithm(*52*) to determine the purity and ploidy of samples within our dataset (Figure 1, Methods). Compared to other classes of primary tumor, and GBM specifically(*53*), our pHGG were relatively pure and, consistent with prior studies, most of them were close to diploid. ABSOLUTE also predicts whether a sample has undergone a whole genome doubling (WGD) event during the course of its evolution. We found no significant difference in the rate of WGD between DIPG and GBM (11.8% vs 20.8%, Fisher's exact test p=0.15). Without solving for tumor purity and ploidy, it can be difficult to distinguish samples with stable genomes from those with very low tumor purity. Additionally, copy neutral loss of heterozygosity is a frequent event in cancer, which can go undetected without solving for allelic copy ratios of WGS reads. By measuring allelic fraction of reads within our WGS data, we found two GBM samples in our dataset that had undergone almost genome-wide loss of heterozygosity, in addition to WGD (SFig 1), an event which has not previously been appreciated in this tumor type. We also found 11 diploid samples with extremely stable genomes, defined as <1% of genome altered (7 samples had no observed copy level changes). These stable genomes occurred in about equal proportions between DIPG and GBM samples (6.7% vs 6.25% respectively; FE test p=0.59). This represents a decrease in the rate of stable genomes compared to

other studies, which is likely due to the increased resolution copy number determination through WGS provides.

Signatures of distinct processes of mutation and structural rearrangement

Whole genome sequencing across our dataset provided an unparalleled assessment of background rates of somatic alterations, shedding new light on potential mechanisms of genetic instability within this disease (Figure 2). The vast majority of mutations detected via WGS do not have an associated effect on cellular fitness, in contrast to exome sequencing approaches. This allows us to more accurately assess the background rate of mutations within each sample. The overall rate of SNV within our dataset was significantly less than adult GBM (median = 958 SNV/sample vs. ~6,600 SNV/sample in aGBM(*54*)). However, there were 5 hyper-mutated samples within our dataset (>100,000 mutations/sample), all of which were pGBM samples (fisher exact test: p = 0.0093). Even excluding these samples, GBM had a significantly higher rate of SNV than DIPGs (GBM: min=220, max=25440, median=1194; DIPG: min=70, max = 6180, median=749; Wilcoxon rank sum test: p = 0.0013, Supplementary Figure 2).

Recent work by Alexandrov(*41*, *42*) and others have shown the importance of how specific mutational signatures within and across tumor types can provide insight into the causes of single nucleotide instability and the process of tumorigenesis(*55*). We applied these signature analysis to each tumor within our dataset (Figure 3). The sample with the most mutations (n=955,233) was almost entirely driven by the signature W2, at signature primary composed of Cytosine to Adenine transversion (Supplementary Figure 3 for in-depth signature contribution information) while the rest of the hypermutated samples primarily were composed of mutations consistent with signature W3, a signature associated with BRCA1/2 deficiency and almost exclusively composed of C-T transitions within the context of a neighboring guanine (SFig 3). Samples that strongly carry this signature of spontaneous deamination of C>T in NpCpG trinucleotides is observed in samples across tumor types(*42*) but is also a frequent mutational process within germline and normal somatic cells(*56*). Interestingly, this process

has previously been most highly linked to normal cell aging, as opposed to specific deficits in DNA repair. Of interest, both of these mutation types (C>A and C>T) were recently observed as occurring with increased frequency within individual cells during early development in normal human fetuses(*48*).

Recurrent regions of DNA copy number changes

While subject to less overall DNA copy number variation than its adult counterpart, recurrent regions of copy change are likely to be a major driver in pHGG development. Similar to previous studies, we found that chromosomes 1 and 2 were the most frequently amplified chromosomes (Supplementary Table 1), while chromosomes 13 and 14 were the most frequently deleted chromosomes. After accounting for arm and chromosome level changes, we applied the GISTIC algorithm, which looks for recurrent patterns of focal DNA copy number changes, to the WGS data from our tumors (Supplementary Figure 4). There were 12 regions of amplification with q-value<0.25, including regions with the known pHGG-associated genes *PDGFRA, MYCN, MET*, and a region on 8q24 just outside of *MYC* (see Supplementary Table 2). There were also 46 regions of significant deletion with q<0.25, including the most significant region containing *CDKN2A,* as well as regions containing CDKN1B and NF2 (Supplementary Table 3). There were a total of 68 chromothripsis events detected in 52 of our 178 samples with 40% and 23% of GBMs and DIPGs, respectively, having at least one event (FE test p=0.01; Supplementary Table 4). However, the most frequently affected chromosome (chr17) was affected primarily in DIPG samples (7 in DIPG, 1 in GBM; FE test p=0.014).

Recurrent Somatic Single Nucleotide variation

After excluding our 5 hypermutated samples to improve overall signal, our significance analysis showed heavy involvement of genes within the Histone 3, PI3K, and MYC transcriptional regulatory pathway. MUTSIG(*54*), which performs a gene level significance analysis looking for genes mutated significantly more than expected based on background rates of mutation, found 48 genes with q-value < 0.25 (Figure 4 shows top 30 genes; Supplementary Table 5 for complete list). Besides the known mutations in *H3F3A, HIST1H3B,* we also found that gene *HIST1H3C* was significantly mutated. Our significance analysis was also populated by many members of the PI3K pathway (*PIK3CA, PIK3R1, PTEN*), PDGF-R pathway (*PDGFRA, BRAF, EGFR, NF1*) that have been previously described in pHGG. New genes that this analysis has allowed us to detect include *TERF1*, which acts as an inhibitor of telomerase(*57*) suggesting similar function to *ATRX* (which was also significantly mutated) in these tumors. There were also several *MYC* associated genes, including previously described *MYC* and *MYCN*, as well as *MAX* (*MYC*-associated factor X) and that have not been previously described. We also performed MUTSIG on the 11 samples we identified as having very stable genomes above (Supplementary Table 6). Interestingly, the most significant gene was PDGFR, which was mutated in 4 out of the 11 of these stable samples (FE test p=0.008).

Significant regions of structural rearrangement

WGS provides a unique glimpse into the complex rearrangements DNA undergoes over the course of tumor evolution, and we found many regions that were recurrently affected by structural variation. The length of rearrangements affecting pHGG genomes had a bimodal distribution, with peaks at around 500 bases and 10mb (Supplementary Figure 5). The majority

of samples had less than 100 SV, with the sample with the largest number having over 1000 (Figure 2). Conceptually, the process of creating a rearrangement can be thought of in two steps. In the first, a strand of DNA is broken and in the second, two previously disparate strands are brought together. Similarly, we assessed statistical significance with regards to structural rearrangements in two distinct ways (methods developed and described in Wala et al(*43*)). In the first, we looked for regions of the genome that were subjected to increased frequency of breakpoints and structural rearrangements compared to a background model, in a method similar to the way significance is calculated for copy number change and SSNV (a so-called a 1-dimensional analysis). We also interrogated our dataset to look for any TWO specific, normally separate, regions of the genome that are brought together via structural rearrangement more often than we would expect by chance in our dataset, given a background model (a 2D analysis)(*43*). In our 1D analysis, we found eleven regions recurrently affected by DNA breakpoints within our dataset. This included regions containing *MET*, and *KIT,* genes whose alterations are known to play an important role in pHGG(*26*). In our 2D analysis, we found 12 region-pairs that underwent SV to bring then within close proximity more frequently than you would expect. Eleven of our DIPG samples had a structural rearrangement between an area near *MYC* and a gene 20mb downstream from it near a gene called *GSDMC*. This rearrangement was not observed in any of our GBM samples and, interestingly, this region was significant in both our 1D analysis, and our analysis of significantly amplified focal regions (see Supplementary tables 2 and 7). Additionally, we found 4 of our DIPG samples held structural rearrangements between a region containing *NMYC* and a region containing *ID2*. *ID2* is an NMYC target gene and is known to promote NMYC transformation in human neuroblastomas.

Given the downstream nature of this gene from *NMYC*, it's conceivable that these rearrangements lead to a positive feedback loop where n-MYC is able to drive its own expression through rearrangement of regions that regulate *ID2* to allow them to directly regulate *NMYC* transcription. Further suggesting the functional role of these rearrangements, samples with MYC-GSDMC rearrangements tended to have increased expression of the *MYC* (Supplementary Figure 6). Together, these two rearrangements occurred in 15 of our DIPG samples, yet were not observed in any of our GBM samples (FE test: $p<0.001$).

Discussion and Future directions

The landscape of somatic alterations that drive pHGG development

Unlike many other pediatric malignancies that are primarily driven by a single alteration or pathway, the landscape of driver mutations in pediatric high grade glioma seems to be drawn from many distinct cellular processes. Through a WGS of the largest cohort of pHGG to date, we have further confirmed the role of pathways involving H3.3 regulation, AKT signaling, and PDGFRA and members of its downstream pathway. This is the first study to have the power to show that *HIST1H3C* is significantly mutated along with *H3F3A, HIST1H3B,* reinforcing that there are likely many ways to accomplish the reduction of the repressive H3K27me3 that seems to drive tumorigenesis. An exciting possibility is that, like the difference observed in methylation patterns between *H3F3A* and *HIST1H3B*(*17*)*,* the mutations we observed in *HIST1H3C* my produce yet a third epigenetic signature. However, as all these mutations seem to lead to transformation, understanding the commonality between these 3 epigenetic fingerprints may help us identify precisely what about them causes transformation.

The novel rearrangements we describe in regions containing *MYC*, and *MYCN* provide yet another example of the importance of MYC/n-MYC in malignant transformation across many malignancies. It's worth noting that all of these rearrangements, (as well as 3 out of the 4 mutations in the MYC associated gene MAX) occurred within DIPG samples, suggested an outsized role for MYC within these tumors as compared to other pHGG. Further experimental validation of the exact structure of these rearrangements, as well as their influence on MYC/n-MYC expression will be required to confirm their effects, but this provides an important lead into the differences between these two groups of tumors. Additionally, the identification of somatic alterations that might drive MYC family/pathway expression

could be important for both future prognostics and therapeutics, as in medulloblastoma for example, where MYC expression is associated with worse overall prognosis(*58*).

Patterns of DNA alteration as a window into tumorigenesis

In adults, malignancies with high rates of somatic alterations are usually driven by known environmental factors (such as smoke in Lung cancer or UV radiation in Melanoma), or by some error in DNA repair machinery (such as BRCA 1/2 loss, defects in Mismatch repair, etc). While pHGG may have relatively stable genomes compared to their adult counterparts, the amount of genomic alteration we observed far outstrips most pediatric malignancies, yet the age of the children makes environmental exposure exceedingly unlikely, and out study did not show significant mutations within DNA repair machinery. Our SSNV signature analysis found relatively large contributions of signatures W3, W9 and W11 (signatures originally described it Alexandrov(*42*)). W9 is associated with failure of BRCA 1/2 DSB repair though no mutations in these genes were found in our dataset. Given recent work showing the increased of somatic mosaicism within early neural development(*48*) it is possible that this increased mutation rate is due to failure to eventually inhibit these natural mutational processes, as presumably happens in normal neural tissue. If that is the case, then there should be significant mutational burden within somatic cells directly adjacent to tumor tissue. We are now in the process of trying to “extract” these somatic mutations in pre-malignant cells by backcalculating the overall fraction of cells harboring or observed mutations versus the purity of the tumor within our sample. It is also becoming increasingly appreciated that many of the seeds for pediatric tumors may have occurred as mutations very early in embryonic development, and therefore would be in all cells within the individual. Our future goals also include interrogating the DNA from mono-nucleated blood cells from individuals with pHGG to look for recurrent somatic mutations present throughout the body.

Methods

Whole-genome sequencing

Libraries were prepared according to methods in (Saito et al, 2016)(*59*). Briefly, DNA was randomly fragmented and paired-end sequencing was performed on Illumina HiSeq at The Broad Institute of MIT and Harvard. Pooled libraries were normalized to 2 nM and denatured using 0.1 N NaOH before sequencing. Flowcell cluster amplification and sequencing were performed according to the manufacturer's protocols using HiSeq 2500. Each run was a 76-bp paired-end with an 8-base index barcode read. Data was analysed using the Broad Picard Pipeline which includes de-multiplexing and data aggregation (http://picard.sourceforge.net/). In addition, BAM sequencing files were retrieved from recently published PLGG datasets(*60*, *61*). Read pairs were aligned to reference genome hg19 (Build 37) using the Burrows-Wheeler Aligner (bwa) with options −q 5 −l 32 −k 2 −o 1 (*62*). Reads were sorted by coordinates, normalized, cleaned and duplicates were marked using SAMtools and Picard(*63*). Base quality score assignments were recalibrated to control for biases due to flow cell, lane, dinucleotide context and machine cycle using the Genome Analysis Toolkit (GATK4). DNA oxidation artifacts introduced through sequencing were computationally removed using a filter-based methods. The MuTect algorithm was used to call somatic mutations and subsequently filtered against a panel of normals(*64*). Base context frequency rates and mutation signature analyses were performed as described previously(*42*, *54*). Copy-number alterations were evaluated using SegSeq(*65*). GISTIC2 was used to identify recurrent copy-number alterations(*66*, *67*). Somatic point mutations and short indels were called using Mutect9(*68*) and IndelLocator. Mutsig (version 2.0)(*54*) was applied to detect significantly recurrent mutations and annotation of known varients was accomplished using Oncotator(*69*). Rearrangements and breakpoints were identified using dRanger and BreakPointer(*70*). Significant regions of recurrent breakpoints were identified as described previously(*43*). Quality control

of significant mutations and breakpoint regions was done through manual visual inspection in IGV(*71*). All analyses were performed within Firehose(*64*).

Mutation calling and significance analysis

MuTect(*68*) was applied to identify somatic single-nucleotide variants. Strelka(*72*) was applied to identify small insertions or deletions. Artifacts introduced by DNA oxidation during sequencing were computationally removed using a filter-based methods. Mutations with allelic fractions of less than 0.05 excluded. Annotation of identified variants was done using Oncotator(*69*). Mutational significance analysis was performed using MutSigCV(*73*). Manual review of mutations in putative significantly mutated genes was subsequently performed using Integrated Genomics Viewer(*73*). Comparison of mutations between samples from the same patient was accomplished using previously described methods(*74*, *75*), including application of force-calling to recover evidence of mutations called in one sample from other samples in the same individual. we used a previously described method(*76*) designed to recover evidence for mutations called in one sample in all other samples derived from the same individual, increasing sensitively to detect and recover mutations that might otherwise be missed.

Copy number determination

Copy-ratio profiles were inferred using GATK4 Allelic CNV algorithm. Read depth along the genome was normalized using panel of normals created from patients in our dataset. The resulting normalized copy ratios are then segmented using the circular binary segmentation algorithm(*77*). These data were used to produce allelic copy number data via through combination of allele fraction data with informative germline SNPs(*78*). Finally allelic copy number data was integrated with data from point mutations and short deletions and insertions as input to ABSOLUTE(*52*). The ABSOLUTE algorithm was used as previously described to generate purity and ploidy solutions for each tumour samples(*52*, *79*). These solutions provided estimates of total allelic copy number and cancer cell fraction for mutations and copy number events in the tumour samples. ABSOLUTE solutions were selected using manual curation. We

defined copy number events as arm level if the event spanned at least 80% of that arm and affected at least one allele. Chromothripsis determination was done using probability of event count on that chromosome versus other chromosomes in the sample based on a joint binomial distribution of event rate within sample and across samples on that chromosome. Then each chromosome with a p-value<$1e10^{-10}$ was manually curated for evidence of chromothripsis as described in Korbel and Campbell(*33*)

RNA-seq

RNA extraction of tumor samples were preformed using RNeasy (Qiagen), followed by library construction using a non-strand specific Illumunia TruSeq protocol. Flowcell cluster amplification and sequencing were performed using HiSeq 2000/2500 according to the manufacturer's protocol with a 76 bp paired-end run including an eight-base index barcode read. In addition, RNA sequencing files were downloaded from already published datasets(*19*, *20*). RNA-seq BAM files were transformed to fastq files using the Picard SamToFastq algorithm (http://picard.sourceforge.net/). Raw paired-end reads were aligned to reference genome hg19. Preprocessing, gene expression levels, quality metrics and fusion transcript calling were accomplished using PRADA (Pipeline for RNA-sequencing Data Analysis)(*34*) within Firehose.

Rearrangement significance and signature analysis

SNV signature analysis done as reported previously(*55*). Rearrangement significance done as reported in Wala et al with background rates and covariate regression parameters calculated from the ICGC pan-cancer dataset(*43*).

Phylogenetic analysis

We used the SSNV and allelic copy number information to draw phylogenetic trees for each patient with multiple tumor samples. A copy number event was considered shared between two related samples if

each of those samples contained an event whose start points fell within two probes of one another, whose end points fell within two probes of one another, and whose event amplitude was identical. Arm lengths are proportional to number of events delineating the samples on the ends of the branch. Phylogenetic tree construction was done as previously described(*74*, *75*).

Bibliography

1. Q. T. Ostrom *et al.*, *Neuro. Oncol.*, in press, doi:10.1093/neuonc/not151.

2. C. Jones, S. J. Baker, Unique genetic and epigenetic mechanisms driving paediatric diffuse high-grade glioma. *Nat. Rev. Cancer*. **14**, 651–661 (2014).

3. Q. T. Ostrom *et al.*, CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014. *Neuro. Oncol.* **19**, 1–88 (2017).

4. C. Jones, L. Perryman, D. Hargrave, Paediatric and adult malignant glioma: Close relatives or distant cousins? *Nat. Rev. Clin. Oncol.* **9**, 400–413 (2012).

5. I. Qaddoumi, I. Sultan, A. Gajjar, Outcome and prognostic features in pediatric gliomas: A review of 6212 cases from the surveillance, epidemiology, and end results database. *Cancer*. **115**, 5761–5770 (2009).

6. A. Broniscer *et al.*, Clinical and molecular characteristics of malignant transformation of low-grade glioma in children. *J. Clin. Oncol.* **25**, 682–689 (2007).

7. J. Rineer, D. Schreiber, K. Choi, M. Rotman, Characterization and outcomes of infratentorial malignant glioma: A population-based study using the Surveillance Epidemiology and End-Results database. *Radiother. Oncol.* **95**, 321–326 (2010).

8. A. L. Albright *et al.*, Magnetic resonance scans should replace biopsies for the diagnosis of diffuse brain stem gliomas: a report from the Children's Cancer Group. *Neurosurgery*. **33**, 1026-9-30 (1993).

9. H. S. Panitch, B. O. Berg, Brain Stem Tumors of Childhood and Adolescence. *Arch. Pediatr. Adolesc. Med.* **119**, 465 (1970).

10. G. M. Hunninghake, J. E. Gadek, The New England Journal of Medicine Downloaded from nejm.org at HARVARD UNIVERSITY on August 25, 2013. For personal use only. No other uses without permission. From the NEJM Archive. Copyright © 2009 Massachusetts Medical Society. All rights reserved. (2009).

11. D. Hargrave, Diffuse brainstem gliomas in children: should we or shouldn't we biopsy? *Br. J. Neurosurg.* **22**, 624–624 (2008).

12. T. J. MacDonald, Diffuse intrinsic pontine glioma (DIPG): Time to biopsy again? *Pediatr. Blood Cancer*. **58**, 487–488 (2012).

13. P. A. Northcott, S. M. Pfister, D. T. W. Jones, Next-generation (epi)genetic drivers of childhood

brain tumours and the outlook for targeted therapies. *Lancet Oncol.* **16**, 293–302 (2015).

14. J. Zhang *et al.*, Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nat. Genet.* **45**, 602–612 (2013).

15. G. Wu *et al.*, Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat. Genet.* **44**, 251–253 (2012).

16. J. Schwartzentruber *et al.*, Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*. **482**, 226–231 (2012).

17. D. Sturm *et al.*, Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell*. **22**, 425–437 (2012).

18. P. W. Lewis *et al.*, Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma. *Science (80-.).* **340**, 857–861 (2013).

19. S. Bender *et al.*, Reduced H3K27me3 and DNA Hypomethylation Are Major Drivers of Gene Expression in K27M Mutant Pediatric High-Grade Gliomas. *Cancer Cell*. **24**, 660–672 (2013).

20. J. Schwartzentruber *et al.*, Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*. **482**, 226–231 (2012).

21. A. D. Goldberg *et al.*, Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell*. **140**, 678–691 (2010).

22. P. W. Lewis, S. J. Elsaesser, K.-M. Noh, S. C. Stadler, C. D. Allis, Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc. Natl. Acad. Sci.* **107**, 14075–14080 (2010).

23. L. Bjerke *et al.*, Histone H3.3 mutations drive pediatric glioblastoma through upregulation of MYCN. *Cancer Discov.* **3**, 512–519 (2013).

24. K. R. Taylor *et al.*, Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma. *Nat. Genet.* **46**, 457–461 (2014).

25. T. Katagiri, Recent topics in fibrodysplasia ossificans progressiva. *J. Oral Biosci.* **54**, 119–123 (2012).

26. S. Bender *et al.*, Recurrent MET fusion genes represent a drug target in pediatric glioblastoma. *Nat. Med.* **22**, 1314–1320 (2016).

27. R. Beroukhim *et al.*, Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 20007–20012 (2007).

28. D. A. Bax *et al.*, A distinct spectrum of copy number aberrations in pediatric high-grade gliomas. *Clin. Cancer Res.* **16**, 3368–3377 (2010).

29. H. Qu *et al.*, Novel Chromosomal Imbalances in Pediatric glioma. *Neuro. Oncol.* **12**, 153–163 (2010).

30. R. McLendon *et al.*, Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. **455**, 1061–1068 (2008).

31. J. Barrow *et al.*, Homozygous loss of ADAM3A revealed by genome-wide analysis of pediatric

high-grade glioma and diffuse intrinsic pontine gliomas. *Neuro. Oncol.* **13**, 212–222 (2011).

32. S. Puget *et al.*, Use of integrated genomics to identify three high-grade glioma subtypes with distinct genetic profiles, pathway signatures, and clinicopathologic features. *J. Clin. Oncol.* **28**, e12500–e12500 (2010).

33. J. O. Korbel, P. J. Campbell, Criteria for inference of chromothripsis in cancer genomes. *Cell*. **152**, 1226–1236 (2013).

34. M. L. Leibowitz, C. Zhang, D. Pellman, Chromothripsis : A New Mechanism for Rapid Karyotype Evolution, 1–29 (2015).

35. C. Z. Zhang, M. L. Leibowitz, D. Pellman, Chromothripsis and beyond: Rapid genome evolution from complex chromosomal rearrangements. *Genes Dev.* **27**, 2513–2530 (2013).

36. H. Cai *et al.*, Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens. *BMC Genomics*. **15**, 82 (2014).

37. P. J. Stephens *et al.*, Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell*. **144**, 27–40 (2011).

38. C.-Z. Zhang, M. L. Leibowitz, D. Pellman, Chromothripsis and beyond: rapid genome evolution from complex chromosomal rearrangements. *Genes Dev.* **27**, 2513–2530 (2013).

39. B. S. Paugh *et al.*, Genome-Wide Analyses Identify Recurrent Amplifications of Receptor Tyrosine Kinases and Cell-Cycle Regulatory Genes in Diffuse Intrinsic Pontine Glioma. *J. Clin. Oncol.* **29**, 3999–4006 (2011).

40. F. Mertens, B. Johansson, T. Fioretos, F. Mitelman, The emerging complexity of gene fusions in cancer. *Nat. Rev. Cancer*. **15**, 371–381 (2015).

41. S. Nik-Zainal *et al.*, The life history of 21 breast cancers. *Cell*. **149**, 994–1007 (2012).

42. L. B. Alexandrov, S. Nik-zainal, D. C. Wedge, S. A. J. R. Aparicio, Signatures of mutational processes in human cancer. *Nature*. **500**, 415–421 (2014).

43. J. A. Wala *et al.*, Selective and mechanistic sources of recurrent rearrangements across the cancer genome. *bioRxiv*, 187609 (2017).

44. M. J. Garnett *et al.*, Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*. **483**, 570–575 (2012).

45. P. C. Fong *et al.*, Poly(ADP)-Ribose Polymerase Inhibition: Frequent Durable Responses in *BRCA* Carrier Ovarian Cancer Correlating With Platinum-Free Interval. *J. Clin. Oncol.* **28**, 2512–2519 (2010).

46. E. M. Van Allen *et al.*, Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov.* **4**, 1140–1153 (2014).

47. S. Nik-Zainal *et al.*, Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. **534**, 47–54 (2016).

48. T. Bae *et al.*, Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science (80-.).* **8690**, 1–10 (2017).

49. J. D. Rowley, Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*. **243**, 290–293 (1973).

50. C. Dupain, A. C. Harttrampf, G. Urbinati, B. Geoerger, L. Massaad-Massade, Relevance of Fusion Genes in Pediatric Cancers: Toward Precision Medicine. *Mol. Ther. Nucleic Acid*. **6**, 315–326 (2017).

51. P. Bandopadhayay *et al.*, MYB-QKI rearrangements in angiocentric glioma drive tumorigenicity through a tripartite mechanism. *Nat. Genet.* **48**, 273–282 (2016).

52. G. C. Scott L, C. K, H. E, Getz, Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

53. T. I. Zack *et al.*, Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).

54. M. S. Lawrence *et al.*, Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. **505**, 495–501 (2014).

55. J. Kim *et al.*, Somatic ERCC2 mutations are associated with a distinct genomic signature in urothelial tumors. *Nat. Genet.* **48**, 600–606 (2016).

56. J. S. Welch *et al.*, The Origin and Evolution of Mutations in Acute Myeloid Leukemia. *Cell*. **150**, 264–278 (2012).

57. X. Z. Zhou, K. P. Lu, The Pin2/TRF1-interacting protein PinX1 is a potent telomerase inhibitor. *Cell*. **107**, 347–59 (2001).

58. M. F. Roussel, G. W. Robinson, Role of MYC in Medulloblastoma. *Cold Spring Harb. Perspect. Med.* **3** (2013), doi:10.1101/cshperspect.a014308.

59. T. Saito, J. Sadoshima, MYB-QKI rearrangements in Angiocentric Glioma drive tumorigenicity through a tripartite mechanism. *Nat. Genet.* **48**, 273–282 (2016).

60. L. A. Ramkissoon *et al.*, Genomic analysis of diffuse pediatric low-grade gliomas identifies recurrent oncogenic truncating rearrangements in the transcription factor MYBL1. *Proc. Natl. Acad. Sci.* **110**, 8188–8193 (2013).

61. J. Zhang *et al.*, Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nat. Genet.* **45**, 602–612 (2013).

62. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).

63. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).

64. M. A. Chapman *et al.*, Initial genome sequencing and analysis of multiple myeloma. *Nature*. **471**, 467–472 (2011).

65. D. Y. Chiang *et al.*, High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat. Methods*. **6**, 99–103 (2009).

66. R. Beroukhim *et al.*, The landscape of somatic copy-number alteration across human cancers.

Nature. **463**, 899–905 (2010).

67. C. H. Mermel *et al.*, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, 1–14 (2011).

68. K. Cibulskis *et al.*, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–9 (2013).

69. A. H. Ramos *et al.*, Oncotator: Cancer Variant Annotation Tool. *Hum. Mutat.* **36**, E2423–E2429 (2015).

70. Y. Drier *et al.*, Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res.* **23**, 228–235 (2013).

71. J. T. Robinson *et al.*, Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

72. C. T. Saunders *et al.*, Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics*. **28**, 1811–1817 (2012).

73. M. S. Lawrence *et al.*, Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. **499**, 214–8 (2013).

74. W. J. Gibson *et al.*, The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis. *Nat. Genet.* **48**, 848–855 (2016).

75. P. K. Brastianos *et al.*, Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets. *Cancer Discov.* **5**, 1164–1177 (2015).

76. M. D. Stachler *et al.*, Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nat. Genet.* **47**, 1047–1055 (2015).

77. A. B. Olshen, E. S. Venkatraman, R. Lucito, M. Wigler, Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. **5**, 557–572 (2004).

78. S. L. Carter *et al.*, Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

79. T. I. Zack *et al.*, Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).

Figures

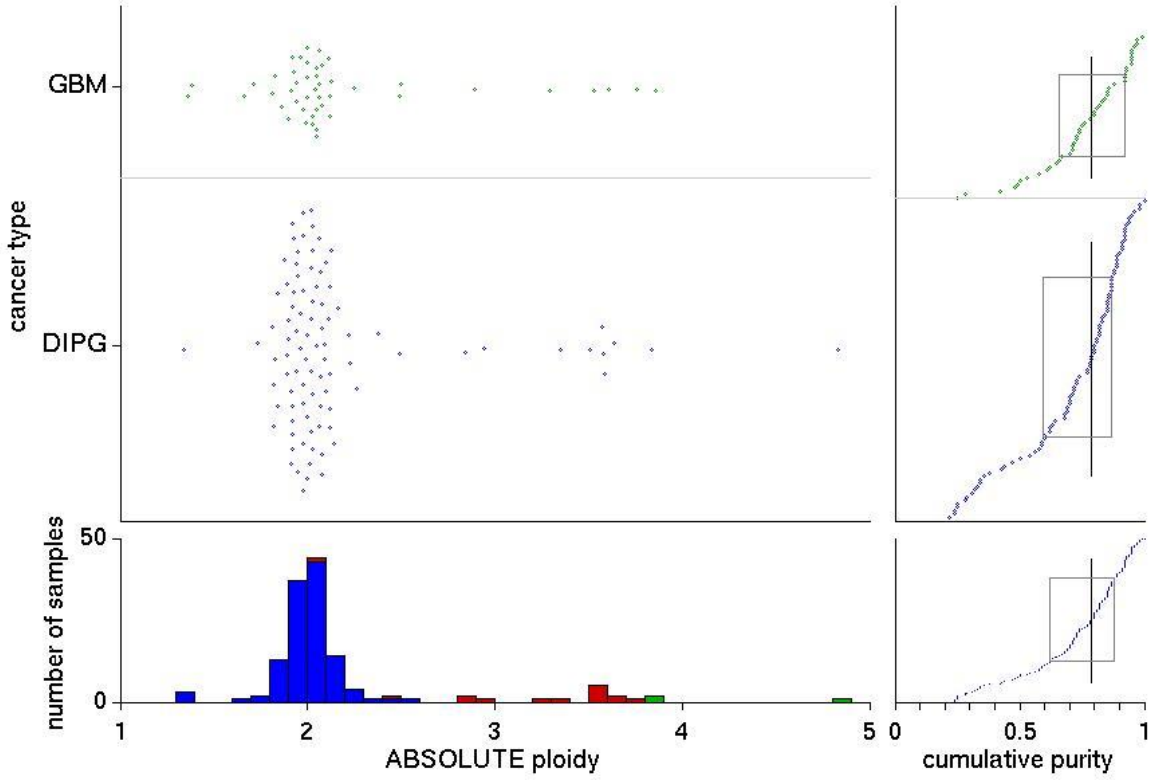


Figure 1 Purity and ploidy determination of samples across our dataset. The top right of the graph plots the overall ploidy across primary tumors within our dataset (GBM above, DIPG below), with a cumulative distribution of their purity levels on the top left. The bottom panel shows a histogram of ploidy across our entire cohort. Red in this histogram indicates samples that have undergone a genome doubling and green indicating samples that have undergone more than one genome doubling.

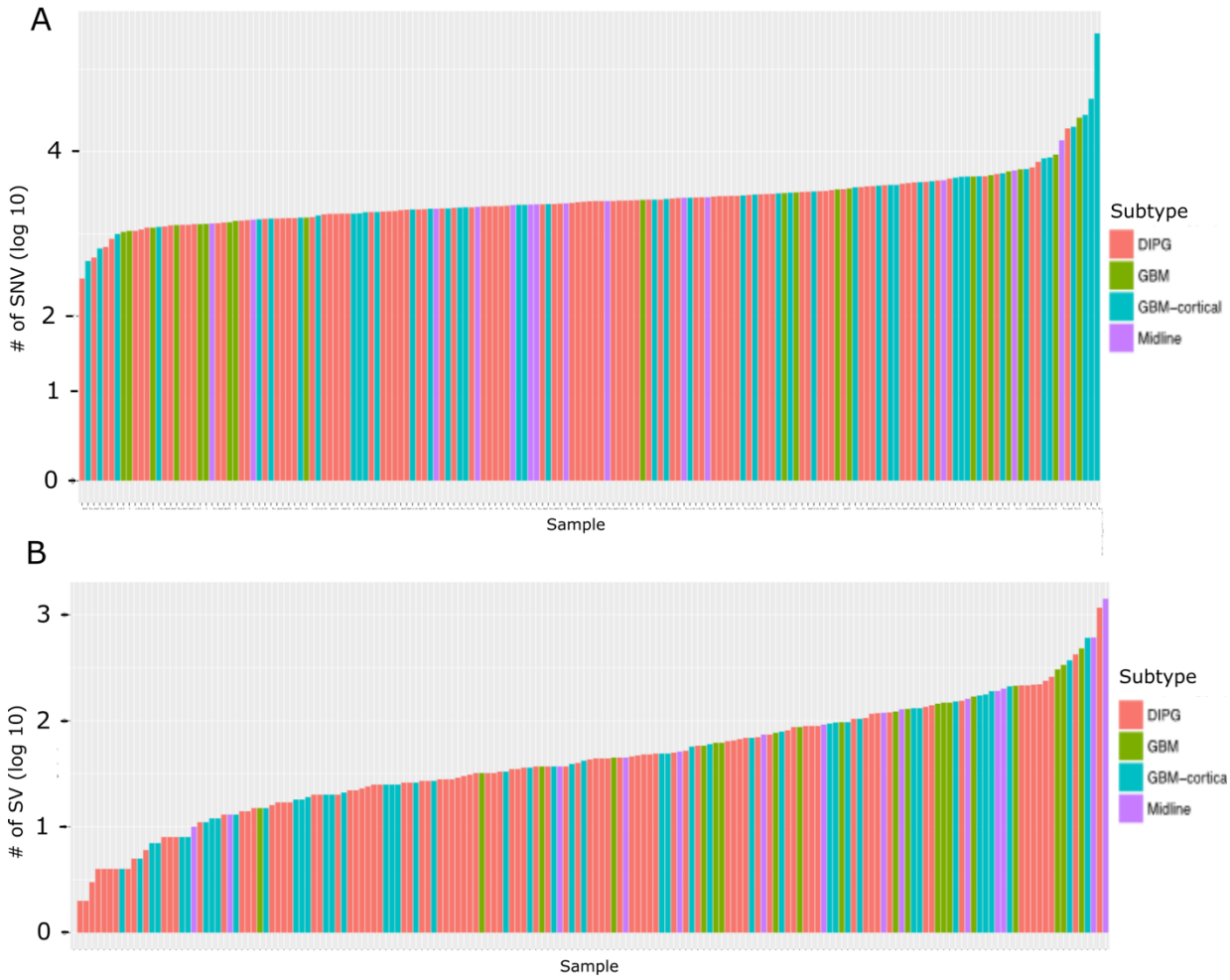


Figure 2 Rate of Single Nucleotide Variations (A) and Structural Variations (B) within samples in our dataset.

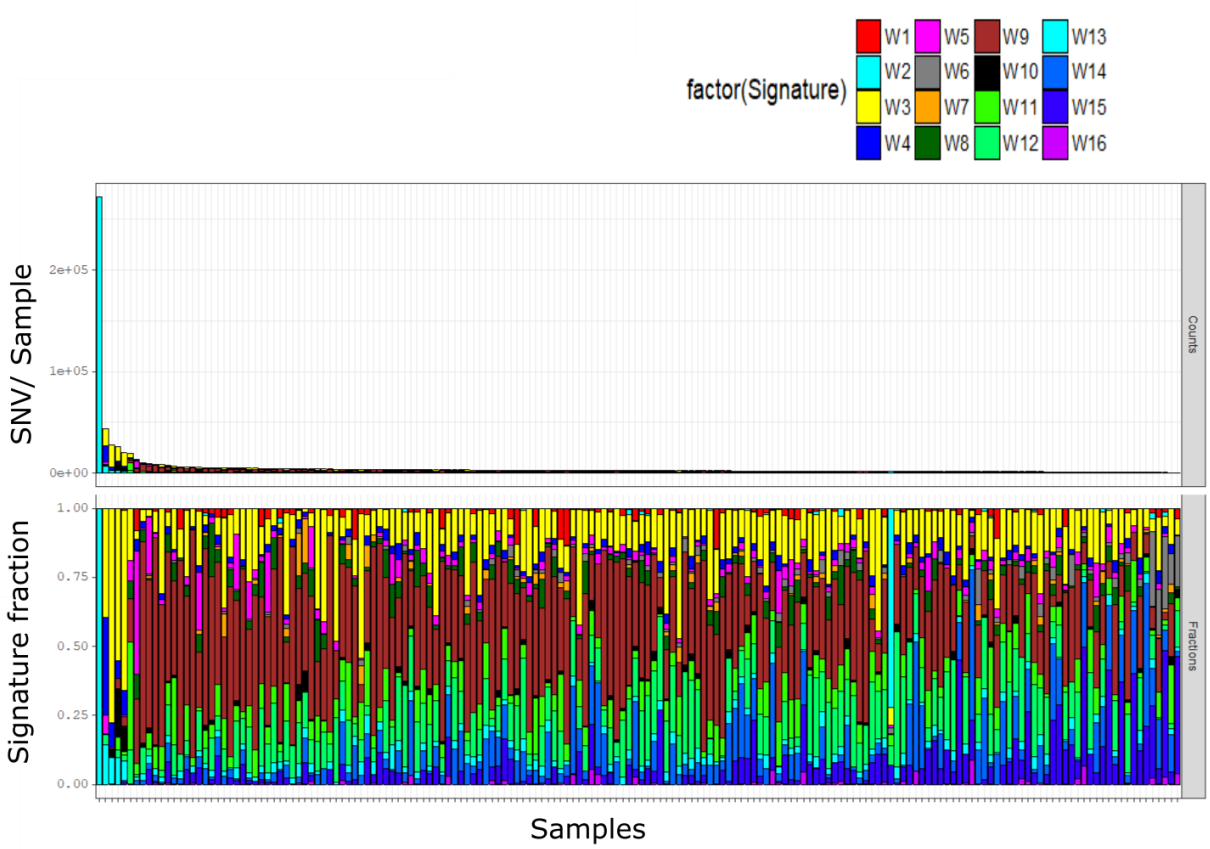


Figure 3 SNV signature contributions by sample. Signatures were derived through Negative matrix factorization based on base change and immediately adjacent base context (see Methods). Definitions of signatures W1-W16 listed in Supplementary Figure 3.

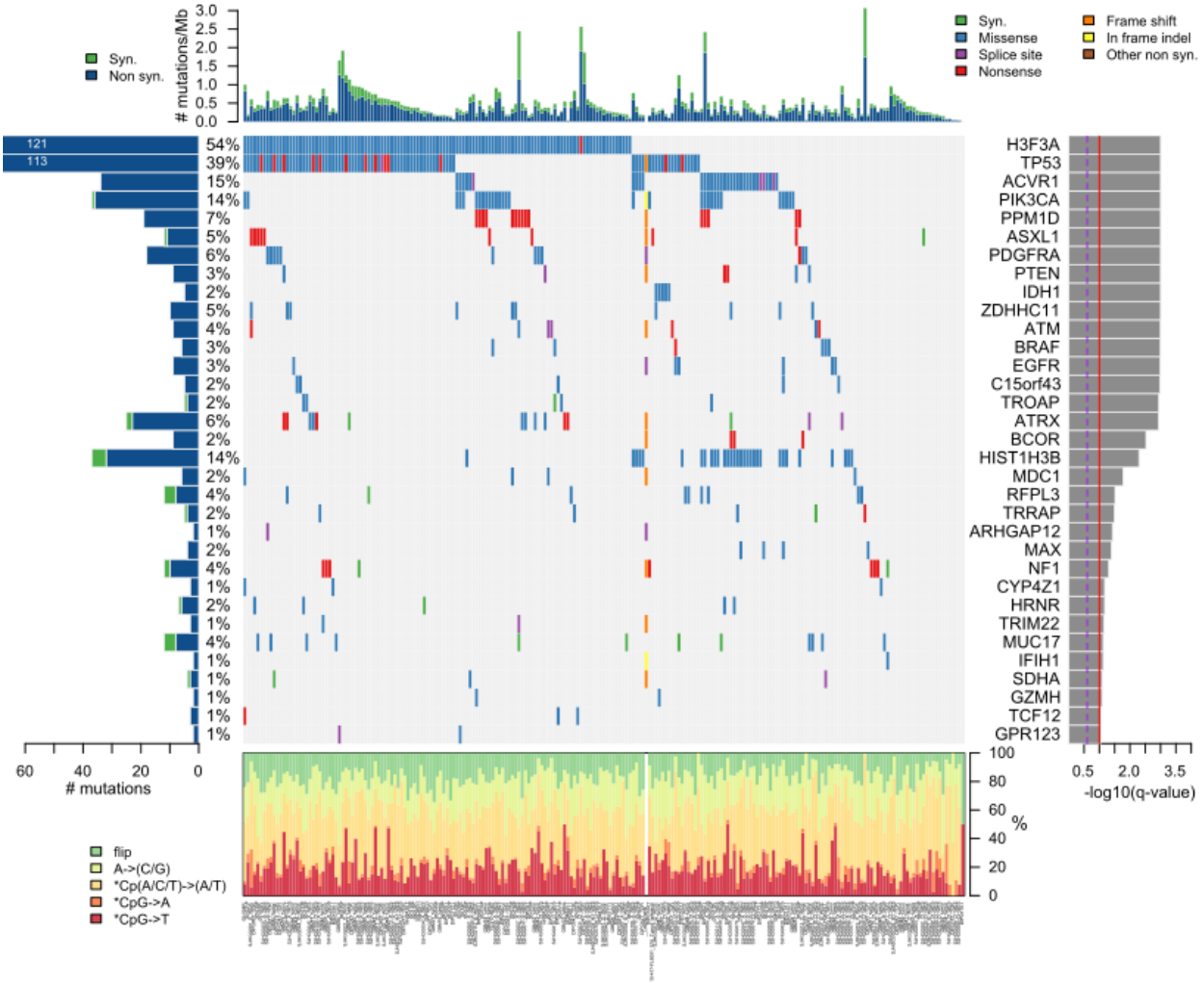
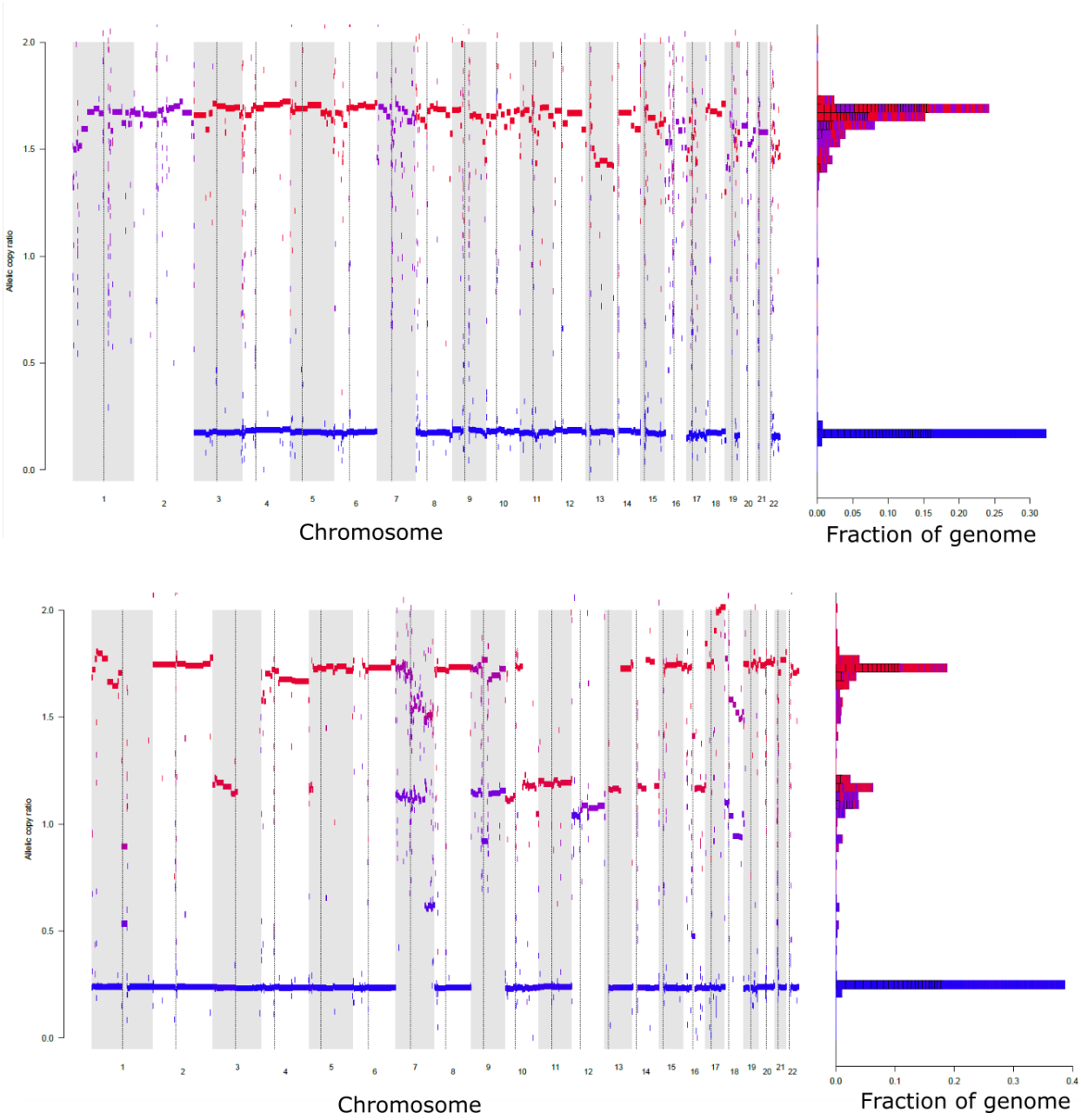
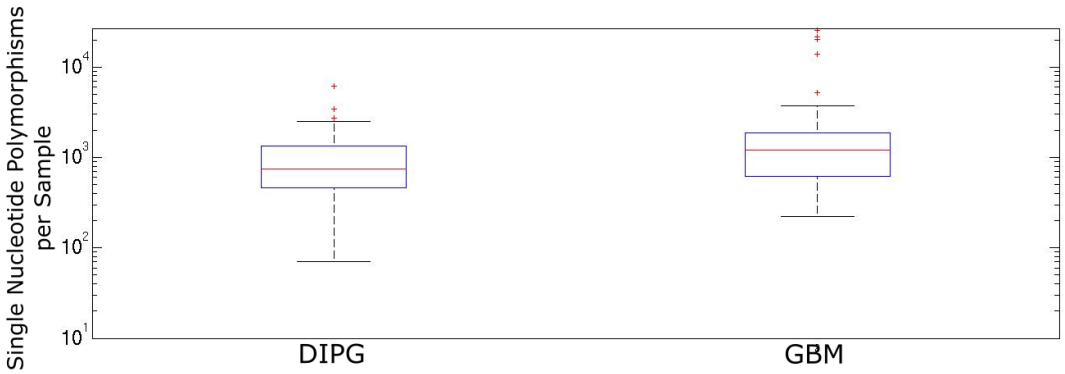


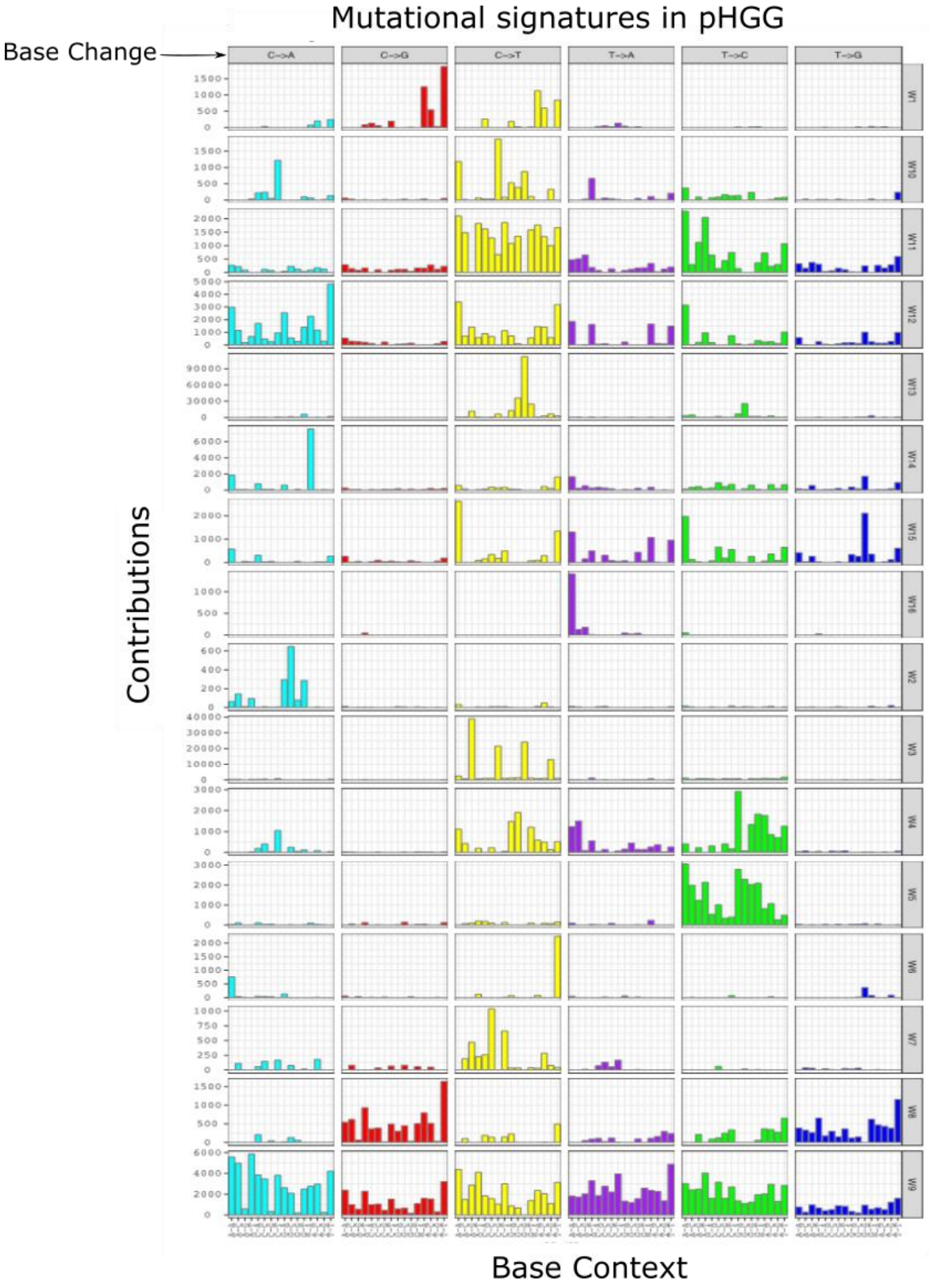
Figure 4: Significant mutations within pHGG. Mutations ordered by significance in MUTSIG algorithm (shown on far right panel). Number of samples with mutation shown in far left panel. Middle section shows specific mutations in each sample (column).



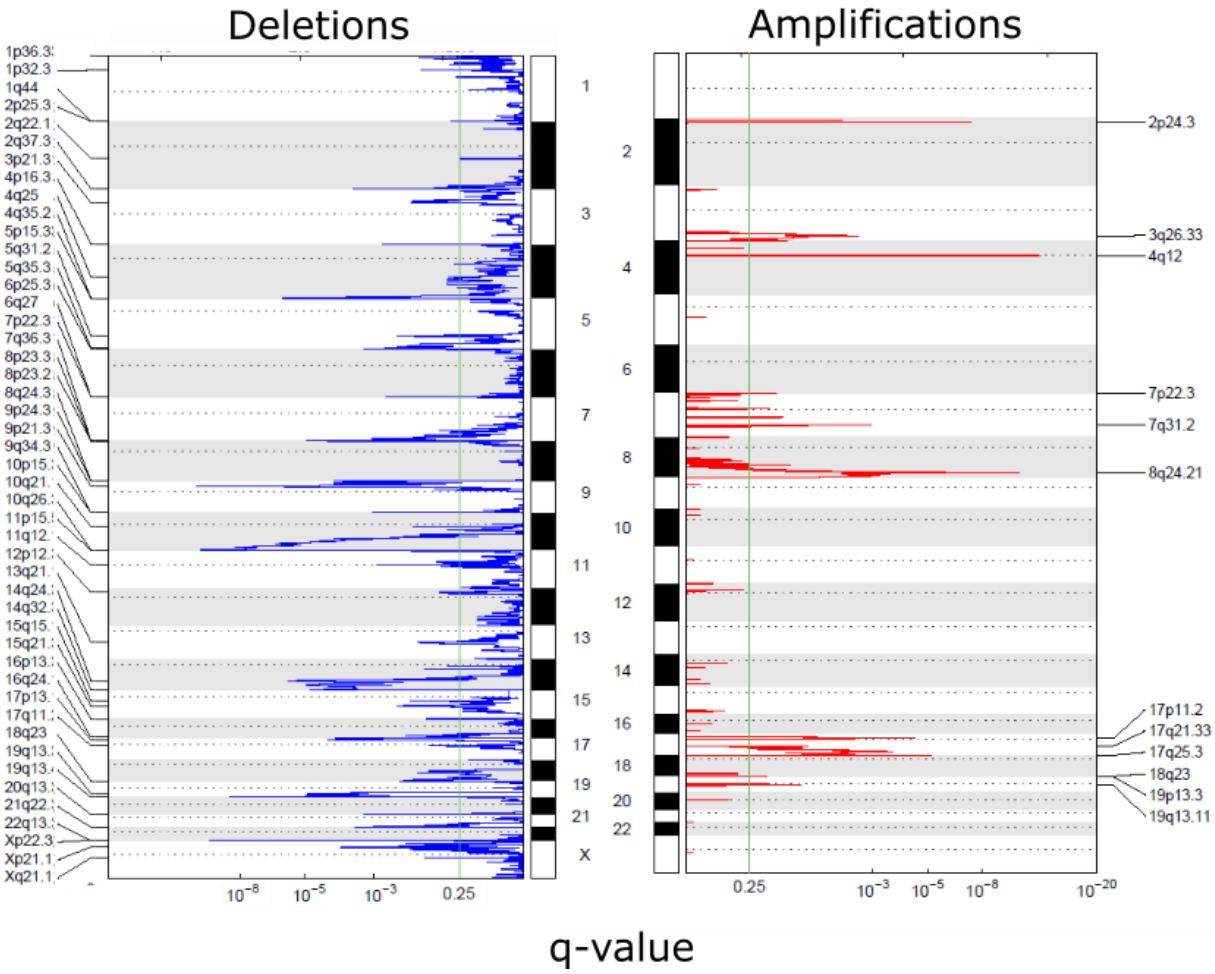
Supplementary Figure 1: GBM samples that have undergone almost genome-wide loss of heterozygosity. Top: Sample 10-417-2263 with genome-wide loss of heterozygosity excluding chromsomes 1,2 and 7. Bottom: Sample GBM11 with genome-wide loss of heterzygoisty excluding chromosome 7, 9, and 12. The Y-axis represents allelic copy ratio in that sample, where the allelic segment with the highest measured copy ratio at that location is in red and the allelic segment with the lowest copy ratio at that segment is in blue (purple indicates the two alleles have overlapping copy ratios). The far right is a histogram of all segments in the sample, as measured by the fraction of the 2N genome.



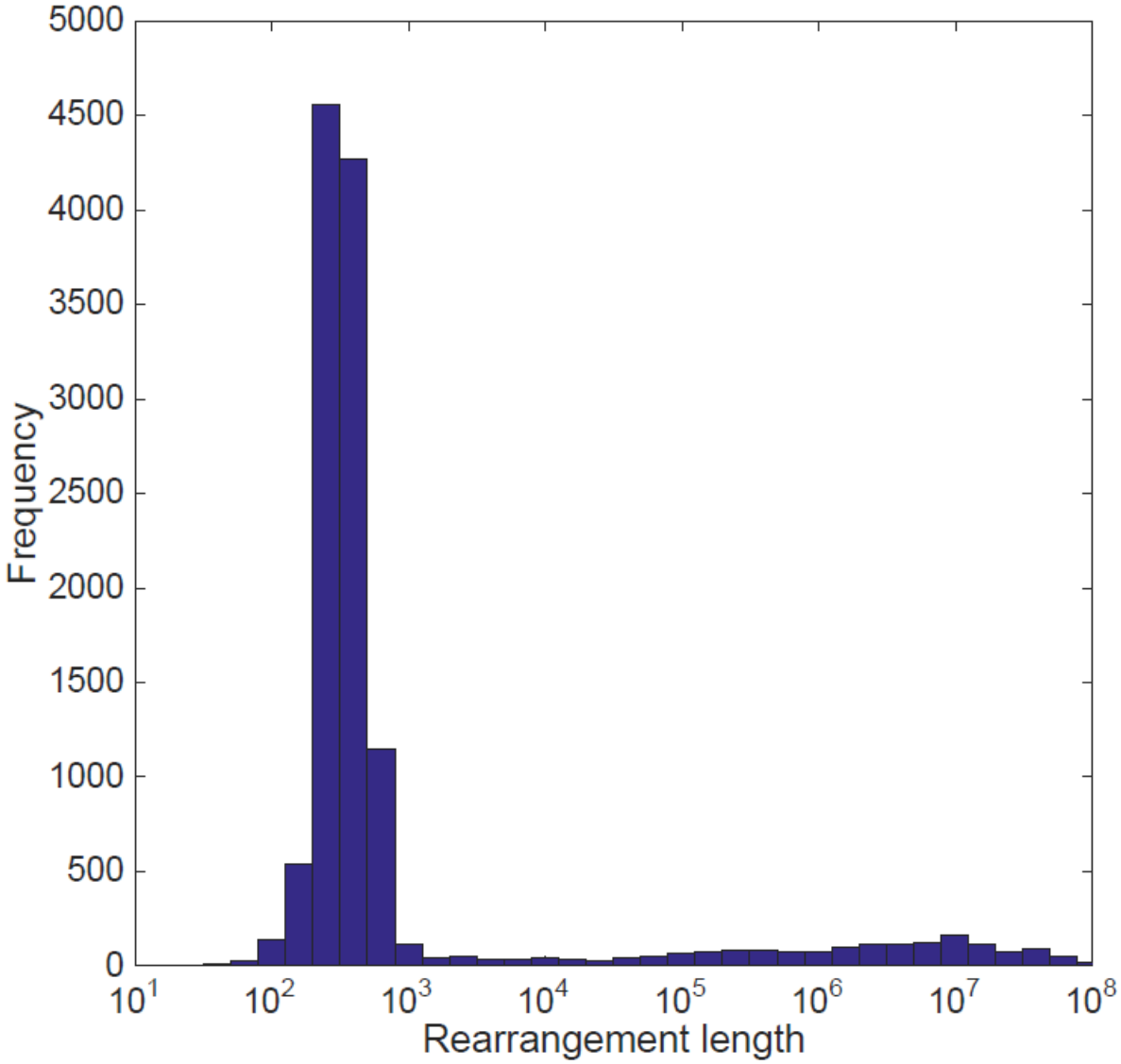
Supplementary Figure 2 Mutation rate in DIPG vs GBM. Hypersegmented samples removed from this plot for clarity (and to conform with Wilcoxan rank sum test done in main text).



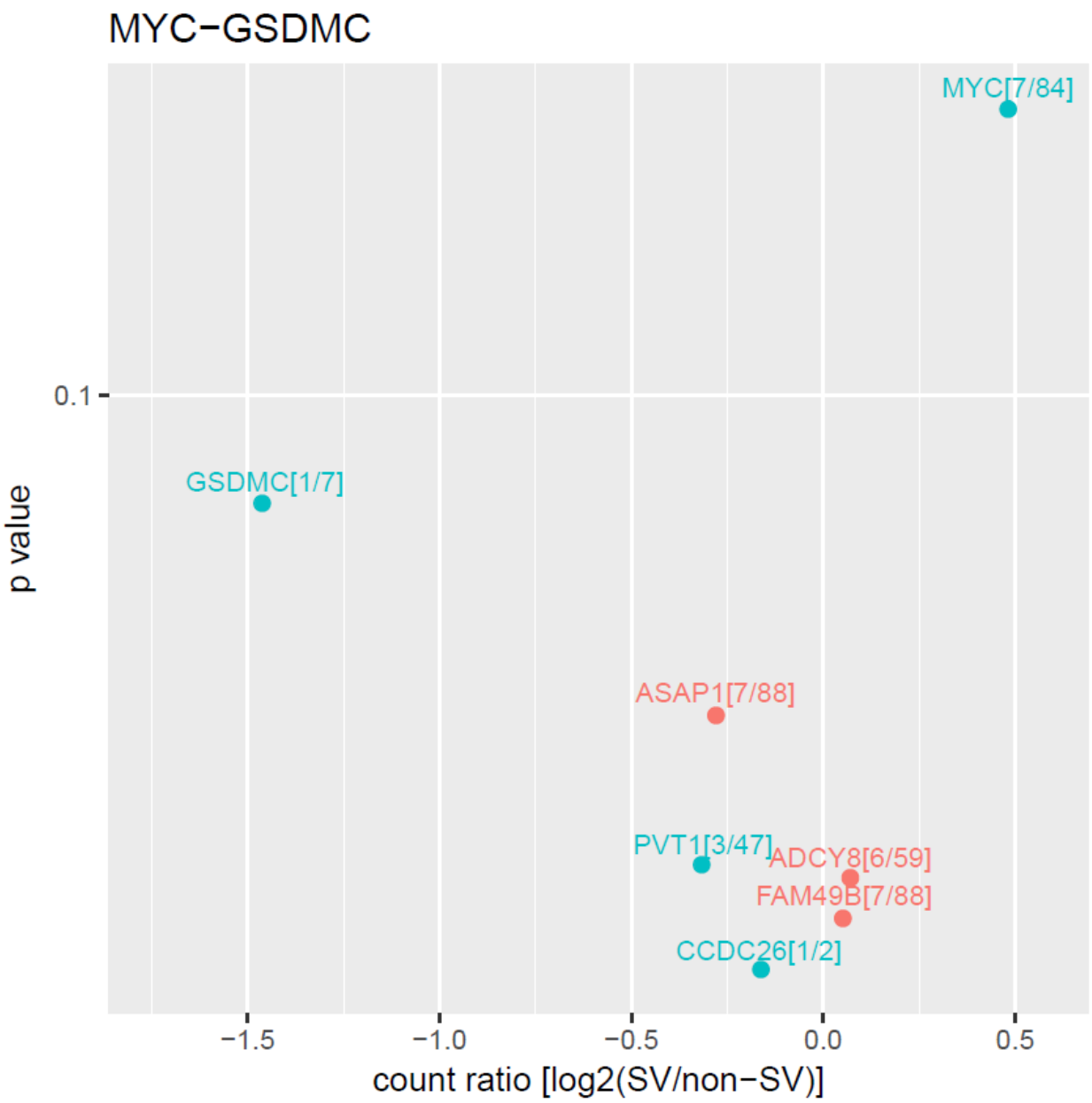
Supplementary Figure 3: Mutational signature contributions by base change and context (neighboring bases)



Supplementary Figure 4: Significant regions of amplification (red, right) and deletion (blue, left).



Supplementary Figure 5: Histogram of observed rearrangements, by length.



Supplementary Figure 6: MYC expression in samples with a MYC-GSDMC rearrangement vs those without.

Supplementary Table 1 - Chromosome level copy number change results									
Arm	# Genes	Amp frequency	Amp frequency score	Amp z-score	Amp q-value	Del frequency	Del frequency score	Del z-score	Del q-value
1p	2121	0.28	0.31	8.75	0	0.08	0.11	0.323	0.687
1q	1955	0.41	0.42	13.6	0	0.02	0.04	-1.99	0.996
2p	924	0.17	0.19	3.25	0.00524	0.08	0.1	-0.249	0.86
2q	1556	0.18	0.19	3.55	0.00293	0.05	0.06	-1.46	0.996
3p	1062	0.09	0.1	-0.357	1	0.08	0.09	-0.593	0.951
3q	1139	0.08	0.08	-0.836	1	0.09	0.1	-0.364	0.891
4p	489	0.03	0.03	-3.06	1	0.09	0.09	-0.733	0.967
4q	1049	0.04	0.05	-2.39	1	0.1	0.1	-0.0302	0.789
5p	270	0.05	0.06	-1.98	1	0.11	0.12	0.323	0.687
5q	1427	0.05	0.05	-1.97	1	0.12	0.13	1.14	0.286
6p	1173	0.04	0.04	-2.51	1	0.14	0.14	1.51	0.177
6q	839	0.04	0.05	-2.29	1	0.14	0.15	1.69	0.13
7p	641	0.17	0.17	2.64	0.0271	0.04	0.05	-2.24	0.996
7q	1277	0.18	0.18	3.3	0.00524	0.03	0.04	-2.63	0.996
8p	580	0.07	0.08	-1.18	1	0.1	0.11	-0.0177	0.789
8q	859	0.08	0.09	-0.642	1	0.09	0.1	-0.408	0.891
9p	422	0.09	0.1	-0.205	1	0.09	0.1	-0.205	0.86
9q	1113	0.12	0.12	0.814	1	0.05	0.06	-1.78	0.996
10p	409	0.02	0.02	-3.27	1	0.2	0.2	3.67	0.000405
10q	1268	0.02	0.02	-2.99	1	0.24	0.25	6.02	5.78E-09
11p	862	0.01	0.02	-3.43	1	0.21	0.22	4.55	1.05E-05
11q	1515	0.01	0.01	-3.77	1	0.15	0.15	2.23	0.0399
12p	575	0.09	0.1	-0.248	1	0.08	0.08	-0.945	0.996
12q	1447	0.08	0.08	-0.836	1	0.07	0.08	-1.07	0.996
13p	0	0.02	0.02	-3.11	1	0.27	0.27	6.47	3.84E-10
13q	654	0.02	0.02	-2.95	1	0.28	0.29	7.5	3.59E-13
14p	0	0.02	0.03	-2.77	1	0.28	0.29	7.25	1.87E-12
14q	1341	0.02	0.02	-2.85	1	0.28	0.29	7.83	3.75E-14
15p	0	0.02	0.02	-3.53	1	0.13	0.13	0.817	0.433
15q	1355	0.02	0.02	-3.29	1	0.14	0.14	1.47	0.179
16p	872	0.02	0.03	-3.18	1	0.11	0.12	0.347	0.687
16q	702	0.03	0.04	-2.59	1	0.22	0.23	4.86	3.42E-06
17p	683	0.04	0.05	-2.34	1	0.21	0.22	4.64	8.04E-06
17q	1592	0.07	0.07	-1.24	1	0.09	0.1	-0.0361	0.789
18p	143	0.04	0.05	-2.41	1	0.22	0.23	4.68	7.29E-06
18q	446	0.04	0.05	-2.38	1	0.21	0.22	4.55	1.05E-05
19p	995	0.17	0.17	2.8	0.0197	0.05	0.06	-1.9	0.996
19q	1709	0.12	0.13	1.17	0.694	0.09	0.1	-0.0284	0.789
20p	355	0.09	0.1	-0.533	1	0.08	0.09	-0.764	0.967
20q	753	0.08	0.09	-0.82	1	0.05	0.06	-1.99	0.996
21p	13	0.07	0.08	-1.11	1	0.21	0.22	4.34	2.50E-05

21q	509	0.08	0.09	-0.807	1	0.13	0.14	1.28	0.245
22p	0	0.03	0.03	-2.99	1	0.14	0.14	1.12	0.286
22q	921	0.04	0.04	-2.66	1	0.11	0.11	0.163	0.77
Xp	834	0.04	0.08	-0.961	1	0.46	0.48	15.2	0
Xq	1312	0.04	0.07	-0.967	1	0.44	0.46	14.7	0

Supplementary Table 2 - Significantly amplified regions			
cytoband	q value	residual q value	wide peak boundaries
4q12	3.54E-14	3.54E-14	chr4:54467130-55410905
8q24.21	1.78E-09	1.78E-09	chr8:130579494-130672420
2p24.3	4.79E-08	4.79E-08	chr2:16080723-16175697
17p11.2	5.65E-05	5.65E-05	chr17:17313400-18553898
17q25.3	9.82E-06	7.97E-05	chr17:80646099-81195210
7q31.2	0.001587	0.001587	chr7:116332937-116521447
3q26.33	0.0040514	0.0040514	chr3:170801975-198022430
19q13.11	0.058424	0.058424	chr19:32641004-32667088
17q21.33	0.051528	0.063581	chr17:46086326-80229226
7p22.3	0.098535	0.098535	chr7:1-40721434
18q23	0.15777	0.15777	chr18:43067507-78077248
14q32.2	0.16607	0.16607	chr14:66188507-107349540

Supplementary Table 3 - Significantly Deleted regions			
cytoband	q value	residual q value	wide peak boundaries
9p21.3	1.11E-11	1.11E-11	chr9:21558222-22447896
10q26.3	1.11E-11	1.11E-11	chr10:133794683-135534747
Xp22.33	3.15E-09	1.54E-08	chrX:1-150070792
14q24.3	3.65E-07	4.84E-06	chr14:72205417-78871492
19q13.43	4.38E-09	1.53E-05	chr19:58661093-59128983
7q36.3	9.47E-05	9.47E-05	chr7:158379504-159138663
11p15.5	0.000412	0.000412	chr11:1-2291360
17p13.1	0.000558	0.000558	chr17:1-62227316
5p15.33	0.001057	0.001057	chr5:1-988568
19q13.31	3.40E-05	0.001144	chr19:41312680-49793303
4q34.3	0.000111	0.001809	chr4:177713504-191154276
2q37.3	0.003192	0.003192	chr2:242447057-243199373
21q22.3	0.004108	0.004052	chr21:33039190-48129895
6p25.3	0.004537	0.00462	chr6:1-1313166
11q12.1	0.004966	0.005139	chr11:50003977-56021568
20q13.33	0.006982	0.006963	chr20:61050432-63025520
4q35.2	1.06E-06	0.009593	chr4:188425210-191154276
10p15.3	0.011526	0.011526	chr10:1-697679
14q32.33	4.45E-06	0.011526	chr14:81999730-107349540
13q21.1	0.01483	0.014845	chr13:53626144-78118898
16q24.1	0.018485	0.018485	chr16:70834128-90354753
8p23.2	0.000557	0.03335	chr8:2078769-6261748
1p36.33	0.03483	0.035659	chr1:1-48999162
4p16.3	0.047737	0.047788	chr4:1-1065630
6q27	0.057913	0.056823	chr6:170145909-171115067
8p23.3	0.001673	0.056823	chr8:1-1993202
15q15.1	0.059392	0.059704	chr15:1-83381345
3p21.31	0.064002	0.064184	chr3:1-63992869
10q21.1	0.068484	0.067549	chr10:52643502-72238741
17q11.2	0.068631	0.067549	chr17:27226618-45331993
5q31.2	0.035752	0.070719	chr5:128729143-180915260
12q24.33	0.082163	0.083154	chr12:104231615-133851895

8q24.3	0.086598	0.084178	chr8:6821836-146364022
5q35.3	0.012276	0.10354	chr5:156900318-180915260
16p13.3	0.10713	0.10714	chr16:1-1665521
Xp21.1	0.00161	0.11524	chrX:30869677-34646164
1q44	0.1408	0.1408	chr1:209848374-249250621
Xq21.1	0.14777	0.15078	chrX:76710681-77155515
13q34	0.15518	0.15649	chr13:79231544-115169878
2q22.1	0.21635	0.22163	chr2:139655535-143635744
9p24.3	0.001144	0.23238	chr9:1-7796635
12q23.2	0.24023	0.23238	chr12:1-132288242
7p22.3	0.24233	0.24287	chr7:1-7223879
9q34.3	0.24233	0.24287	chr9:123935549-141213431
22q13.33	0.24606	0.2465	chr22:18573479-51304566
4q25	0.20875	0.95864	chr4:1-191154276

Supplementary Table 4 - Chromothripsis counts			
Chromosome	# of DIPG samples	# of GBM samples	Total
1	4	2	6
2	1	1	2
3	1	2	3
4	1		1
5	0	0	0
6		1	1
7	1	6	7
8	1	0	1
9	0	5	5
10	2	1	3
11	2	1	3
12	1	3	4
13	2	1	3
14	0	4	4
15	0	2	2
16	3	3	6
17	7	1	8
18	1	2	3
19	3	1	4
20	0	1	1
21	1	0	1
22	0	0	0
	31	37	68

Supplementary Table 5 -Significant SSNV gene mutation list							
rank	gene	longname	codelen	npat	nsite	p	q
1	H3F3A	H3 histone, family 3A	431	87	5	1.00E-16	2.67E-13
2	TP53	tumor protein p53	1955	79	52	1.00E-16	2.67E-13
3	ACVR1	activin A receptor, type I	1583	26	6	1.00E-16	2.67E-13
4	HIST1H3B	histone cluster 1, H3b	808	20	2	1.00E-16	2.67E-13
5	FAM186A	family with sequence similarity 186, member A	7088	14	13	1.00E-16	2.67E-13
6	PPM1D	protein phosphatase 1D magnesium-dependent, delta isoform	1842	14	13	1.00E-16	2.67E-13
7	BCOR	BCL6 co-repressor	5325	11	11	1.00E-16	2.67E-13
8	PDGFRA	platelet-derived growth factor receptor, alpha polypeptide	3364	15	20	3.33E-16	7.79E-13
9	ATRX	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)	7615	18	20	5.55E-16	1.15E-12
10	PIK3CA	phosphoinositide-3-kinase, catalytic, alpha polypeptide	3288	28	17	2.44E-15	4.57E-12

11	ASXL1	additional sex combs like 1 (Drosophila)	4683	10	7	4.83E-13	8.21E-10
12	PTEN	phosphatase and tensin homolog (mutated in multiple advanced cancers 1)	1559	8	6	4.61E-10	7.18E-07
13	NF1	neurofibromin 1 (neurofibromatosis, von Recklinghausen disease, Watson disease)	12110	12	20	2.99E-09	4.30E-06
14	GTF2I	general transcription factor II, i	3133	6	3	7.84E-08	1.05E-04
15	PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	2361	11	13	1.78E-07	2.21E-04
16	DRD5	dopamine receptor D5	1438	9	5	4.44E-07	5.19E-04
17	SUSD2	sushi domain containing 2	2529	5	3	6.26E-07	6.89E-04
18	IDH1	isocitrate dehydrogenase 1 (NADP+), soluble	1292	4	2	1.04E-06	1.08E-03
19	ZFP36L2	zinc finger protein 36, C3H type-like 2	1492	3	2	2.08E-06	1.96E-03
20	UNC93B1	unc-93 homolog B1 (C. elegans)	1835	7	5	2.09E-06	1.96E-03
21	RFPL3	ret finger protein-like 3	968	6	5	4.80E-06	4.27E-03
22	FAM72A	family with	462	4	1	5.86E-06	4.98E-03

		sequence similarity 72, member A					
23	MDC1	mediator of DNA damage checkpoint 1	6395	9	8	9.07E-06	7.38E-03
24	AGAP6	ArfGAP with GTPase domain, ankyrin repeat and PH domain 6	2090	7	4	1.24E-05	9.36E-03
25	BRAF	v-raf murine sarcoma viral oncogene homolog B1	2371	5	4	1.25E-05	9.36E-03
26	EGFR	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	4007	9	14	1.38E-05	9.95E-03
27	TUBA3D	tubulin, alpha 3d	1373	4	3	1.87E-05	1.30E-02
28	KIAA0947	KIAA0947	6873	3	7	2.50E-05	1.67E-02
29	ZNF98	zinc finger protein 98 (F7175)	1732	8	4	2.83E-05	1.82E-02
30	TCF12	transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)	2279	4	5	3.45E-05	2.15E-02
31	CACNA1C	calcium channel, voltage-dependent, L type, alpha 1C subunit	7311	8	6	3.87E-05	2.34E-02

32	HIST1H3C	histone cluster 1, H3c	796	3	1	6.18E-05	3.61E-02
33	PCDHB13	protocadherin beta 13	2695	6	1	1.24E-04	7.05E-02
34	DPP4	dipeptidyl-peptidase 4 (CD26, adenosine deaminase complexing protein 2)	2401	4	4	1.45E-04	8.00E-02
35	FGFR1	fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome)	2963	4	4	1.57E-04	8.16E-02
36	MAX	MYC associated factor X	1962	4	2	1.57E-04	8.16E-02
37	ZNF100	zinc finger protein 100	1647	4	3	1.63E-04	8.25E-02
38	TERF1	telomeric repeat binding factor (NIMA-interacting) 1	1358	3	2	1.74E-04	8.55E-02
39	CYP4A11	cytochrome P450, family 4, subfamily A, polypeptide 11	1606	8	7	2.69E-04	1.29E-01
40	CSAD	cysteine sulfinic acid decarboxylase	1652	4	4	3.29E-04	1.54E-01
41	STIP1	stress-induced-phosphoprotein 1 (Hsp70/Hsp90-organizing protein)	1687	3	3	3.76E-04	1.68E-01

42	KRTAP5-5	keratin associated protein 5-5	718	3	2	3.76E-04	1.68E-01
43	POTEH	POTE ankyrin domain family, member H	1679	8	4	4.01E-04	1.71E-01
44	FAM153B	family with sequence similarity 153, member B	1250	7	2	4.02E-04	1.71E-01
45	SNX3	sorting nexin 3	505	3	3	4.44E-04	1.85E-01
46	STARD6	StAR-related lipid transfer (START) domain containing 6	682	3	2	4.54E-04	1.85E-01
47	SEMG1	semenogelin I	1405	4	4	4.95E-04	1.97E-01
48	LATS2	LATS, large tumor suppressor, homolog 2 (Drosophila)	3301	4	3	5.28E-04	2.06E-01

Supplementary Table 6 -Significant SSNV gene mutation list for patients with stable genomes							
rank	gene	longname	codelen	npat	nsite	p	q
1	PDGFRA	platelet-derived growth factor receptor, alpha polypeptide	3364	4	10	1.87E-04	1
2	ASXL1	additional sex combs like 1 (Drosophila)	4683	3	3	2.01E-03	1
3	RSPH10B	radial spoke head 10 homolog B (Chlamydomonas)	26226	1	3	2.11E-03	1
4	PGLYRP3	peptidoglycan recognition protein 3	1054	1	1	2.73E-03	1
5	SPDYE2	speedy homolog E2 (Xenopus laevis)	2476	1	1	3.31E-03	1
6	UBE2S	ubiquitin-conjugating enzyme E2S	684	1	1	3.80E-03	1
7	KIAA0408	KIAA0408	2108	1	2	4.06E-03	1
8	DUS3L	dihydrouridine synthase 3-like (S. cerevisiae)	2005	1	3	4.39E-03	1