Medulloblastoma Group 3 and 4 Tumors Comprise a Clinically and Biologically Significant Expression Continuum Reflecting Human Cerebellar Development

Daniel Williamson[1*], Edward C Schwalbe[1,2], Debbie Hicks[1], Kimberly A Aldinger[3], Janet C Lindsey[1], Stephen Crosier[1], Stacey Richardson[1], Jack Goddard[1], Rebecca M Hill[1], Jemma Castle[1], Yura Grabovska[1,4], James Hacking[1], Barry Pizer[5], Stephen B Wharton[6], Thomas S Jacques[7], Abhijit Joshi[8], Simon Bailey[1], Steven C Clifford[1*].

[1] Wolfson Childhood Cancer Research Centre, Translational and Clinical Research Institute, Newcastle University Centre for Cancer, Newcastle University, Newcastle upon Tyne, UK.

[2] Department of Applied Sciences, Northumbria University, Newcastle upon Tyne, UK.

[3] Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA, USA

[4] Division of Molecular Pathology, Institute of Cancer Research, London, UK

[5] Institute of Translational Research, University of Liverpool, Liverpool, UK.

[6] Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK.

[7] Developmental Biology and Cancer Programme, UCL GOS Institute of Child Health, London and Department of Histopathology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

[8] Department of Neuropathology, Royal Victoria Infirmary (RVI), Newcastle University Teaching Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

*** Corresponding Authors:** e-mail: daniel.williamson@newcastle.ac.uk, steve.clifford@newcastle.ac.uk

Summary:

Medulloblastoma is currently sub-classified into distinct DNA methylation subgroups/subtypes with particular clinico-molecular features. Using RNA-seq in large well annotated cohorts of medulloblastoma we show that transcriptionally Group3 and Group4 medulloblastomas exist not as discrete types but as intermediates on a bipolar continuum between archetypal Group3 and Group4 entities. Continuum position is prognostic, reflects propensity for specific DNA copy-number changes, key switches in isoform/enhancer usage and RNA-editing. Examining scRNA-seq profiles we show intra-tumoral transcriptional heterogeneity along the continuum is limited in a subtype-dependent manner. By integrating with a human scRNA-seq reference atlas we show this continuum is mirrored by an equivalent continuum of transcriptional cell types in early fetal cerebellar development. We identify unique developmental niches for all four major subgroups and link each to a common developmental antecedent. Our findings show a transcriptional continuum arising from oncogenic disruption of highly specific fetal cerebellar cell types, linked to almost every aspect of Group3/Group4 molecular biology and clinico-pathology.

Introduction

The division of medulloblastoma into molecular subgroups has defined the past decade of medulloblastoma research making it all but impossible to interpret new findings except through the prism of these fundamental biological subdivisions. Medulloblastoma was first divided into subgroups on the basis of profiling by expression array (Cho et al., 2011; Fattet et al., 2009; Kool et al., 2008; Northcott et al., 2011; Thompson et al., 2006) and subsequently DNA methylation array (Hovestadt et al., 2014; Schwalbe et al., 2013). The current consensus is that there exist four major medulloblastoma subgroups (MB_{SHH}, MB_{WNT}, MB_{Grp3}, MB_{Grp4}) each with unique clinico-biological characteristics (Taylor et al., 2012); MB_{WNT} and MB_{SHH} are named after characteristic disruptions in the WNT (*CTNNB1* mutation (Ellison et al., 2005)) and SHH (*PTCH*, *SUFU*, *SMO* mutation or *GLI2* amplification (Kool et al., 2014)) pathways respectively. MB_{WNT} denotes an almost entirely curable disease (Ellison et al., 2005) and MB_{SHH} occur more frequently in infants (Kool et al., 2014). The remaining two subgroups Group3 (MB_{Grp3}) and Group4 (MB_{Grp4}) do not exhibit subgroup defining mutations (Northcott et al., 2017) but nonetheless possess distinct clinico-biological characteristics; MB_{Grp3} patients have a greater incidence of "high-risk" features such as LCA (large-cell/anaplastic) histology and *MYC* amplification (Kool et al., 2012; Northcott et al., 2012; Ryan et al., 2012; Taylor et al., 2012). MB_{Grp4} tumors more frequently demonstrate isochromosome 17q (i17q) (Sharma et al., 2019). The advent of routine medulloblastoma molecular subgrouping has enabled the current generation of molecularly driven trials (e.g. NCT02066220, NCT01878617, NCT02724579, NCT01125800) (Li et al., 2019; G. W. Robinson et al., 2015) which exploit MB_{WNT}/MB_{SHH} biology to stratify treatments or direct biological therapeutics.

Further elaborations of the consensus subgroups were published, based primarily upon methylomic definitions (Cavalli et al., 2017; Northcott et al., 2017; Schwalbe et al., 2017). These were followed by a second consensus study which defined 8 subtypes within MB_{Grp3}/MB_{Grp4} named I-VIII; a number of which comprised a mix of MB_{Grp3} and MB_{Grp4} tumors (Sharma et al., 2019). Furthermore, MB_{SHH} can be further divided into subtypes broadly associated with age at diagnosis[10,18,20].

Based on murine modelling, expression and imaging studies (Gibson et al., 2010), MB_{WNT} and MB_{SHH} are believed to derive from two spatially distinct developmental origins in the early hindbrain; lower rhombic lip/dorsal brainstem and upper rhombic lip/early cerebellum, respectively. The developmental origins of MB_{Grp3}

and MB_{Grp4} were investigated in a study mapping subgroup-specific super-enhancer elements, suggesting deep cerebellar nuclei residing in the nuclear transitory zone as the cell of origin for MB_{Grp4} (Lin et al., 2016). More recently, two studies which compared bulk and single-cell transcriptomic (scRNA-seq) MB profiles with developing murine cerebellar scRNA-seq reference datasets described MB_{Grp3} and MB_{Grp4} as most closely resembling Nestin positive stem cells (Vladoiu et al., 2019) and Unipolar Brush Cells (UBC) respectively, highlighting putative cells of origin (Hovestadt et al., 2019; Vladoiu et al., 2019). It is notable that the conclusions of each of these studies rely principally upon cross-species comparisons with murine as opposed to human developmental references. Human rhombic lip development is more complex and prolonged than that of mouse possessing unique features not shared with any other vertebrates (Haldipur et al., 2019).

Here, we characterize the transcriptomic landscape of 331 primary medulloblastoma, with clinico-pathological annotation, DNA methylation and copy number profiles, and we catalogue subgroup-specific isoforms and RNA-editing events. We show that, despite the discrete methylomic subdivisions of the MB_{Grp3}/MB_{Grp4} methylation subtypes I-VIII, these tumors manifest transcriptionally on a bipolar continuum between MB_{Grp3} and MB_{Grp4} archetypes. Moreover, an individual tumor's position on this continuum is predictive of methylation subtype, prognosis, specific copy number and mutational alterations, and activation of key molecular pathways and regulatory events. By using human scRNA-seq fetal cerebellar reference data, we show that this continuum mirrors and recapitulates the major developmental trajectories within early human cerebellar development allowing us to map the interplay between key oncogenic events and putative cells of origin for each medulloblastoma subtype.

Results

Medulloblastoma shows a continuum of expression between MB_{Grp3} and MB_{Grp4}

RNA-seq (~90M paired-end reads) was performed on 331 snap-frozen primary samples from patients with a diagnosis of medulloblastoma (Supplementary Table 1). Transformed gene-level read counts were subject to consensus NMF clustering with resampling to determine the most stable number of clusters and metagenes i.e. major biological effects described by multiple genes and summarized as a single score. As expected, a 4-

metagene/4-cluster solution was optimally stable, reflecting the four major consensus subgroups as currently understood (Figure 1A). ~3% (10/331) of samples were defined as non-classifiable i.e. low probability of classification. ~4% (13/331) samples could only be classified as indeterminate $MB_{Grp3}/MB_{Grp4,}$ i.e. confidently classifiable as either MB_{Grp3} or MB_{Grp4} but not specific as to which. The distribution of clinico-biological features was consistent with previously described features of the consensus MB subgroups (Figure 1A & 1SA); for instance, Chromosome 6 loss in 83% (24/29) of MB_{WNT}.

The two metagenes which described MB_{Grp3} and MB_{Grp4} samples were notably gradated and overlapping in an anti-correlative manner (Figure 1A) implying that, contrary to some previous descriptions using expression microarrays (Cavalli et al., 2017), MB_{Grp3} and MB_{Grp4} are not distinct transcriptional entities but rather exist as a continuum between two transcriptional polarities we refer to here as G3 and G4. To describe this continuum, we created a continuous score (G3/G4-score) scaled between 0 and 1 to reflect the proportionate amount of G3/G4 metagene expression in each MB_{Grp3}/MB_{Grp4} (i.e. all non-WNT/non-SHH tumors) whereby a score of '0' indicates a 100% G4 tumor and '1' 100% G3 (Figure 1A). This was applied to the 223 samples classified as MB_{Grp3}, MB_{Grp4} or intermediate MB_{Grp3}/MB_{Grp4}.

For convenient comparison, we sub-divided the expression continuum (G3/G4-score) into five notional groups: HighG4 (0-0.2, n=69/223 (31%)), LowG4 (0.2-0.4, n=60/223 (27%), G3.5 (0.4-0.6, n=39/223 (17%)), LowG3 (0.6-0.8, n=22/223 (10%)) & HighG3 (0.8-1 G3/G4-score, n=33/223 (15%)). All samples with >0.5 G3/G4 score were classified as $MB_{Grp3.}$ Notably, 15/20 (75%) MB_{Grp3}/MB_{Grp4} samples which showed disagreement in classification between RNA-seq and DNA methylation array were classified as indeterminate MB_{Grp3}/MB_{Grp4} by RNA-seq (Figure 1A). Examining the MB_{Grp3}/MB_{Grp4} subtype (I-VIII) calls by t-SNE (Figure 1B) shows clustering by subtype, suggesting that each methylation subtype imparts distinct secondary expression characteristics beyond the primary G3/G4 continuum metagene. Regardless, the MB_{Grp3}/MB_{Grp4} subtypes may be broadly ordered upon the G3/G4 continuum in partially overlapping domains from most Group4-like to most Group3-like (VIII, VI, VII, V, I, III, IV, II respectively) (Figure 1C).

Specific clinico-biological features were significantly non-randomly distributed across the G3/G4 continuum (Figure 2A). For instance, LCA pathology is significantly enriched at the G3 end of the continuum (D=0.339, p=0.046, n=158); MB_{Grp3} patients display LCA (Large Cell Anaplasia) pathology ~3 times more frequently if

HighG3 as opposed to LowG3. Large (arm level/chromosomal) copy number alterations are likewise non-randomly distributed with respect to the G3/G4 continuum. Most notably, i17q is proportionately greater in HighG4 individuals (75% compared to 36% in Low G4, D=0.402, p<0.001, n = 201) and chromosome 8 gain more frequent in HighG3 (44% in High G3, 5% in Low G3, D=0.69, p<0.001, n=201) (Figure 1S). Mutations are not frequent in MB_{Grp3}/MB_{Grp4} (Northcott et al., 2017), however non-synonymous mutations of *ZMYM3*, *KDM6A* are significantly non-randomly distributed with respect to the continuum (each p<0.01)(Figure 2S).

We examined the relationship between the G3/G4 score and prognosis. Treating the G3/G4 score as categorical showed a progressively poorer 5-year overall survival (OS) across the continuum; Log-rank p=0.026, n=191, HighG3=46%, LowG3=57%, G3.5=71%, LowG4=81%, HighG4=76% (Figure 2B). Cox-regression indicates the continuous G3/G4 score is highly significant (p=0.003, n=191) showing an increase in relative risk of death (RR) of 4.7 times greater for a patient with a G3/G4 score 0 compared to a score of 1. Using the G3/G4 score as a categorical variable, only HighG3 patients have significantly poorer prognosis (RR=2.8, p=0.001, n=191), Multivariable Cox-regression analysis of progression free survival including other risk factors *MYC* amplification, LCA histology and metastatic disease showed that the HighG3 category retains significance (RR=2.4, p=0.014, n=135) indicating the G3/G4 score possesses significant independent prognostic power distinct from its association with other "high-risk" disease features (Figure 2C).

A G3/G4 continuum score can be reverse-engineered from DNA methylation profiles to validate clinico-pathological associations

A series of sample cohorts of MB_{Grp3}/MB_{Grp4} with DNA methylation profiles have previously been published by ourselves and others (Cavalli et al., 2017; Northcott et al., 2017; Schwalbe et al., 2017; Sharma et al., 2019). To these we added 166 profiles to produce a large cohort (n=1670) better powered to validate and further expand our findings made using transcriptomic datasets. We therefore explored the possibility of reverse-engineering a G3/G4 score from DNA methylation data. Using the same method as used for expression was impossible, given that the constrained range (i.e. 0 (fully unmethylated) to 1 (fully methylated)) and bimodal distribution of CpG methylation does not lend itself straightforwardly to a continuous score (Figure 3SA). Unlike expression, which tends to follow a log-linear association with G3/G4 score, methylation follows a sigmoidal distribution from hypo- to hyper-methylation or *vice versa*. The inflection point along the G3/G4 continuum at which these CpGs

"switch" from one state to the other varies by CpG (Figure 3A, 3SB,C). We trained a classifier using a training cohort of MB_{Grp3}/MB_{Grp4} samples for which we possessed both RNA-seq and DNA methylation profiles (n=192). Pre-selecting 400 cross-validated CpG features which distinguish between each of the G3/G4 categorical states we used these to train a random forest classifier to accurately predict (RMSE = 0.036) a G3/G4 score from DNA methylation data alone (Figure 3B).

Using this larger MB_{Grp3}/MB_{Grp4} methylation cohort we were able to demonstrate significant non-random distribution with respect to the continuum of infant status (<3 years), metastases, LCA, and *MYC* amplification (each more frequent in HighG3 patients) and mutations of *PRDM6*, *KDM6A*, *KMT2C* and *ZMYM3* (more frequent in HighG4 patients) (each $p<0.001$, Figure 4SA). Likewise, chromosomal gains of 1q, 5, 6, 8 and 16q (each $p<0.001$ and most frequent in HighG3) and i17q ($p<0.001$ most frequent in HighG4) (Figure 4SA). These findings thus validated our findings from the initial RNA-seq cohort.

The larger cohort size allowed us to also explore the relationship between the G3/G4 continuum and the MB_{Grp3}/MB_{Grp4} subtypes (I-VIII) as well as their previously reported clinico-pathological/mutational characteristics (Sharma et al., 2019). The MB_{Grp3}/MB_{Grp4} subtypes as predicted from DNA methylation data once again occupy discrete but partly overlapping domains within the G3/G4 continuum; broadly ordered - as per the RNA-seq only cohort - from most archetypally MB_{Grp4} to MB_{Grp3} VIII, VI, VII, V, I, III, IV, II (Figure 3C).

We next asked whether the variation in distribution of clinicopathological features and mutation previously described as being characteristic of MB_{Grp3}/MB_{Grp4} subtypes (I-VIII) (Sharma et al., 2019) were attributable to position on the G3/G4 continuum, MB_{Grp3}/MB_{Grp4} subtype (I-VIII) or indeed both. Certain frequent clinicopathological features and copy number changes (e.g. metastatic disease, *MYC* amplification, LCA histology, i17q, loss of chromosome 8, gain of chromosome 5) are significantly non-randomly distributed with respect to G3/G4 continuum, even within individual subtypes (Figure 4SB,C). For example, 100% (11/11) of subtype III with *MYC* amplifications are HighG3 compared to 59% (69/117) without *MYC* amplification. The presence of i17q as the only major chromosomal alteration is a highly characteristic change in subtype VIII but when considering only medulloblastoma subtype VIII is still significantly enriched at the High G4 end of the continuum ($D=0.162$, $p=0.014$). The relative contribution of MB_{Grp3}/MB_{Grp4} subtype and G3/G4 continuum to

predicting presence/absence of a clinicopathological or mutational change was additionally demonstrated by logistic regression (Figure 4SD). This showed that, in several instances (e.g. LCA, *MYC* amplification, i17q as only chromosomal aberration) the response variable was better described using the G3/G4 continuum rather than the medulloblastoma subtype (I-VIII) as an explanatory variable.

The relationship between G3/G4 score and risk of death is significant and striking, allowing us to validate the findings of our RNA-seq cohort with greater confidence; patients age >3 years Log-rank p<0.0001 n=589, HighG3=49%, LowG3=59%, G3.5=64%, LowG4=77%, HighG4=83% (Figure 3D, 5SA). A similar result is found in patients of all ages; Log-rank p<0.0001 n=654 (Figure 5SA). Modelling G3/G4 score as a continuous variable using a Cox proportional hazards again in patients age > 3years shows a 3x increased risk of death from one end of the continuum to the other (RR=3, n=589, p<0.001) (Figure 5SB). MB_{Grp3}/MB_{Grp4} subtypes (I-VIII) are also significantly associated with overall survival (n=524, p<0.001) (Figure 5SC).

The G3/G4 continuum is associated with differential regulation of oncogenic/developmental pathways.

The expression of 590 genes are significantly correlated with the G3/G4 score in our RNA-seq cohort (p<0.01, $\log_2$ fold change>10, n=223), increasing/decreasing log-linearly across the continuum. Most notably, *MYC* expression correlates significantly with G3/G4 score (Rho=0.73, p<0.001, n = 223). HighG3 samples had a mean expression of *MYC* 46 times greater than HighG4 and 2 times greater than LowG3 (Figure 4A). Performing Gene Set Enrichment Analysis (GSEA) we observed that transcriptional targets of MYC were also significantly upregulated (NES = 3.37, p=0.007) (Figure 4B). ssGSEA analysis (Hänzelmann et al., 2013) was used to represent activation/repression of pathway/signatures for each individual and found several oncogenic pathways which were progressively activated or repressed in a manner significantly correlated (each p<0.001) with the G3/G4 continuum including MYC, Cell Cycle, MTOR, TGF-Beta (activated at the G3 pole) and NOTCH (activated at the G4 pole) (Figure 4C). In addition, a broad pattern of progressive neuronal differentiation at the G4 pole and photoreceptor (CRX/NRL) characteristics at the G3 pole of the G3/G4 continuum were observed.

We examined differentially methylated regions (DMRs) within previously identified MB_{Grp3}/MB_{Grp4} specific enhancer loci(Lin et al., 2016), identifying 45 which also overlapped with gene promoters; each “switched” from hypomethylated to hypermethylated or *vice versa* at specific points along the G3/G4 continuum. The

195 expression of 33/45 of these genes are significantly correlated with the G3/G4 continuum ($p<0.01$). This
196 switching appears progressive, with certain MB_{Grp3}/MB_{Grp4} enhancer loci "switching" earlier and others later.
197 For instance, the enhancer/DMR loci overlapping with the promoters of medulloblastoma lineage
198 development/differentiation genes *LHX1*, *NEUROD2*, *LMX1A*, and *HLX* on average "switch" at points 0.23,
199 0.49, 0.56, 0.87 on the G3/G4 continuum (Figure 4D,E). We note also that expression of each of these genes is
200 significantly correlated with the G3/G4 continuum and DMR methylation ($p<0.01$). This is consistent with a
201 developmental identity controlled by cumulative changes in underlying epigenetic architecture driving
202 progressive transition from an MB_{Grp3} to a MB_{Grp4} cell state.

203 **The G3/G4 continuum is associated with post-transcriptional regulation of isoform expression and RNA-**
204 **editing**

205 To explore the clinico-biological significance of differentially expressed transcriptional isoforms across
206 subgroups, Kallisto (Bray et al., 2016) was used to estimate abundance. Taking TPM (Transcripts Per Million)
207 >10 as indicative of a moderate-highly expressed isoform it is notable that the diversity of isoforms being
208 expressed across subgroups was significantly greater in MB_{Grp4} than MB_{Grp3} ($p<0.001$, $F=9.877$) (Figure 6SA).

209 153 genes were identified whose expression overall is invariant but for which expression of specific isoforms
210 correlates significantly with G3/G4 score (Figure 5A). For instance, overall expression of *GTF2I (General*
211 *Transcription Factor IIi)* is ubiquitous but a progressive isoform switch corresponding to the balance between
212 β/δ (*GTF2I-215*/*GTF2I-218*) and α/γ (*GTF2I-221*/*GTF2I-212*) isoforms correlates significantly to G3/G4 score
213 (Figure 5B). These isoforms switches are known to alter protein stability (Shirai et al., 2015) and subcellular
214 localization (Shirai et al., 2017).

215 4,668,508 established RNA editing sites were profiled using the QEdit/Reditools pipeline (Giudice et al., 2020);
216 we observed significant differences in overall A-I editing level. The OEI (Overall Editing Index - i.e. total
217 number of reads with G at all known editing positions over the number of all reads covering the positions)
218 differs significantly with respect to subgroup ($F=9.761$, $n=223$, $p<0.001$); post-hoc testing showed RNA editing
219 events in MB_{Grp4} to be significantly more numerous than MB_{Grp3} and MB_{SHH} (each $p<0.01$) (Figure 6SB).
220 Analysis of 5174 non-synonymous RNA editing sites showed 32 significantly differentially edited with respect
221 to the G3/G4 continuum ($p<0.05$, Figure 5C). One such RNA editing site is *AZIN1* chr8:103841636T>C, known

to result in a S367G substitution which causes conformational changes, cytoplasmic-to-nuclear translocation and gain of function increasing tumor potential in hepatocellular carcinoma (Chen et al., 2013), non-small-cell lung cancer (Hu et al., 2017), colorectal cancer (Shigeyasu et al., 2018) and gastric cancer (Okugawa et al., 2018) (Figure 5D). It is also notable that *ADAR1* and *ADAR2* expression are both correlated with G3/G4 score (rho=0.54, p<0.001 and rho=0.33, p<0.001, n=223, respectively).

Intra-tumoral cellular heterogeneity with respect to the G3/G4 continuum is apparent but constrained by subtype

We projected our MB_{Grp3}/MB_{Grp4} metagenes onto a MB_{Grp3}/MB_{Grp4} scRNA-seq dataset comprising 4,256 cells from 15 individuals (5xSubtype-II, 2xSubtype-III, 1xSubtype-I, 2xSubtype-V, 4xSubtype-VIII) previously published by Hovestadt *et al* (Hovestadt et al., 2019). This allowed us to estimate a G3/G4 score for each cell within a given sample. MB_{Grp3} individuals were described by Hovestadt *et al* as being dominated by cells with an undifferentiated progenitor-like expression program and MB_{Grp4} dominated by a differentiated neuronal-like program; these appear to broadly equate with our MB_{Grp3} and MB_{Grp4} metagenes. MB_{Grp3}/MB_{Grp4} intermediate tumors were described by Hovestadt *et al* as a partial admixture of cell types. The distribution of G3/G4 scores at the single cell level shows that, whilst there is a certain amount of intra-tumoral cellular variation (Figure 6A), the majority of cells fall within the same G3/G4 range observed in the equivalent subtype bulk RNA-seq profiles (Figure 6B). For example, amongst medulloblastoma subtype VIII individuals 78% (667/853) of cells fall within the HighG4 range, 22% (184/853) within LowG4 and 0.2% (2/853) within G3.5. This is equivalent to the bulk profiles for which 83% (40/48) of individuals fall within the HighG4 range and 17% (8/48) fall within the LowG4 range (Figure 6B). In short, the phenomenon of a G3/G4 continuum observed in bulk RNA-seq analysis seems not to be produced by differently proportioned admixtures of otherwise wholly Group3 and Group4 cells but rather medulloblastoma are composed of populations of individual cells, which themselves display continuous G3/G4 expression characteristics; these being constrained to occupy a discrete part of the G3/G4 continuum as dictated by their MB_{Grp3}/MB_{Grp4} (I-VIII) subtype.

Medulloblastoma subtypes and the G3/G4 continuum are mirrored in early human cerebellar development

The origins of medulloblastoma within spatially and temporally distinct regions of the fetal cerebellum (Upper Rhombic Lip/Granule Cell lineage for MB_{SHH} and Lower Rhombic Lip for MB_{WNT}) have been primarily

established by mouse modelling (Gibson et al., 2010; Lin et al., 2016) and, more recently, by comparison with reference to mouse fetal cerebellum scRNA-seq datasets which suggest a unipolar brush cell (UBC) origin for MB_{Grp4} (Hovestadt et al., 2019; Vladoiu et al., 2019). Such comparisons in embryonal tumors are predicated on the idea that partial transformation in an early prenatal cell interrupts development/differentiation resulting in a proportion of the expression characteristics of the tumor initiating cell being retained.

Here, we avoid any cross-species comparisons by using instead a human fetal cerebellum scRNA-seq reference set (69,174 cerebellar cells 9-21 post conception weeks (PCW)). We reconstructed a pseudotemporal cellular trajectory within a broadly defined rhombic lip lineage (12,243 cells, comprising rhombic lip precursors (RL), excitatory cerebellar nuclei/unipolar brush cells (eCN/UBC), granule cell precursors (GCP), and granule cell neurons sub-divided into four clusters (GN)) (Figure 7A). We projected our 4 subgroup metagenes onto these cerebellar cells, identifying those cells which showed highest expression of each metagene. These cells occupy distinct branches of our lineage. High MB_{WNT} metagene expressing cells, as expected, occupy a discrete subset of the RL precursors (Figure 7B). High MB_{Grp3}/MB_{Grp4} metagene expressing cells occupy a distinct eCN/UBC branch beginning with RL precursors (highly expressing MB_{Grp3} metagenes) and transitioning midway to eCN/UBC cells highly expressing the MB_{Grp4} metagene (Figure 7B); this cell trajectory in effect mirrors the G3/G4 continuum. This can be demonstrated formally by calculating a projected per-cell G3/G4 score, revealing a smooth transition from a MB_{Grp3}-like to a MB_{Grp4}-like expression state (Figure 7C). More straightforwardly, this is demonstrated by observing the significant change in expression with respect to pseudotime of those G3/G4 continuum-associated genes whose expression is sufficiently high to be consistently detectable within the relatively low depth scRNA-seq data (each $p<0.01$, Figure 7SA).

Cells which express the MB_{SHH} metagene most highly, as expected, occupy a granule cell developmental branch beginning with GCPs and extending partly into the earliest GN cell types (Figure 7B). Two metagenes representing $MB_{SHH\text{-}Infant}$ (primarily patients <4 years) and $MB_{SHH\text{-}Child}$ (primarily patients > 4 years) - as described in previous studies (Kool et al., 2014; Schwalbe et al., 2017) - were also projected onto the cells in this branch. This indicated a switch midway through the granule cell pseudotemporal lineage from a predominantly $MB_{SHH\text{-}Infant}$ metagene to a predominantly $MB_{SHH\text{-}Child}$ metagene expression; this coincided approximately with the first transition from GCPs to GNs (Figure 7D). Again, where expression of individual

genes which distinguish infant MB_{SHH} from childhood MB_{SHH} were sufficiently detectable within the scRNA-seq profiles they were significantly associated with pseudotime (each $p<0.01$, Figure 7SB).

Thus, by aligning the oncogenic G3/G4 scale with the pseudotemporal scale we were able to order and assign tumorigenic events to specific points within fetal cerebellar developmental lineages (Figure 8). *MYC* amplification, for instance, tends to coincide with the earlier RL pseudotemporal space as opposed to *KDM6A* mutation which occupies the later more differentiated eCN/UBC space. Likewise for aneuploidies, gain of chromosome 8 coincides with the earlier RL developmental space and i17q (as the sole copy number alteration) with the later eCN/UBC cell types.

We note that as well as the pseudotemporal transition from MB_{Grp3} to MB_{Grp4} or $MB_{SHH\text{-}Infant}$ to $MB_{SHH\text{-}Child}$ there is also a literal temporal transition as well. The cerebellar cells most closely associated with the archetypal MB_{Grp3} are predominant at 11PCW (and possibly before). By 18PCW, those most closely associated with the archetypal MB_{Grp4} predominate; this persists until at least 20PCW and between 12PCW and 14PCW, cells associated with G3.5 and LowG4 predominate. This temporal staging from early to late forms of MB_{Grp3}/MB_{Grp4} is also mirrored in the average age of onset of disease. The distribution of age at diagnosis of each MB_{Grp3}/MB_{Grp4} (I-VIII) subtype closely parallels the distribution across the G3/G4 continuum (Figure 8SA) and there is a significant correlation between G3/G4 score and age at diagnosis (Figure 8SB). On the RL to GN branch, cells most closely associated with $MB_{SHH\text{-}Infant}$ are predominant at PCW11 and reduced by PCW20 at which point $MB_{SHH\text{-}Child}$ associated cells predominate (Figure 8B).

We also note that the relative frequencies of $MB_{Grp3,}$ MB_{Grp4} and the distribution of the patient population across the G3/G4 continuum is mirrored in the size of the relevant pseudotemporal space; by space we mean the estimated proportion of putative cells of origin for each subtype (adjusted by number of cells sampled) as a function of the length of time they persist throughout development (9-21 PCW) (Figure 8SC,D). For instance, in our methylation cohort the incidence of MB_{Grp3} relative to MB_{Grp4} is approximately 1:2 (35% 585/1670 MB_{Grp3} vs 65% 1085/1670). The overlap with the relevant proportion of matched pseudotemporal space indicates the same 1:2 ratio (38% MB_{Grp3} matched space vs 62% MB_{Grp4} matched space). The bimodal empirical distribution of the G3/G4 continuum is also paralleled by the pseudotemporal space occupied (Figure 8SD). We regard the above observations as likely explanatory of the relative differences in age of onset and frequency of

303 the different MB_{Grp3}/MB_{Grp4} subtypes and reflective of the relative window of opportunity in terms of time and
304 developmental space for transformation to occur.

305 Discussion

306 Here we show that, with regards to their transcriptomes, the primary inter-tumoral variation in MB_{Grp3}/MB_{Grp4}
307 patients is continuous, in contrast to the discrete nature of the methylation MB_{Grp3}/MB_{Grp4} subtypes (I-VIII)
308 (Cavalli et al., 2017; Northcott et al., 2017; Schwalbe et al., 2013; Sharma et al., 2019). This is not in itself
309 contradictory, as we show that the MB_{Grp3}/MB_{Grp4} methylation subtypes are ordered along the G3/G4 continuum
310 in discrete but partially overlapping domains (Figure 1D). Furthermore, as has been demonstrated previously
311 (Cavalli et al., 2017; Sharma et al., 2019), the methylation subtypes are reflected to some extent in their
312 expression profiles (Figure 1C). Nonetheless, these are shown here to be secondary expression characteristics
313 subordinate to the overarching primary expression characteristic which is the G3/G4 continuum.

314 The position of an individual MB_{Grp3}/MB_{Grp4} tumor upon the continuum is significantly associated with certain
315 mutations, copy number aberrations, clinico- and histopathology. This is to be expected, as many of these have
316 been shown to be non-randomly associated with MB_{Grp3}/MB_{Grp4} subtypes (Sharma et al., 2019). That both
317 methylation subtype and the expression continuum are related to key tumor characteristics and indeed to one
318 another is clear. The question remains to what extent the inter-tumoral variation in such characteristics may be
319 better explained by position upon the continuum than by methylation subtype. For at least some of these
320 characteristics - those which are frequent and not specific to single subtypes (e.g. *MYC* amplification, LCA,
321 i17q, gain of chromosome 5 and loss of chromosome 8) - the answer appears to be that they are better explained
322 by the continuum (Figure 4SB, C, D).

323 The most striking association is between the G3/G4 continuum and risk of death; at least during the first 5 years
324 post-diagnosis. Risk increases continuously with the G3/G4 continuum (Figure 3D), the documented
325 phenomenon (Sharma et al., 2019) of late (>5-year post diagnosis) relapse in subtype VIII notwithstanding. We
326 regard this study as a description of an extremely close and therefore important relationship between biology
327 and clinical course rather than an advocation for its use as a clinical biomarker. Those judgements should be
328 made using prospective clinical trials and the cohort used here, whilst sizeable and carefully reviewed, is a

retrospective cohort with all the limitations and caveats that implies. Nevertheless, we note that when it comes to incorporating molecular data into risk stratification schemes the use of a single G3/G4 risk score for all MB_{Grp3}/MB_{Grp4} patients has a certain pragmatic logic over atomizing a rare cancer into 8 separate subtypes.

Pathway analysis of the G3/G4 continuum shows a concomitant activation of oncogenic processes (e.g. MYC, MTOR, TP53) as tumors become more MB_{Grp3}-like, which itself suggests a more aggressive phenotype. The influence of the G3/G4 continuum also extends to post-transcriptional regulation i.e. isoform usage and RNA-editing. A close relationship with cell differentiation (e.g. CRX/NRL, neuronal differentiation) is also evident and consistent with previous descriptions of MB_{Grp3}/MB_{Grp4} biology cell identity and differentiation (Bandopadhayay et al., 2019; Garancher et al., 2018). This is further reflected in the progressive switches in methylation status we observe within MB_{Grp3}/MB_{Grp4} specific enhancers (Lin et al., 2016).

Importantly, we show here that the MB_{Grp3}/MB_{Grp4} continuum is not produced as the result of an admixture of discrete cell types in different proportions but by individual cells which themselves exist upon the same expression continuum as the bulk tumors. In part, this was observed by Hovestadt *et al* (Hovestadt et al., 2019) in their original analysis of their pooled MB_{Grp3}/MB_{Grp4} scRNA-seq data. They described two metagenes diverging according to *MYC* expression and describe bulk tumors as composed of cells of either a predominately MB_{Grp3}, MB_{Grp4} or intermediate type which themselves represent a continuum of neuronal differentiation (Hovestadt et al., 2019). We have expanded this by fitting individual cells onto the same metagene scale used to define the bulk tumor transcriptome, thereby defining more precisely the range of transcriptional intra-tumoral heterogeneity within MB_{Grp3}/MB_{Grp4} tumors and showing that it appears to be confined to certain limits prescribed by MB_{Grp3}/MB_{Grp4} subtype. This is in turn consistent with the finding that medulloblastoma sampled from different areas of the tumor or at diagnosis and relapse rarely alter subgroup (Kumar et al., 2021; Morrissy et al., 2016; Ramaswamy et al., 2013).

Unlike previous studies which attempted to define cells of origin we used a human rather than a mouse scRNA-seq reference set for comparison. The use of a human atlas is significant because human RL persists longer through cerebellar development than the mouse and has unique cytoarchitectural features not shared with any other vertebrates (Haldipur et al., 2019). Mouse RL is a transient, proliferative stem cell zone present between E12.5 and E17.5 whereas human RL begins as a progenitor niche and is later compartmentalized into ventricular

and subventricular zones, forming a human-specific progenitor pool within the posterior lobule, which persists until birth (Haldipur et al., 2019). We show that the MB_{Grp3}/MB_{Grp4} continuum is paralleled by a fetal cerebellar lineage that begins with an RL progenitor and ends with eCN/UBC. This points to highly specific windows of opportunity - which we define in part here - within developmental pseudotemporal space for oncogenic events to successfully provoke medulloblastoma of a given subtype. These windows of opportunity are highly coherent with both the age of onset of the disease and the relative frequency of the individual subtypes. Importantly, we were able to identify a developmental niche for each of the 4 main medulloblastoma subgroups including a separate space for $MB_{SHH\text{-}Child}$ and $MB_{SHH\text{-}Infant}$. Each of these is contained within a branch of the same early cerebellar lineage explicitly unifying each of the 4 subgroups to a common developmental antecedent; something not reported in previous studies. For instance, Hovestadt *et al* (Hovestadt et al., 2019) were unable to identify a significant matching reference cell type for MB_{Grp3} and MB_{WNT} whereas Vladoiu *et al* (Vladoiu et al., 2019) did not analyze MB_{WNT} and note a prosaic resemblance of MB_{Grp3} to $Nestin^{+}$ early neural stem-like cells. We should note that we were selective, albeit based on prior knowledge, in the subset of cell types we considered as potential candidate cells of origin: figuratively by assigning them to what we broadly described as the Rhombic lip lineage and literally by the physical process of cell extraction and the points in early human development for which sampling was possible (PCW 9-21). MB_{WNT} in particular is thought to originate in the dorsal brainstem, and it may be that certain alternative cells of origin were excluded or curtailed on that basis. Nevertheless, previous studies follow a similar logic to our own and the coherent picture of the relationships between the subgroups would seem to bear out our choices.

In conclusion, our findings point to the following important insights. First, that Group3/Group4 MB and their methylation subtypes exist transcriptionally upon a continuum and that this is mirrored entirely by an equivalent continuum of transcriptional cell types in early human fetal cerebellar development. Second, that by using a human scRNA-seq reference all four MB subtypes can be linked to a common developmental antecedent within the RL lineage. Third, that transcriptional intra-tumoral heterogeneity is limited to certain domains within the continuum as dictated by subtype and finally, that the continuum is linked with almost every aspect of Group3/Group4 molecular biology and clinico-pathology. We anticipate this to have implications for the future

382 treatment and modelling of the disease, most pressingly a need to match cell type with specific timing of
383 mutations to develop faithful models.

384 Acknowledgements

385 Supporting Grants: This work was supported by Cancer Research UK (C8464/A13457 & C8464/A2339), Tom
386 Graham Trust/CCLG and by the INSTINCT network, co-funded by The Brain Tumour Charity, Great Ormond
387 Street Children's Charity, and Children with Cancer UK (grant 16/193). TSJ is grateful for additional funding
388 from the Olivia Hodson Cancer Fund, Cancer Research UK and the National Institute of Health Research. All
389 research at Great Ormond Street Hospital NHS Foundation Trust and UCL Great Ormond Street Institute of
390 Child Health is made possible by the NIHR Great Ormond Street Hospital Biomedical Research Centre. The
391 views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department
392 of Health.

393 Author Contributions

394 Conceptualization D.W., E.C.S., S.B. and S.C.C; Methodology D.W., E.C.S., S.B. and S.C.C; Formal Analysis
395 D.W., E.C.S, K.A.A. and Y.G.; Investigation D.W., E.C.S, D.H, K.A.A, J.C.L, S.C., S.R., J.G., R.M.H., J.C.,
396 S.B.W., T.S.J., A.J., S.B. and S.C.C.; Resources K.A.A., B.P., S.B.W., T.S.J. and A.J.; Data Curation D.W.,
397 E.C.S., D.H., J.C.L., S.R., J.G., R.M.H., J.C., Y.G., and J.H.; Writing – Original Draft D.W., E.C.S and S.C.C;
398 Writing – Review & Editing D.H., K.A.A., S.R., J.G., R.M.H, B.P., and S.B.; Visualization D.W. and J.H.;
399 Supervision D.W., S.B. and S.C.C.; Funding Acquisition D.W., S.B. and S.C.C. All authors approved final
400 version of the manuscript.

401 References

402 Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq--a Python framework to work with high-throughput sequencing data.
403 Bioinformatics 31, 166–169. doi:10.1093/bioinformatics/btu638

404 Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., Irizarry, R.A., 2014. Minfi:
405 a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays.
406 Bioinformatics 30, 1363–1369. doi:10.1093/bioinformatics/btu049

407 Bandopadhayay, P., Piccioni, F., O'Rourke, R., Ho, P., Gonzalez, E.M., Buchan, G., Qian, K., Gionet, G., Girard, E.,
408 Coxon, M., Rees, M.G., Brenan, L., Dubois, F., Shapira, O., Greenwald, N.F., Pages, M., Balboni Iniguez, A., Paolella,
409 B.R., Meng, A., Sinai, C., Roti, G., Dharia, N.V., Creech, A., Tanenbaum, B., Khadka, P., Tracy, A., Tiv, H.L., Hong,
410 A.L., Coy, S., Rashid, R., Lin, J.-R., Cowley, G.S., Lam, F.C., Goodale, A., Lee, Y., Schoolcraft, K., Vazquez, F., Hahn,
411 W.C., Tsherniak, A., Bradner, J.E., Yaffe, M.B., Milde, T., Pfister, S.M., Qi, J., Schenone, M., Carr, S.A., Ligon, K.L.,

Kieran, M.W., Santagata, S., Olson, J.M., Gokhale, P.C., Jaffe, J.D., Root, D.E., Stegmaier, K., Johannessen, C.M., Beroukhim, R., 2019. Neuronal differentiation and cell-cycle programs mediate response to BET-bromodomain inhibition in MYC-driven medulloblastoma. Nat Commun 10, 2400–16. doi:10.1038/s41467-019-10307-9

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527. doi:10.1038/nbt.3519

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420. doi:10.1038/nbt.4096

Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., Kratz, A., Wefers, A.K., Huang, K., Pajtler, K.W., Schweizer, L., Stichel, D., Olar, A., Engel, N.W., Lindenberg, K., Harter, P.N., Braczynski, A.K., Plate, K.H., Dohmen, H., Garvalov, B.K., Coras, R., Hölsken, A., Hewer, E., Bewerunge-Hudler, M., Schick, M., Fischer, R., Beschorner, R., Schittenhelm, J., Staszewski, O., Wani, K., Varlet, P., Pages, M., Temming, P., Lohmann, D., Selt, F., Witt, H., Milde, T., Witt, O., Aronica, E., Giangaspero, F., Rushing, E., Scheurlen, W., Geisenberger, C., Rodriguez, F.J., Becker, A., Preusser, M., Haberler, C., Bjerkvig, R., Cryan, J., Farrell, M., Deckert, M., Hench, J., Frank, S., Serrano, J., Kannan, K., Tsirigos, A., Brück, W., Hofer, S., Brehmer, S., Seiz-Rosenhagen, M., Hänggi, D., Hans, V., Rozsnoki, S., Hansford, J.R., Kohlhof, P., Kristensen, B.W., Lechner, M., Lopes, B., Mawrin, C., Ketter, R., Kulozik, A., Khatib, Z., Heppner, F., Koch, A., Jouvet, A., Keohane, C., Mühleisen, H., Mueller, W., Pohl, U., Prinz, M., Benner, A., Zapatka, M., Gottardo, N.G., Driever, P.H., Kramm, C.M., Müller, H.L., Rutkowski, S., Hoff, von, K., Frühwald, M.C., Gnekow, A., Fleischhack, G., Tippelt, S., Calaminus, G., Monoranu, C.-M., Perry, A., Jones, C., Jacques, T.S., Radlwimmer, B., Gessi, M., Pietsch, T., Schramm, J., Schackert, G., Westphal, M., Reifenberger, G., Wesseling, P., Weller, M., Collins, V.P., Blümcke, I., Bendszus, M., Debus, J., Huang, A., Jabado, N., Northcott, P.A., Paulus, W., Gajjar, A., Robinson, G.W., Taylor, M.D., Jaunmuktane, Z., Ryzhova, M., Platten, M., Unterberg, A., Wick, W., Karajannis, M.A., Mittelbronn, M., Acker, T., Hartmann, C., Aldape, K., Schüller, U., Buslei, R., Lichter, P., Kool, M., Herold-Mende, C., Ellison, D.W., Hasselblatt, M., Snuderl, M., Brandner, S., Korshunov, A., Deimling, von, A., Pfister, S.M., 2018. DNA methylation-based classification of central nervous system tumours. Nature 555, 469–474. doi:10.1038/nature26000

Cavalli, F.M.G., Remke, M., Rampasek, L., Peacock, J., Shih, D.J.H., Luu, B., Garzia, L., Torchia, J., Nor, C., Morrissy, A.S., Agnihotri, S., Thompson, Y.Y., Kuzan-Fischer, C.M., Farooq, H., Isaev, K., Daniels, C., Cho, B.-K., Kim, S.-K., Wang, K.-C., Lee, J.-Y., Grajkowska, W.A., Perek-Polnik, M., Vasiljevic, A., Faure-Conter, C., Jouvet, A., Giannini, C., Nageswara Rao, A.A., Li, K.K.W., Ng, H.-K., Eberhart, C.G., Pollack, I.F., Hamilton, R.L., Gillespie, G.Y., Olson, J.M., Leary, S., Weiss, W.A., Lach, B., Chambless, L.B., Thompson, R.C., Cooper, M.K., Vibhakar, R., Hauser, P., van Veelen, M.-L.C., Kros, J.M., French, P.J., Ra, Y.-S., Kumabe, T., López-Aguilar, E., Zitterbart, K., Sterba, J., Finocchiaro, G., Massimino, M., Van Meir, E.G., Osuka, S., Shofuda, T., Klekner, A., Zollo, M., Leonard, J.R., Rubin, J.B., Jabado, N., Albrecht, S., Mora, J., Van Meter, T.E., Jung, S., Moore, A.S., Hallahan, A.R., Chan, J.A., Tirapelli, D.P.C., Carlotti, C.G., Fouladi, M., Pimentel, J., Faria, C.C., Saad, A.G., Massimi, L., Liau, L.M., Wheeler, H., Nakamura, H., Elbabaa, S.K., Perezpeña-Diazconti, M., Chico Ponce de León, F., Robinson, S., Zapotocky, M., Lassaletta, A., Huang, A., Hawkins, C.E., Tabori, U., Bouffet, E., Bartels, U., Dirks, P.B., Rutka, J.T., Bader, G.D., Reimand, J., Goldenberg, A., Ramaswamy, V., Taylor, M.D., 2017. Intertumoral Heterogeneity within Medulloblastoma Subgroups. Cancer Cell 31, 737–754.e6. doi:10.1016/j.ccell.2017.05.005

Chang, C.H., Housepian, E.M., Herbert, C., 1969. An operative staging system and a megavoltage radiotherapeutic technic for cerebellar medulloblastomas. Radiology 93, 1351–1359. doi:10.1148/93.6.1351

Chen, L., Li, Y., Lin, C.H., Chan, T.H.M., Chow, R.K.K., Song, Y., Liu, M., Yuan, Y.-F., Fu, L., Kong, K.L., Qi, L., Zhang, N., Tong, A.H.Y., Kwong, D.L.-W., Man, K., Lo, C.M., Lok, S., Tenen, D.G., Guan, X.-Y., 2013. Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. Nat. Med. 19, 209–216. doi:10.1038/nm.3043

Cho, Y.-J., Tsherniak, A., Tamayo, P., Santagata, S., Ligon, A., Greulich, H., Berhoukim, R., Amani, V., Goumnerova, L., Eberhart, C.G., Lau, C.C., Olson, J.M., Gilbertson, R.J., Gajjar, A., Delattre, O., Kool, M., Ligon, K., Meyerson, M., Mesirov, J.P., Pomeroy, S.L., 2011. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. J. Clin. Oncol. 29, 1424–1430. doi:10.1200/JCO.2010.28.5148

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. doi:10.1093/bioinformatics/bts635

Ellison, D.W., Onilude, O.E., Lindsey, J.C., Lusher, M.E., Weston, C.L., Taylor, R.E., Pearson, A.D., Clifford, S.C., United Kingdom Children's Cancer Study Group Brain Tumour Committee, 2005. beta-Catenin status predicts a favorable outcome in childhood medulloblastoma: the United Kingdom Children's Cancer Study Group Brain Tumour Committee. J. Clin. Oncol. 23, 7951–7957. doi:10.1200/JCO.2005.01.5479

Fattet, S., Haberler, C., Legoix, P., Varlet, P., Lellouch-Tubiana, A., Lair, S., Manie, E., Raquin, M.-A., Bours, D., Carpentier, S., Barillot, E., Grill, J., Doz, F., Puget, S., Janoueix-Lerosey, I., Delattre, O., 2009. Beta-catenin status in

paediatric medulloblastomas: correlation of immunohistochemical expression with mutational status, genetic profiles, and clinical characteristics. J. Pathol. 218, 86–94. doi:10.1002/path.2514

Garancher, A., Lin, C.Y., Morabito, M., Richer, W., Rocques, N., Larcher, M., Bihannic, L., Smith, K., Miquel, C., Leboucher, S., Herath, N.I., Dupuy, F., Varlet, P., Haberler, C., Walczak, C., Tayara, El, N., Volk, A., Puget, S., Doz, F., Delattre, O., Druillennec, S., Ayrault, O., Wechsler-Reya, R.J., Eychène, A., Bourdeaut, F., Northcott, P.A., Pouponnot, C., 2018. NRL and CRX Define Photoreceptor Identity and Reveal Subgroup-Specific Dependencies in Medulloblastoma. Cancer Cell 33, 435–449.e6. doi:10.1016/j.ccell.2018.02.006

Gibson, P., Tong, Y., Robinson, G., Thompson, M.C., Currle, D.S., Eden, C., Kranenburg, T.A., Hogg, T., Poppleton, H., Martin, J., Finkelstein, D., Pounds, S., Weiss, A., Patay, Z., Scoggins, M., Ogg, R., Pei, Y., Yang, Z.-J., Brun, S., Lee, Y., Zindy, F., Lindsey, J.C., Taketo, M.M., Boop, F.A., Sanford, R.A., Gajjar, A., Clifford, S.C., Roussel, M.F., McKinnon, P.J., Gutmann, D.H., Ellison, D.W., Wechsler-Reya, R., Gilbertson, R.J., 2010. Subtypes of medulloblastoma have distinct developmental origins. Nature 468, 1095–1099. doi:10.1038/nature09587

Giudice, Lo, C., Silvestris, D.A., Roth, S.H., Eisenberg, E., Pesole, G., Gallo, A., Picardi, E., 2020. Quantifying RNA Editing in Deep Transcriptome Datasets. Front Genet 11, 194. doi:10.3389/fgene.2020.00194

Haldipur, P., Aldinger, K.A., Bernardo, S., Deng, M., Timms, A.E., Overman, L.M., Winter, C., Lisgo, S.N., Razavi, F., Silvestri, E., Manganaro, L., Adle-Biassette, H., Guimiot, F., Russo, R., Kidron, D., Hof, P.R., Gerrelli, D., Lindsay, S.J., Dobyns, W.B., Glass, I.A., Alexandre, P., Millen, K.J., 2019. Spatiotemporal expansion of primary progenitor zones in the developing human cerebellum. Science 366, 454–460. doi:10.1126/science.aax7526

Hänzelmann, S., Castelo, R., Guinney, J., 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 14, 7–15. doi:10.1186/1471-2105-14-7

Hovestadt, V., Jones, D.T.W., Picelli, S., Wang, W., Kool, M., Northcott, P.A., Sultan, M., Stachurski, K., Ryzhova, M., Warnatz, H.-J., Ralser, M., Brun, S., Bunt, J., Jäger, N., Kleinheinz, K., Erkek, S., Weber, U.D., Bartholomae, C.C., Kalle, von, C., Lawerenz, C., Eils, J., Koster, J., Versteeg, R., Milde, T., Witt, O., Schmidt, S., Wolf, S., Pietsch, T., Rutkowski, S., Scheurlen, W., Taylor, M.D., Brors, B., Felsberg, J., Reifenberger, G., Borkhardt, A., Lehrach, H., Wechsler-Reya, R.J., Eils, R., Yaspo, M.-L., Landgraf, P., Korshunov, A., Zapatka, M., Radlwimmer, B., Pfister, S.M., Lichter, P., 2014. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. Nature 510, 537–541. doi:10.1038/nature13268

Hovestadt, V., Smith, K.S., Bihannic, L., Filbin, M.G., Shaw, M.L., Baumgartner, A., DeWitt, J.C., Groves, A., Mayr, L., Weisman, H.R., Richman, A.R., Shore, M.E., Goumnerova, L., Rosencrance, C., Carter, R.A., Phoenix, T.N., Hadley, J.L., Tong, Y., Houston, J., Ashmun, R.A., DeCuypere, M., Sharma, T., Flasch, D., Silkov, A., Ligon, K.L., Pomeroy, S.L., Rivera, M.N., Rozenblatt-Rosen, O., Rusert, J.M., Wechsler-Reya, R.J., Li, X.-N., Peyrl, A., Gojo, J., Kirchhofer, D., Lötsch, D., Czech, T., Dorfer, C., Haberler, C., Geyeregger, R., Halfmann, A., Gawad, C., Easton, J., Pfister, S.M., Regev, A., Gajjar, A., Orr, B.A., Slavc, I., Robinson, G.W., Bernstein, B.E., Suvà, M.L., Northcott, P.A., 2019. Resolving medulloblastoma cellular architecture by single-cell genomics. Nature 572, 74–79. doi:10.1038/s41586-019-1434-6

Hu, X., Chen, J., Shi, X., Feng, F., Lau, K.W., Chen, Y., Chen, Y., Jiang, L., Cui, F., Zhang, Y., Xu, X., Li, J., 2017. RNA editing of AZIN1 induces the malignant progression of non-small-cell lung cancers. Tumour Biol 39, 1010428317700001. doi:10.1177/1010428317700001

Kool, M., Jones, D.T.W., Jäger, N., Northcott, P.A., Pugh, T.J., Hovestadt, V., Piro, R.M., Esparza, L.A., Markant, S.L., Remke, M., Milde, T., Bourdeaut, F., Ryzhova, M., Sturm, D., Pfaff, E., Stark, S., Hutter, S., Seker-Cin, H., Johann, P., Bender, S., Schmidt, C., Rausch, T., Shih, D., Reimand, J., Sieber, L., Wittmann, A., Linke, L., Witt, H., Weber, U.D., Zapatka, M., König, R., Beroukhim, R., Bergthold, G., van Sluis, P., Volckmann, R., Koster, J., Versteeg, R., Schmidt, S., Wolf, S., Lawerenz, C., Bartholomae, C.C., Kalle, von, C., Unterberg, A., Herold-Mende, C., Hofer, S., Kulozik, A.E., Deimling, von, A., Scheurlen, W., Felsberg, J., Reifenberger, G., Hasselblatt, M., Crawford, J.R., Grant, G.A., Jabado, N., Perry, A., Cowdrey, C., Croul, S., Zadeh, G., Korbel, J.O., Doz, F., Delattre, O., Bader, G.D., McCabe, M.G., Collins, V.P., Kieran, M.W., Cho, Y.-J., Pomeroy, S.L., Witt, O., Brors, B., Taylor, M.D., Schüller, U., Korshunov, A., Eils, R., Wechsler-Reya, R.J., Lichter, P., Pfister, S.M., ICGC PedBrain Tumor Project, 2014. Genome sequencing of SHH medulloblastoma predicts genotype-related response to smoothened inhibition. Cancer Cell 25, 393–405. doi:10.1016/j.ccr.2014.02.004

Kool, M., Korshunov, A., Remke, M., Jones, D.T.W., Schlanstein, M., Northcott, P.A., Cho, Y.-J., Koster, J., Schouten-van Meeteren, A., van Vuurden, D., Clifford, S.C., Pietsch, T., Bueren, von, A.O., Rutkowski, S., McCabe, M., Collins, V.P., Bäcklund, M.L., Haberler, C., Bourdeaut, F., Delattre, O., Doz, F., Ellison, D.W., Gilbertson, R.J., Pomeroy, S.L., Taylor, M.D., Lichter, P., Pfister, S.M., 2012. Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. Acta Neuropathol. 123, 473–484. doi:10.1007/s00401-012-0958-8

Kool, M., Koster, J., Bunt, J., Hasselt, N.E., Lakeman, A., van Sluis, P., Troost, D., Meeteren, N.S.-V., Caron, H.N., Cloos, J., Mrsić, A., Ylstra, B., Grajkowska, W., Hartmann, W., Pietsch, T., Ellison, D., Clifford, S.C., Versteeg, R., 2008.

Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. PLoS ONE 3, e3088. doi:10.1371/journal.pone.0003088

Kumar, R., Smith, K.S., Deng, M., Terhune, C., Robinson, G.W., Orr, B.A., Liu, A.P.Y., Lin, T., Billups, C.A., Chintagumpala, M., Bowers, D.C., Hassall, T.E., Hansford, J.R., Khuong-Quang, D.-A., Crawford, J.R., Bendel, A.E., Gururangan, S., Schroeder, K., Bouffet, E., Bartels, U., Fisher, M.J., Cohn, R., Partap, S., Kellie, S.J., McCowage, G., Paulino, A.C., Rutkowski, S., Fleischhack, G., Dhall, G., Klesse, L.J., Leary, S., Nazarian, J., Kool, M., Wesseling, P., Ryzhova, M., Zheludkova, O., Golanov, A.V., McLendon, R.E., Packer, R.J., Dunham, C., Hukin, J., Fouladi, M., Faria, C.C., Pimentel, J., Walter, A.W., Jabado, N., Cho, Y.-J., Perreault, S., Croul, S.E., Zapotocky, M., Hawkins, C., Tabori, U., Taylor, M.D., Pfister, S.M., Klimo, P., Boop, F.A., Ellison, D.W., Merchant, T.E., Onar-Thomas, A., Korshunov, A., Jones, D.T.W., Gajjar, A., Ramaswamy, V., Northcott, P.A., 2021. Clinical Outcomes and Patient-Matched Molecular Composition of Relapsed Medulloblastoma. J. Clin. Oncol. 39, 807–821. doi:10.1200/JCO.20.01359

Li, Y., Song, Q., Day, B.W., 2019. Phase I and phase II sonidegib and vismodegib clinical trials for the treatment of paediatric and adult MB patients: a systemic review and meta-analysis. Acta Neuropathol Commun 7, 123–8. doi:10.1186/s40478-019-0773-8

Lin, C.Y., Erkek, S., Tong, Y., Yin, L., Federation, A.J., Zapatka, M., Haldipur, P., Kawauchi, D., Risch, T., Warnatz, H.-J., Worst, B.C., Ju, B., Orr, B.A., Zeid, R., Polaski, D.R., Segura-Wang, M., Waszak, S.M., Jones, D.T.W., Kool, M., Hovestadt, V., Buchhalter, I., Sieber, L., Johann, P., Chavez, L., Gröschel, S., Ryzhova, M., Korshunov, A., Chen, W., Chizhikov, V.V., Millen, K.J., Amstislavskiy, V., Lehrach, H., Yaspo, M.-L., Eils, R., Lichter, P., Korbel, J.O., Pfister, S.M., Bradner, J.E., Northcott, P.A., 2016. Active medulloblastoma enhancers reveal subgroup-specific cellular origins. Nature 530, 57–62. doi:10.1038/nature16546

Lopes, M.C., Joyce, C., Ritchie, G.R.S., John, S.L., Cunningham, F., Asimit, J., Zeggini, E., 2012. A combined functional annotation score for non-synonymous variants. Hum Hered 73, 47–51. doi:10.1159/000334984

Louis, D.N., Perry, A., Reifenberger, G., Deimling, von, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W., 2016. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 131, 803–820. doi:10.1007/s00401-016-1545-1

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. doi:10.1186/s13059-014-0550-8

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F., 2016. The Ensembl Variant Effect Predictor. Genome Biol. 17, 122–14. doi:10.1186/s13059-016-0974-4

Morrissy, A.S., Garzia, L., Shih, D.J.H., Zuyderduyn, S., Huang, X., Skowron, P., Remke, M., Cavalli, F.M.G., Ramaswamy, V., Lindsay, P.E., Jelveh, S., Donovan, L.K., Wang, X., Luu, B., Zayne, K., Li, Y., Mayoh, C., Thiessen, N., Mercier, E., Mungall, K.L., Ma, Y., Tse, K., Zeng, T., Shumansky, K., Roth, A.J.L., Shah, S., Farooq, H., Kijima, N., Holgado, B.L., Lee, J.J.Y., Matan-Lithwick, S., Liu, J., Mack, S.C., Manno, A., Michealraj, K.A., Nor, C., Peacock, J., Qin, L., Reimand, J., Rolider, A., Thompson, Y.Y., Wu, X., Pugh, T., Ally, A., Bilenky, M., Butterfield, Y.S.N., Carlsen, R., Cheng, Y., Chuah, E., Corbett, R.D., Dhalla, N., He, A., Lee, D., Li, H.I., Long, W., Mayo, M., Plettner, P., Qian, J.Q., Schein, J.E., Tam, A., Wong, T., Birol, I., Zhao, Y., Faria, C.C., Pimentel, J., Nunes, S., Shalaby, T., Grotzer, M., Pollack, I.F., Hamilton, R.L., Li, X.-N., Bendel, A.E., Fults, D.W., Walter, A.W., Kumabe, T., Tominaga, T., Collins, V.P., Cho, Y.-J., Hoffman, C., Lyden, D., Wisoff, J.H., Garvin, J.H., Stearns, D.S., Massimi, L., Schüller, U., Sterba, J., Zitterbart, K., Puget, S., Ayrault, O., Dunn, S.E., Tirapelli, D.P.C., Carlotti, C.G., Wheeler, H., Hallahan, A.R., Ingram, W., MacDonald, T.J., Olson, J.J., Van Meir, E.G., Lee, J.-Y., Wang, K.-C., Kim, S.-K., Cho, B.-K., Pietsch, T., Fleischhack, G., Tippelt, S., Ra, Y.-S., Bailey, S., Lindsey, J.C., Clifford, S.C., Eberhart, C.G., Cooper, M.K., Packer, R.J., Massimino, M., Garre, M.L., Bartels, U., Tabori, U., Hawkins, C.E., Dirks, P., Bouffet, E., Rutka, J.T., Wechsler-Reya, R.J., Weiss, W.A., Collier, L.S., Dupuy, A.J., Korshunov, A., Jones, D.T.W., Kool, M., Northcott, P.A., Pfister, S.M., Largaespada, D.A., Mungall, A.J., Moore, R.A., Jabado, N., Bader, G.D., Jones, S.J.M., Malkin, D., Marra, M.A., Taylor, M.D., 2016. Divergent clonal selection dominates medulloblastoma at recurrence. Nature 529, 351–357. doi:10.1038/nature16478

Northcott, P.A., Buchhalter, I., Morrissy, A.S., Hovestadt, V., Weischenfeldt, J., Ehrenberger, T., Gröbner, S., Segura-Wang, M., Zichner, T., Rudneva, V.A., Warnatz, H.-J., Sidiropoulos, N., Phillips, A.H., Schumacher, S., Kleinheinz, K., Waszak, S.M., Erkek, S., Jones, D.T.W., Worst, B.C., Kool, M., Zapatka, M., Jäger, N., Chavez, L., Hutter, B., Bieg, M., Paramasivam, N., Heinold, M., Gu, Z., Ishaque, N., Jäger-Schmidt, C., Imbusch, C.D., Jugold, A., Hübschmann, D., Risch, T., Amstislavskiy, V., Gonzalez, F.G.R., Weber, U.D., Wolf, S., Robinson, G.W., Zhou, X., Wu, G., Finkelstein, D., Liu, Y., Cavalli, F.M.G., Luu, B., Ramaswamy, V., Wu, X., Koster, J., Ryzhova, M., Cho, Y.-J., Pomeroy, S.L., Herold-Mende, C., Schuhmann, M., Ebinger, M., Liau, L.M., Mora, J., McLendon, R.E., Jabado, N., Kumabe, T., Chuah, E., Ma, Y., Moore, R.A., Mungall, A.J., Mungall, K.L., Thiessen, N., Tse, K., Wong, T., Jones, S.J.M., Witt, O., Milde, T., Deimling, von, A., Capper, D., Korshunov, A., Yaspo, M.-L., Kriwacki, R., Gajjar, A., Zhang, J., Beroukhim, R., Fraenkel, E., Korbel, J.O., Brors, B., Schlesner, M., Eils, R., Marra, M.A., Pfister, S.M., Taylor, M.D., Lichter, P., 2017. The whole-genome landscape of medulloblastoma subtypes. Nature 547, 311–317. doi:10.1038/nature22973

Northcott, P.A., Jones, D.T.W., Kool, M., Robinson, G.W., Gilbertson, R.J., Cho, Y.-J., Pomeroy, S.L., Korshunov, A., Lichter, P., Taylor, M.D., Pfister, S.M., 2012. Medulloblastomics: the end of the beginning. Nat. Rev. Cancer 12, 818–834. doi:10.1038/nrc3410

Northcott, P.A., Korshunov, A., Witt, H., Hielscher, T., Eberhart, C.G., Mack, S., Bouffet, E., Clifford, S.C., Hawkins, C.E., French, P., Rutka, J.T., Pfister, S., Taylor, M.D., 2011. Medulloblastoma comprises four distinct molecular variants. J. Clin. Oncol. 29, 1408–1414. doi:10.1200/JCO.2009.27.4324

Northcott, P.A., Lee, C., Zichner, T., Stütz, A.M., Erkek, S., Kawauchi, D., Shih, D.J.H., Hovestadt, V., Zapatka, M., Sturm, D., Jones, D.T.W., Kool, M., Remke, M., Cavalli, F.M.G., Zuyderduyn, S., Bader, G.D., VandenBerg, S., Esparza, L.A., Ryzhova, M., Wang, W., Wittmann, A., Stark, S., Sieber, L., Seker-Cin, H., Linke, L., Kratochwil, F., Jäger, N., Buchhalter, I., Imbusch, C.D., Zipprich, G., Raeder, B., Schmidt, S., Diessl, N., Wolf, S., Wiemann, S., Brors, B., Lawerenz, C., Eils, J., Warnatz, H.-J., Risch, T., Yaspo, M.-L., Weber, U.D., Bartholomae, C.C., Kalle, von, C., Turányi, E., Hauser, P., Sandén, E., Darabi, A., Siesjö, P., Sterba, J., Zitterbart, K., Sumerauer, D., van Sluis, P., Versteeg, R., Volckmann, R., Koster, J., Schuhmann, M.U., Ebinger, M., Grimes, H.L., Robinson, G.W., Gajjar, A., Mynarek, M., Hoff, von, K., Rutkowski, S., Pietsch, T., Scheurlen, W., Felsberg, J., Reifenberger, G., Kulozik, A.E., Deimling, von, A., Witt, O., Eils, R., Gilbertson, R.J., Korshunov, A., Taylor, M.D., Lichter, P., Korbel, J.O., Wechsler-Reya, R.J., Pfister, S.M., 2014. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature 511, 428–434. doi:10.1038/nature13379

Okugawa, Y., Toiyama, Y., Shigeyasu, K., Yamamoto, A., Shigemori, T., Yin, C., Ichikawa, T., Yasuda, H., Fujikawa, H., Yoshiyama, S., Hiro, J., Ohi, M., Araki, T., Kusunoki, M., Goel, A., 2018. Enhanced AZIN1 RNA editing and overexpression of its regulatory enzyme ADAR1 are important prognostic biomarkers in gastric cancer. J Transl Med 16, 366–11. doi:10.1186/s12967-018-1740-z

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A., Trapnell, C., 2017. Single-cell mRNA quantification and differential analysis with Census. Nat. Methods 14, 309–315. doi:10.1038/nmeth.4150

Ramaswamy, V., Remke, M., Bouffet, E., Faria, C.C., Perreault, S., Cho, Y.-J., Shih, D.J., Luu, B., Dubuc, A.M., Northcott, P.A., Schüller, U., Gururangan, S., McLendon, R., Bigner, D., Fouladi, M., Ligon, K.L., Pomeroy, S.L., Dunn, S., Triscott, J., Jabado, N., Fontebasso, A., Jones, D.T.W., Kool, M., Karajannis, M.A., Gardner, S.L., Zagzag, D., Nunes, S., Pimentel, J., Mora, J., Lipp, E., Walter, A.W., Ryzhova, M., Zheludkova, O., Kumirova, E., Alshami, J., Croul, S.E., Rutka, J.T., Hawkins, C., Tabori, U., Codispoti, K.-E.T., Packer, R.J., Pfister, S.M., Korshunov, A., Taylor, M.D., 2013. Recurrence patterns across medulloblastoma subgroups: an integrated clinical and molecular analysis. Lancet Oncol. 14, 1200–1207. doi:10.1016/S1470-2045(13)70449-2

Robinson, G.W., Orr, B.A., Wu, G., Gururangan, S., Lin, T., Qaddoumi, I., Packer, R.J., Goldman, S., Prados, M.D., Desjardins, A., Chintagumpala, M., Takebe, N., Kaste, S.C., Rusch, M., Allen, S.J., Onar-Thomas, A., Stewart, C.F., Fouladi, M., Boyett, J.M., Gilbertson, R.J., Curran, T., Ellison, D.W., Gajjar, A., 2015. Vismodegib Exerts Targeted Efficacy Against Recurrent Sonic Hedgehog-Subgroup Medulloblastoma: Results From Phase II Pediatric Brain Tumor Consortium Studies PBTC-025B and PBTC-032. J. Clin. Oncol. 33, 2646–2654. doi:10.1200/JCO.2014.60.1591

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. doi:10.1038/nbt.1754

Ryan, S.L., Schwalbe, E.C., Cole, M., Lu, Y., Lusher, M.E., Megahed, H., O'Toole, K., Nicholson, S.L., Bognár, L., Garami, M., Hauser, P., Korshunov, A., Pfister, S.M., Williamson, D., Taylor, R.E., Ellison, D.W., Bailey, S., Clifford, S.C., 2012. MYC family amplification and clinical risk-factors interact to predict an extremely poor prognosis in childhood medulloblastoma. Acta Neuropathol. 123, 501–513. doi:10.1007/s00401-011-0923-y

Schwalbe, E.C., Lindsey, J.C., Nakjang, S., Crosier, S., Smith, A.J., Hicks, D., Rafiee, G., Hill, R.M., Iliasova, A., Stone, T., Pizer, B., Michalski, A., Joshi, A., Wharton, S.B., Jacques, T.S., Bailey, S., Williamson, D., Clifford, S.C., 2017. Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study. Lancet Oncol. doi:10.1016/S1470-2045(17)30243-7

Schwalbe, E.C., Williamson, D., Lindsey, J.C., Hamilton, D., Ryan, S.L., Megahed, H., Garami, M., Hauser, P., Dembowska-Baginska, B., Perek, D., Northcott, P.A., Taylor, M.D., Taylor, R.E., Ellison, D.W., Bailey, S., Clifford, S.C., 2013. DNA methylation profiling of medulloblastoma allows robust subclassification and improved outcome prediction using formalin-fixed biopsies. Acta Neuropathol. 125, 359–371. doi:10.1007/s00401-012-1077-2

Sharma, T., Schwalbe, E.C., Williamson, D., Sill, M., Hovestadt, V., Mynarek, M., Rutkowski, S., Robinson, G.W., Gajjar, A., Cavalli, F., Ramaswamy, V., Taylor, M.D., Lindsey, J.C., Hill, R.M., Jäger, N., Korshunov, A., Hicks, D., Bailey, S., Kool, M., Chavez, L., Northcott, P.A., Pfister, S.M., Clifford, S.C., 2019. Second-generation molecular subgrouping of medulloblastoma: an international meta-analysis of Group 3 and Group 4 subtypes. Acta Neuropathol. 34, 396–18. doi:10.1007/s00401-019-02020-0

633 Shigeyasu, K., Okugawa, Y., Toden, S., Miyoshi, J., Toiyama, Y., Nagasaka, T., Takahashi, N., Kusunoki, M., Takayama,
634 T., Yamada, Y., Fujiwara, T., Chen, L., Goel, A., 2018. AZIN1 RNA editing confers cancer stemness and enhances
635 oncogenic potential in colorectal cancer. JCI Insight 3. doi:10.1172/jci.insight.99976

636 Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L.A., Edwards, K.J., Day, I.N.M., Gaunt, T.R., 2013.
637 Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov
638 models. Hum Mutat 34, 57–65. doi:10.1002/humu.22225

639 Shirai, Y., Li, W., Suzuki, T., 2017. Role of Splice Variants of Gtf2i, a Transcription Factor Localizing at Postsynaptic
640 Sites, and Its Relation to Neuropsychiatric Diseases. Int J Mol Sci 18, 411. doi:10.3390/ijms18020411

641 Shirai, Y., Watanabe, M., Sakagami, H., Suzuki, T., 2015. Novel splice variants in the 5“UTR of Gtf2i expressed in the rat
642 brain: alternative 5”UTRs and differential expression in the neuronal dendrites. J Neurochem 134, 578–589.
643 doi:10.1111/jnc.13136

644 Tamayo, P., Scanfeld, D., Ebert, B.L., Gillette, M.A., Roberts, C.W.M., Mesirov, J.P., 2007. Metagene projection for cross-
645 platform, cross-species characterization of global transcriptional states. Proc. Natl. Acad. Sci. U.S.A. 104, 5959–5964.
646 doi:10.1073/pnas.0701068104

647 Taylor, M.D., Northcott, P.A., Korshunov, A., Remke, M., Cho, Y.-J., Clifford, S.C., Eberhart, C.G., Parsons, D.W.,
648 Rutkowski, S., Gajjar, A., Ellison, D.W., Lichter, P., Gilbertson, R.J., Pomeroy, S.L., Kool, M., Pfister, S.M., 2012.
649 Molecular subgroups of medulloblastoma: the current consensus. Acta Neuropathol. 123, 465–472. doi:10.1007/s00401-
650 011-0922-z

651 Thompson, M.C., Fuller, C., Hogg, T.L., Dalton, J., Finkelstein, D., Lau, C.C., Chintagumpala, M., Adesina, A., Ashley,
652 D.M., Kellie, S.J., Taylor, M.D., Curran, T., Gajjar, A., Gilbertson, R.J., 2006. Genomics identifies medulloblastoma
653 subgroups that are enriched for specific genetic alterations. J. Clin. Oncol. 24, 1924–1931. doi:10.1200/JCO.2005.04.4974

654 Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K.,
655 Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ data to
656 high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43,
657 11.10.1–11.10.33. doi:10.1002/0471250953.bi1110s43

658 Vladoiu, M.C., El-Hamamy, I., Donovan, L.K., Farooq, H., Holgado, B.L., Sundaravadanam, Y., Ramaswamy, V.,
659 Hendrikse, L.D., Kumar, S., Mack, S.C., Lee, J.J.Y., Fong, V., Juraschka, K., Przelicki, D., Michealraj, A., Skowron, P.,
660 Luu, B., Suzuki, H., Morrissy, A.S., Cavalli, F.M.G., Garzia, L., Daniels, C., Wu, X., Qazi, M.A., Singh, S.K., Chan, J.A.,
661 Marra, M.A., Malkin, D., Dirks, P., Heisler, L., Pugh, T., Ng, K., Notta, F., Thompson, E.M., Kleinman, C.L., Joyner,
662 A.L., Jabado, N., Stein, L., Taylor, M.D., 2019. Childhood cerebellar tumours mirror conserved fetal transcriptional
663 programs. Nature 572, 67–73. doi:10.1038/s41586-019-1158-7

Figure Legends

Figure 1

A: Heatmap showing 4 consensus NMF metagenes calculated for n=331 MB and grouped by subgroup. MB_{Grp3}/MB_{Grp4} individuals are ordered by G3/G4 score. Annotation shows subgroup as determined by RNA-seq (Expression Subgroup), subgroup as determined by methylation (Methylation Subgroup), methylation MB_{Grp3}/MB_{Grp4} subtype (I-VIII) as per Sharma *et al* 2019(Sharma et al., 2019) defined using MNPv2 classifier(Capper et al., 2018) (Grp3/4 Subtype). All other characteristics are indicated to be present or not by dark grey shading according to the following scheme: Infant = age at diagnosis < 3 years, Adult = age at diagnosis > 16 years, DN = Desmoplastic/Nodular, LCA = Large-cell/anaplastic, STR = Sub-Total Resection, DOD=Dead of Disease. Side annotation (top left) shows a heatmap of chi-square residuals indicating subgroup enrichment and significance where relevant. The line plot (bottom) shows G3/G4 score. **B:** t-SNE plot showing MB_{Grp3}/MB_{Grp4} samples shaded by subgroup (top) and methylation MB_{Grp3}/MB_{Grp4} subtype (I-VIII) (bottom). **C:** Violin plot showing G3/G4 score by MB_{Grp3}/MB_{Grp4} subtype (I-VIII).

Figure 2

A: Rug plot showing distribution of clinico-path features with respect to G3/G4 score. Summary counts are given according to the divisions of HighG4, LowG4, G3.5, LowG3, HighG3 and reflected by the red line plots. Presence of a feature is indicated by a bold tick mark the color of which indicates methylation MB_{Grp3}/MB_{Grp4} subtype (I-VIII). Adjusted p-values for a Kolmogorov-Smirnoff statistic (D) are shown to denote non-random distribution of features with respect to G3/G4 score. Mismatch=mismatch between methylation and expression call, Infant=age at diagnosis < 3 years, M+=Metastatic, DOD=Dead of Disease, LCA = Large-Cell/anaplastic, PRDM6 = *PRDM6* rearrangement. **B:** Kaplan-Meier plot showing significant differences in MB_{Grp3}/MB_{Grp4} overall survival by G3/G4 continuum position. **C:** Forest plot showing a multivariate Cox model fitted to progression free survival and containing the independently significant variables High G3, *MYC* amplification, LCA and M+.

Figure 3

A: Fitted sigmoid curve representing the relationship between CpG beta-value and G3/G4 Score. Top 40 most discriminatory CpGs distinguishing HighG4 (dark green), LowG4 (light green), Low G3 (yellow) and High G3 (orange) are shown. **B:** The performance of the cross-validated random forest classifier showing predicted G3/G4 score (derived from DNA methylation values) against actual G3/G4 score (derived from RNA-seq) n = 192. **C:** Violin plot showing G3/G4 score (derived from methylation) by MB_{Grp3}/MB_{Grp4} (I-VIII) subtype. **D:** Kaplan-Meier plot showing significant differences in MB_{Grp3}/MB_{Grp4} overall survival in patients aged > 3 years by G3/G4 score (as derived from methylation values) n = 589.

Figure 4

A: Scatterplot showing significant correlation ($p<0.001$) between *MYC* expression and G3/G4 score. Log-linear line of best fit is shown. Dotted lines divide into High G4, Low G4, G3.5, Low G3, High G3 and Log2 fold change for each category relative to HighG4 are shown. **B:** GSEA enrichment plot showing significant enrichment of MYC target genes. Genes were ranked by correlation with G3/G4 score. **C:** Heatmap of ssGSEA results showing level of pathway enrichment for 223 MB_{Grp3}/MB_{Grp4} individuals ordered by G3/G4 score. MsigDB pathways are curated into pathways (see methods). **D:** Lollipop plot showing mean beta fold change for DMRs within MB_{Grp3}/MB_{Grp4} specific enhancers/superenhancers. The position on the x axis reflects the average point on the continuum at which the methylation level switches from hypo- to hypermethylation. **E:** Plot showing an MB_{Grp3}/MB_{Grp4} specific enhancer within the MB_{Grp3} specific gene *LMX1A* which overlaps with a differentially methylated region significantly associated with the G3/G4 continuum. The mean beta value per G3/G4 category (High G4, Low G4, G3.5, Low G3, High G3) and MB_{Grp3}/MB_{Grp4} subtype (I-VIII) is shown by line and the 95%CI by shaded area.

Figure 5

A: Heatmap showing expression of top significantly differentially expressed isoforms of genes whose overall expression is otherwise not significantly differentially expressed with respect to G3/G4 score. **B:** Schematic showing exon structure of 4 *GTF2I* isoforms significantly differentially expressed with respect to G3/G4 score (left) and scatterplot showing expression of these *GTF2I* isoforms vs G3/G4 score; line represents fitted log-linear model NB: *GTF2I* is not significantly differentially expressed at the gene level. **C:** Top 10 significantly

715 differentially edited non-synonymous RNA editing positions with respect to G3/G4 score. **D:** Boxplot showing
716 level of T>C RNA editing at a non-synonymous position S367G within *AZIN1;* level of editing is significantly
717 associated with G3/G4 score.

718 **Figure 6**

719 **A:** Violin plot showing per-cell G3/G4 score (derived from projection onto scRNA-seq data) for 15
720 MB_{Grp3}/MB_{Grp4} patients aggregated by subtype. **B:** Ridgeplot showing distribution of per-cell G3/G4 score
721 (derived from projection onto scRNA-seq data) for each of 15 MB_{Grp3}/MB_{Grp4} patients shown alongside the
722 G3/G4 score distribution of equivalent subtype bulk tumors. n=x refers to number of individuals for bulk tumors
723 and number of cells for the scRNA-seq data.

724 **Figure 7**

725 **A:** UMAP plot of scRNA-seq profiles showing 12,243 cells of the rhombic lip lineage arranged according to
726 developmental trajectory which is indicated by black line. Color denotes cell type as determined by graph based
727 clustering; RL=Rhombic Lip precursors, GCP=Granule Cell Precursors, GN-I, GN-II, GN-III, GN-IV = four
728 Granule Neuron cell types, eCN/UBC=excitory Cerebellar Neurons/Unipolar Brush Cells. **B:** UMAP plot of
729 the rhombic lip lineage with those cells within the top decile of metagene expression are marked with the
730 following colors: MB_{Grp4}=green, MB_{Grp3}=yellow, MB_{SHH}=red, MB_{WNT}=blue. **C:** Scatterplot showing per-cell
731 scaled metagene expression along the RL to eCN/UBC branch. Fitted sigmoid curves are shown with SD
732 indicated as dashed lines. The grey line represents a sigmoid curve fitted to per-cell G3/G4 score as a function
733 of pseudotime. **D:** Scatterplot showing per-cell scaled metagene expression along the GCP to GN branch. Fitted
734 curves are shown with SD shown as dashed lines. Curves are scaled to be constrained to a range of 0 and 1, in
735 order to be coherent with bulk analysis. For this reason, by definition, some individual cells lie outside the 0
736 and 1 range.

737 **Figure 8**

738 **A:** Schema showing the RL to eCN/UBC developmental branch, the relationship between pseudotime and
739 G3/G4 score and the staging of key tumor characteristics. From top to bottom: a violin plot showing pseudotime
740 distribution of cells by time of sampling, color transition red to purple marks the point along the developmental
741 trajectory where cells are defined as eCN/UBC. A fitted sigmoid curve showing the relationship between

pseudotime and G3/G4 score. Tumor characteristics are transformed from the G3/G4 scale to the pseudotime scale and marked at the appropriate points. Color bars represent subgroups. Methylation subtypes (I-VIII), mutations and copy number changes are marked by box and whisker. Dot represents median distribution thick line represent the interquartile range and the thinner lines corresponds to range. Dotted lines denote where the range extends up to a G3/G4 score of 0 and 1, i.e. matching the *ne plus ultra* pseudotime after which G3/G4 score is unchanged and exact relationship must be extrapolated. **B:** Schematic showing the GCP to GN developmental branch, the relationship between pseudotime and $MB_{SHH\text{-}Infant}$ or $MB_{SHH\text{-}Child}$ metagene. From top to bottom: a violin plot showing pseudotime distribution of cells by time of sample, color transition green to blue marks the point along the developmental trajectory where cells become defined as GN. A loess curve showing the relationship between pseudotime and $MB_{SHH\text{-}Infant}$ (red) or $MB_{SHH\text{-}Child}$ metagene (dark red). Color bars show parts of trajectory paralleled by $MB_{SHH\text{-}Infant}$ or $MB_{SHH\text{-}Child}$ tumors. **C:** UMAP of developmental trajectory marked with colors to denote parts most associated with each MB subgroup and the relevant pseudotime (PsT) scale.

Figure 1S

A: Heatmap showing copy number changes by chromosome for n=331 MB and grouped by subgroup. MB_{Grp3}/MB_{Grp4} individuals are ordered by G3/G4 score. Column annotation shows subgroup as determined by RNA-seq (Expression Subgroup) as determined by DNA methylation array (Methylation Subgroup), and DNA methylation MB_{Grp3}/MB_{Grp4} subtype (I-VIII) as per Sharma *et al* 2019(Sharma et al., 2019) as defined using MNPv2 classifier(Capper et al., 2018) (Grp3/4 Subtype). The line plot (bottom) shows G3/G4 score. **B:** Rug plot showing distribution of aneuploidy/copy number change with respect to G3/G4 score. Summary counts are given according to the convenient divisions of HighG4, LowG4, G3.5, LowG3, HighG3 and reflected by the red line plots. Presence of a given feature is indicated by a bold tick mark, the color of which indicates methylation MB_{Grp3}/MB_{Grp4} subtype (I-VIII). Adjusted P-values for a Kolmogorov-Smirnoff statistic (D) are shown to denote non-random distribution of features with respect to G3/G4 score.

Figure 2S

A: Heatmap showing 4 consensus NMF metagenes calculated for n=331 MB and grouped by subgroup. MB_{Grp3}/MB_{Grp4} individuals are ordered by G3/G4 score. Column annotation shows subgroup as determined by

RNA-seq (Expression Subgroup) as determined by methylation (Methylation Subgroup), methylation MB_{Grp3}/MB_{Grp4} subtype (I-VIII) as per Sharma *et al* 2019 as defined using MNPv2 classifier (Grp3/4 Subtype). Presence of mutations are indicated to be present or not by dark grey shading. White indicates missing data. **B:** Rug plot showing distribution of mutations with respect to G3/G4 score. Summary counts are given according to the convenient divisions of HighG4, LowG4, G3.5, LowG3, HighG3 and reflected by the red line plots. Presence of a given feature is indicated by a bold tick mark the color of which indicates methylation MB_{Grp3}/MB_{Grp4} subtype (I-VIII). P-values for a Kolmogorov-Smirnoff statistic (D) are shown to denote non-random distribution of features with respect to G3/G4 score.

Figure 3S

A: Heatmap showing top 50 genes most significantly differentially expressed (top) and top 50 CpGs differentially methylated between MB_{Grp3} and MB_{Grp4}. Samples are ordered according to G3/G4 score. Note the difference in gradation for the expression values as opposed to the more binary distribution of DNA methylation beta-values. **B:** Heatmap showing DNA methylation values of the top 40 most discriminatory CpGs distinguishing HighG4 (dark green), LowG4 (light green), Low G3 (yellow) and High G3 (orange). G4 hypermethylated CpGs are shown on the left and hypomethylated CpGs on the right. Samples are ordered according to G3/G4 score and G3/G4 categories (HighG4, LowG4, G3.5, LowG3, HighG3) are annotated. **C:** Scatterplot showing beta-values for CpG "cg19784198" colored by G3/G4 categories (HighG4, LowG4, G3.5, LowG3, HighG3) an example of a CpG differentially expressed between MB_{Grp3} and MB_{Grp4} showing a bimodal methylation distribution. The relationship with G3/G4 score can effectively be modelled by a sigmoid/logistic function.

Figure 3S

A: Rug plot showing distribution of clinicopath features (top) mutations (middle) and copy number (bottom) with respect to G3/G4 score derived from DNA methylation data. Summary counts are given according to the convenient divisions of HighG4, LowG4, G3.5, LowG3, HighG3 and reflected by the red line plots. Presence of a given feature is indicated by a bold tick mark the color of which indicates methylation MB_{Grp3}/MB_{Grp4} subtype (I-VIII). P-values for a Kolmogorov-Smirnoff statistic (D) are shown to denote non-random distribution of features with respect to G3/G4 score. Infant=age at diagnosis < 3 years, Metastases = M+, DOD=Dead of

Disease, LCA = Large Cell Anaplasia, PRDM6 = *PRDM6* rearrangement. **B:** Empirical density and rug plots showing the distribution of M+ in MB_{Grp3}/MB_{Grp4} subtype III, LCA in MB_{Grp3}/MB_{Grp4} subtype II and *MYC* amplification in MB_{Grp3}/MB_{Grp4} subtype III with respect to G3/G4 score. The given clinico-pathological features are significantly non randomly distributed with respect to G3/G4 score even within specific MB_{Grp3}/MB_{Grp4} subtypes as shown by Kolmogorov-Smirnoff test (D). **C:** Empirical density and rug plots showing the distribution of copy number changes i17q in MB_{Grp3}/MB_{Grp4} subtype VIII, Gain of chromosome 5 in MB_{Grp3}/MB_{Grp4} subtype II and loss of chromosome 8 in MB_{Grp3}/MB_{Grp4} subtype VI with respect to G3/G4 score. The given copy number features are significantly non randomly distributed with respect to G3/G4 score even within specific MB_{Grp3}/MB_{Grp4} subtypes as shown by Kolmogorov-Smirnoff test (D). **D:** Heatmap showing (z-scores) of either methylation MB_{Grp3}/MB_{Grp4} subtypes (I-VIII) or G3/G4 score as an explanatory variable in a logistic regression determining presence or absence of the indicated clinico-path, copy number changes or mutational characteristics.

Figure 4S

A: Kaplan-Meier plot showing significant differences in MB_{Grp3}/MB_{Grp4} overall survival (patients of all ages) by G3/G4 continuum position. **B:** Forest plot showing univariate Cox models (patients > 3 years) of overall survival containing the variables G3/G4 score (as predicted by DNA methylation) treated as a categorical variable and **C:** MB_{Grp3}/MB_{Grp4} methylation subtype.

Figure 5S

A: Boxplot showing the distribution by MB subgroup of moderately expressed genes, isoforms, CDS or TSS as defined by a TPM>10. **B:** Boxplot showing significant differences in OEI (Overall Editing Index), i.e. level of RNA-editing by MB subgroup.

Figure 6S

Plots showing the per-cell expression of genes whose expression varies according to pseudotime on the **A:** RL to eCN/UBC branch (MB_{Grp3} specific genes are shown on the left and MB_{Grp4} specific genes shown on the right) and the **B:** GCP to GN branch ($MB_{SHH\text{-}Infant}$ specific genes are show on the left and $MB_{SHH\text{-}Child}$ specific genes

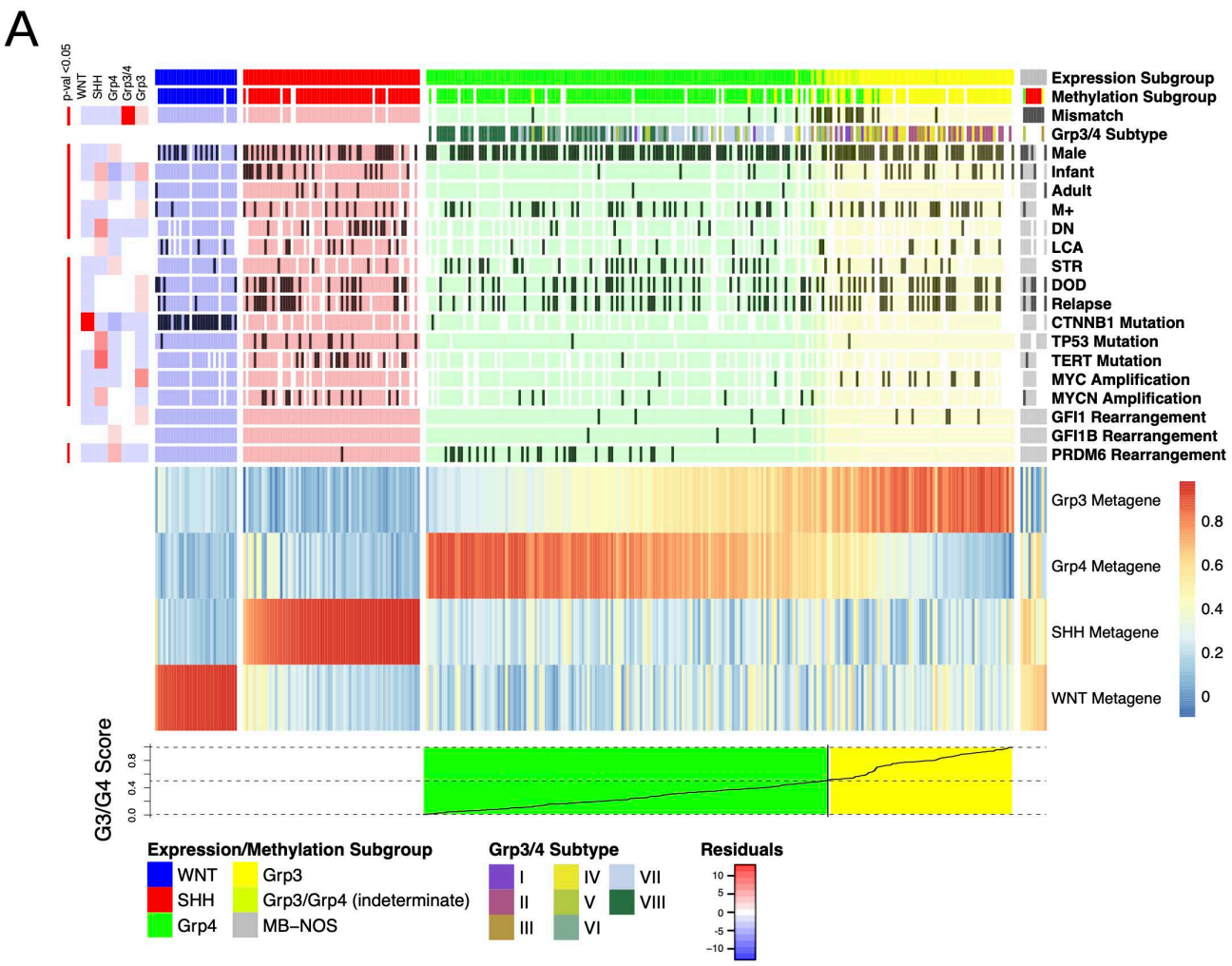
821 are shown on the right). Cell type is denoted by color. Black line represents a loess curve. Expression is
822 represented as normalized count data.

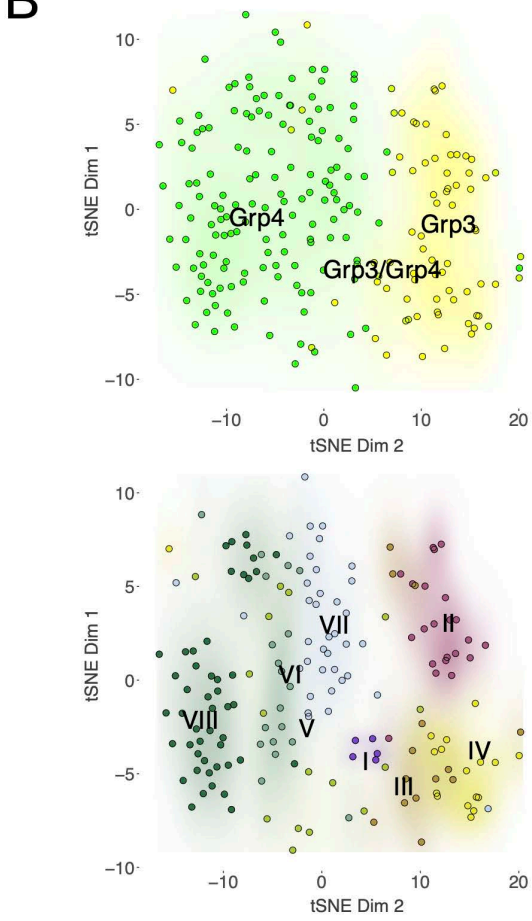
823 **Figure 8S**

824 **A:** Ridgeplots showing (left) distribution of G3/G4 score MB_{Grp3}/MB_{Grp4} patients by methylation subtype (I-
825 VIII) (right) and distribution of age at diagnosis by DNA methylation subtype (I-VIII). **B:** Scatterplot showing
826 age at diagnosis by G3/G4 score (as determined by DNA methylation), 2d empirical density is shown as red
827 shading and a loess curve with 95% CI is shown as blue line with grey shading. **C:** Plot showing 3d weighted
828 empirical density of cells on the RL to eCN/UBC branch in fetal cerebellar development from 9-21 PCW and
829 by pseudotime. Cells were weighted according to the proportion of cerebellar cells sampled at that time point.
830 **D:** Empirical density plots showing (left) distribution of MB_{Grp3}/MB_{Grp4} patients by G3/G4 score and (right) the
831 empirical density cumulatively across 9-21 PCW (essentially the 3d volume under the plane shown in panel C)
832 by pseudotime. Note the similarities in distribution. The dotted line shows the divide between Grp3 and Grp4
833 space on both G3/G4 and pseudotime scales and the percentages relate to the relative amounts of density
834 represented on either side of the divide.

835

Figure 1





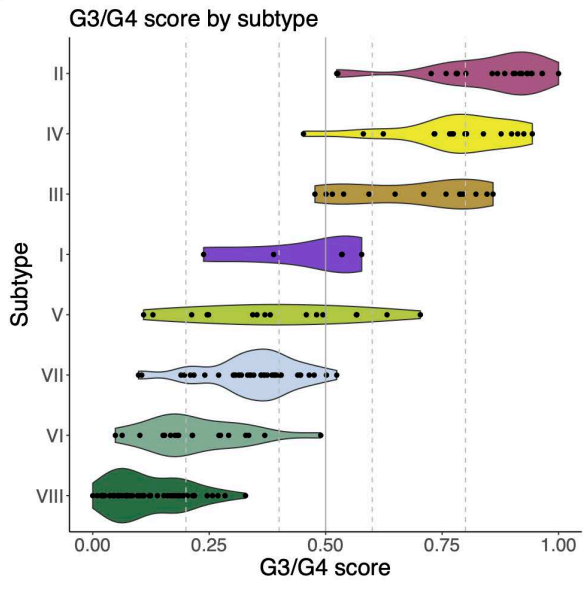


Figure 2

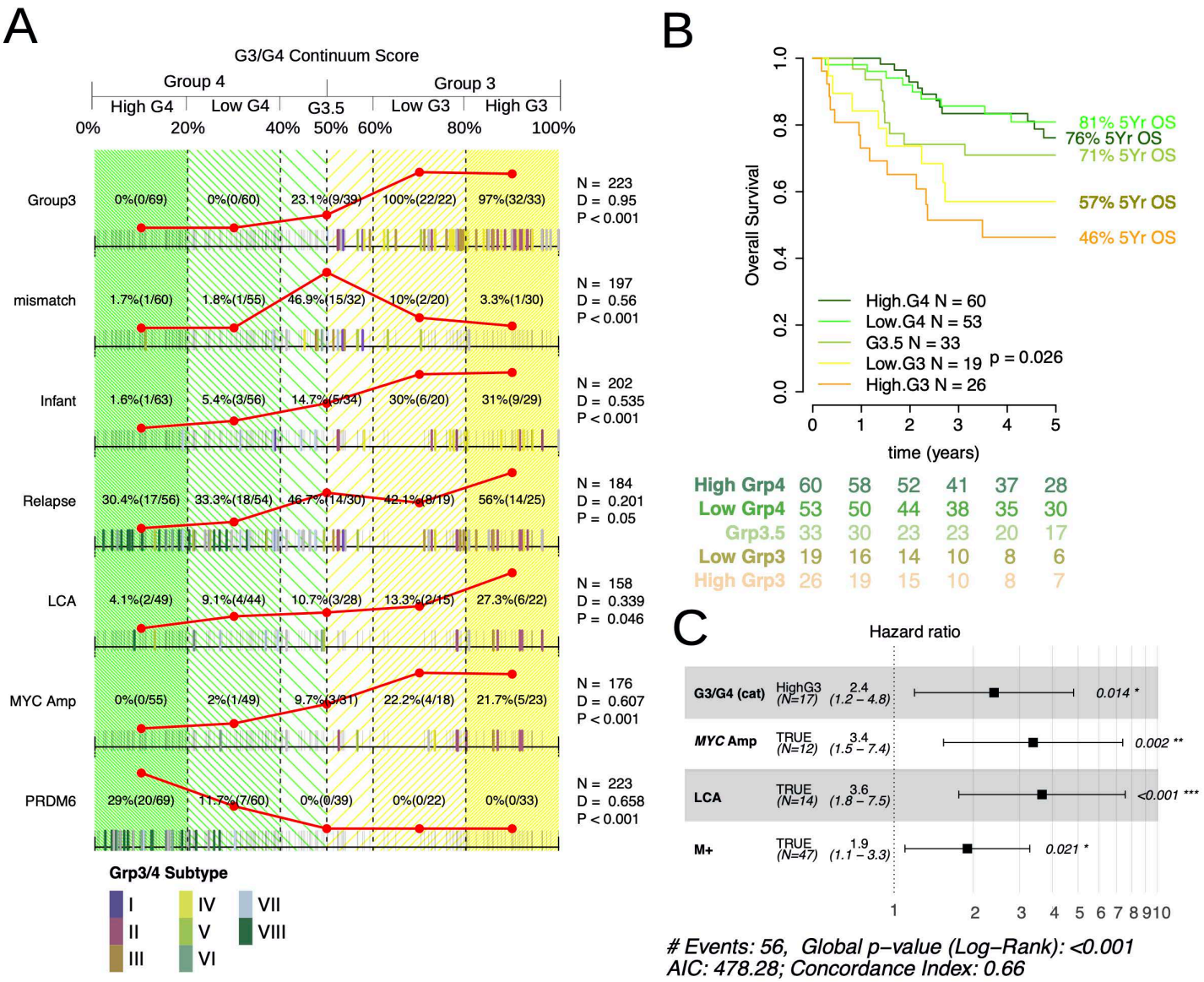
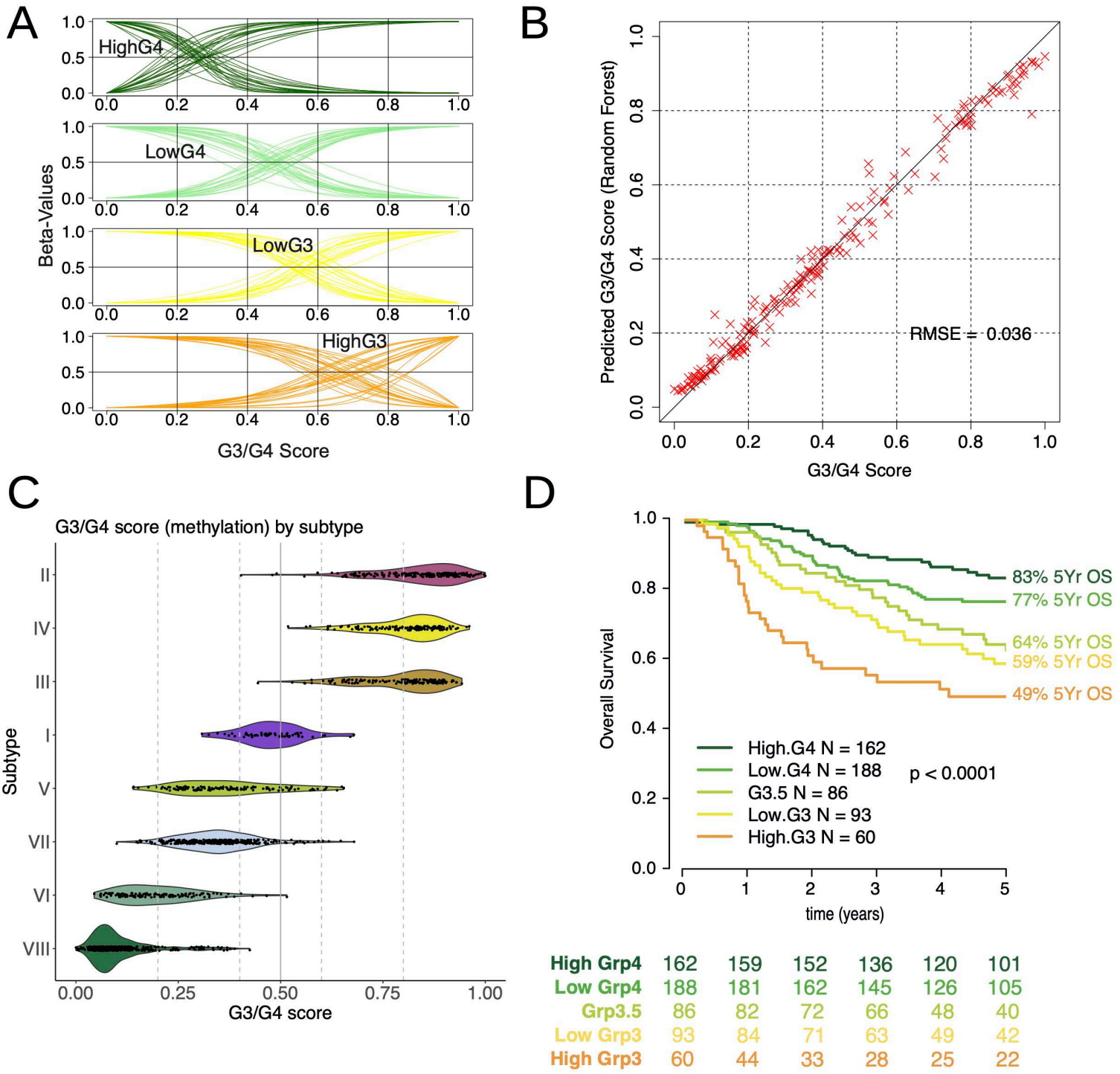


Figure 3



High Grp4	162	159	152	136	120	101
Low Grp4	188	181	162	145	126	105
Grp3.5	86	82	72	66	48	40
Low Grp3	93	84	71	63	49	42
High Grp3	60	44	33	28	25	22

Figure 4

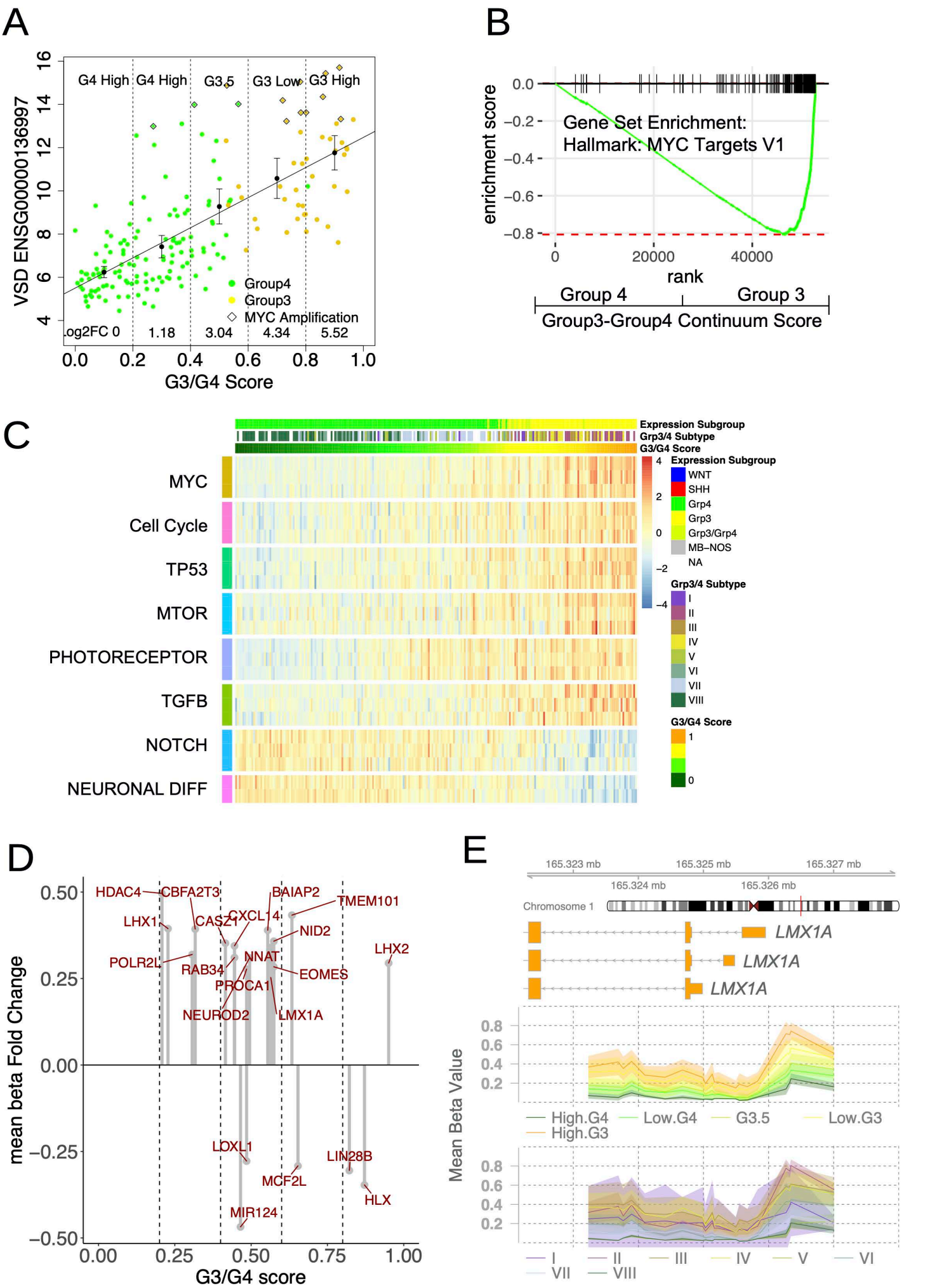


Figure 5

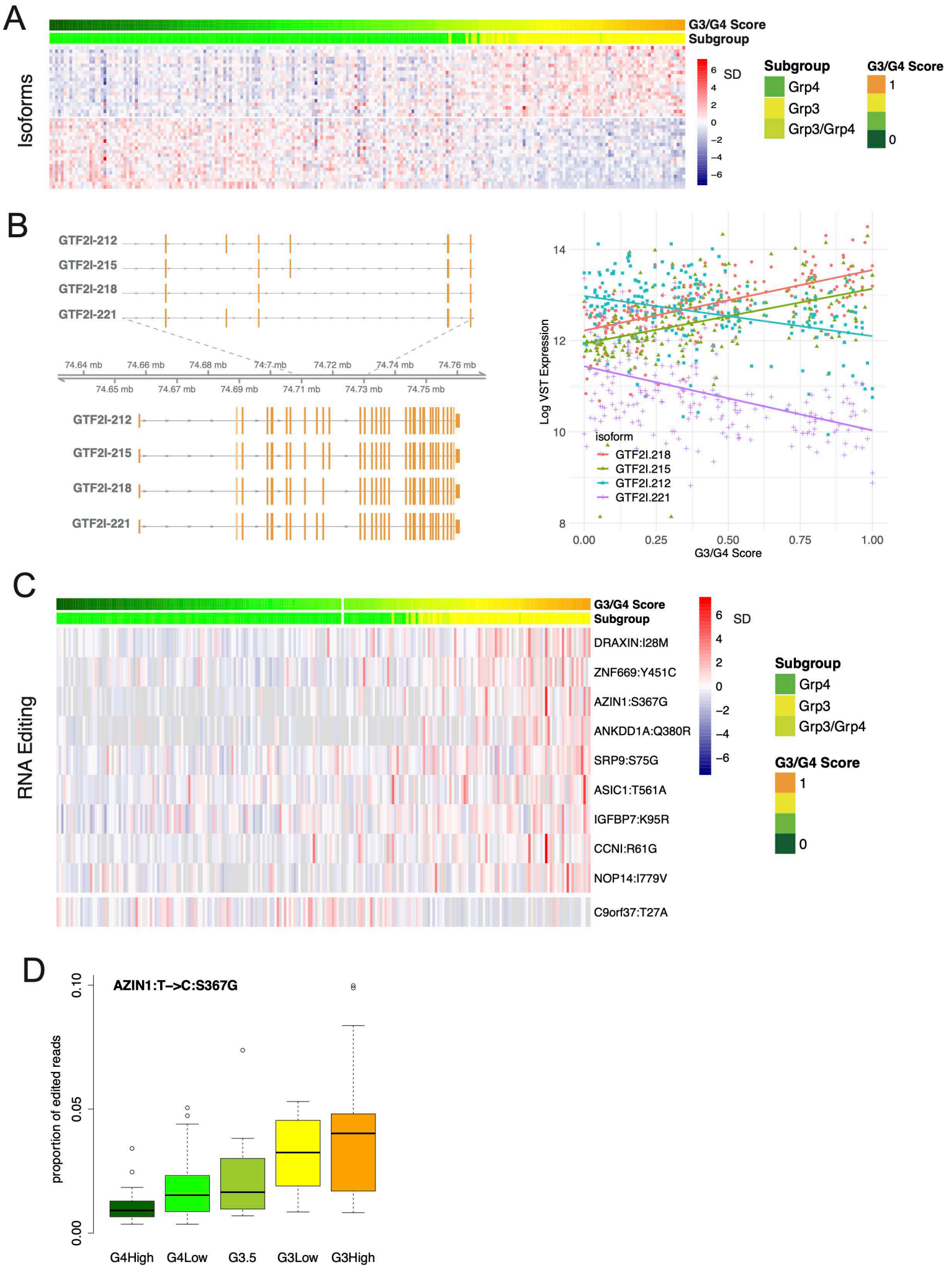


Figure 6

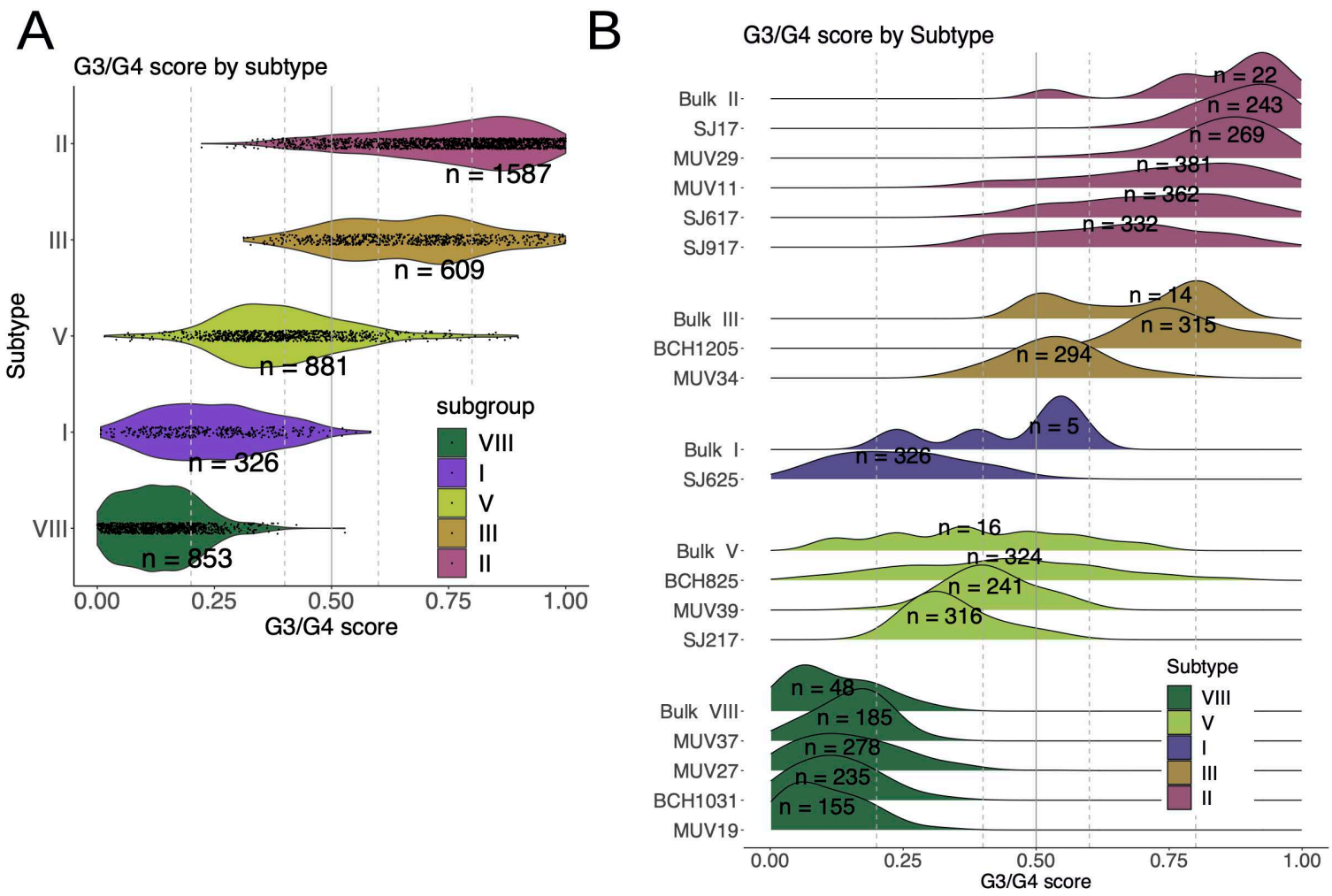


Figure 7

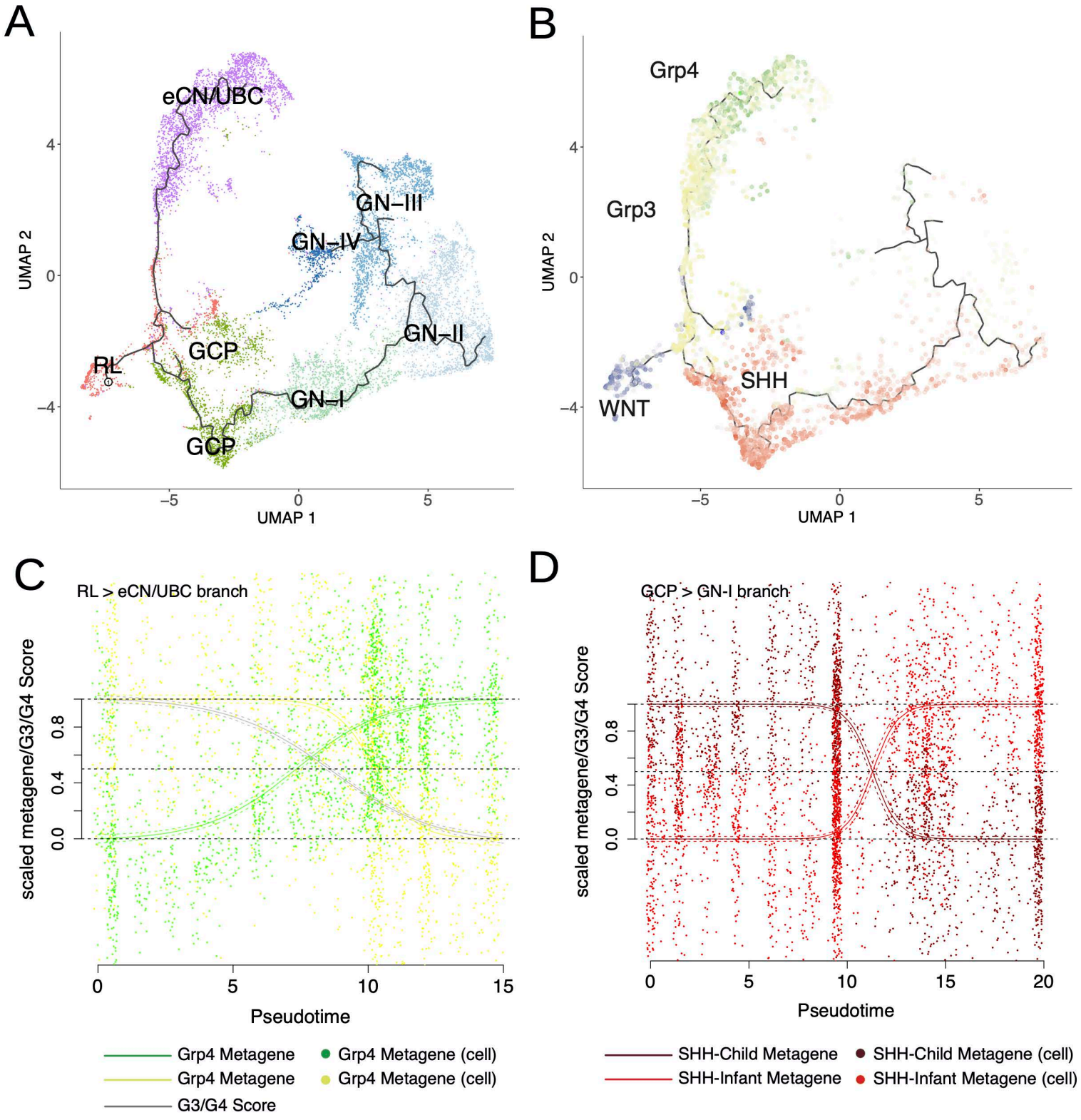


Figure 8

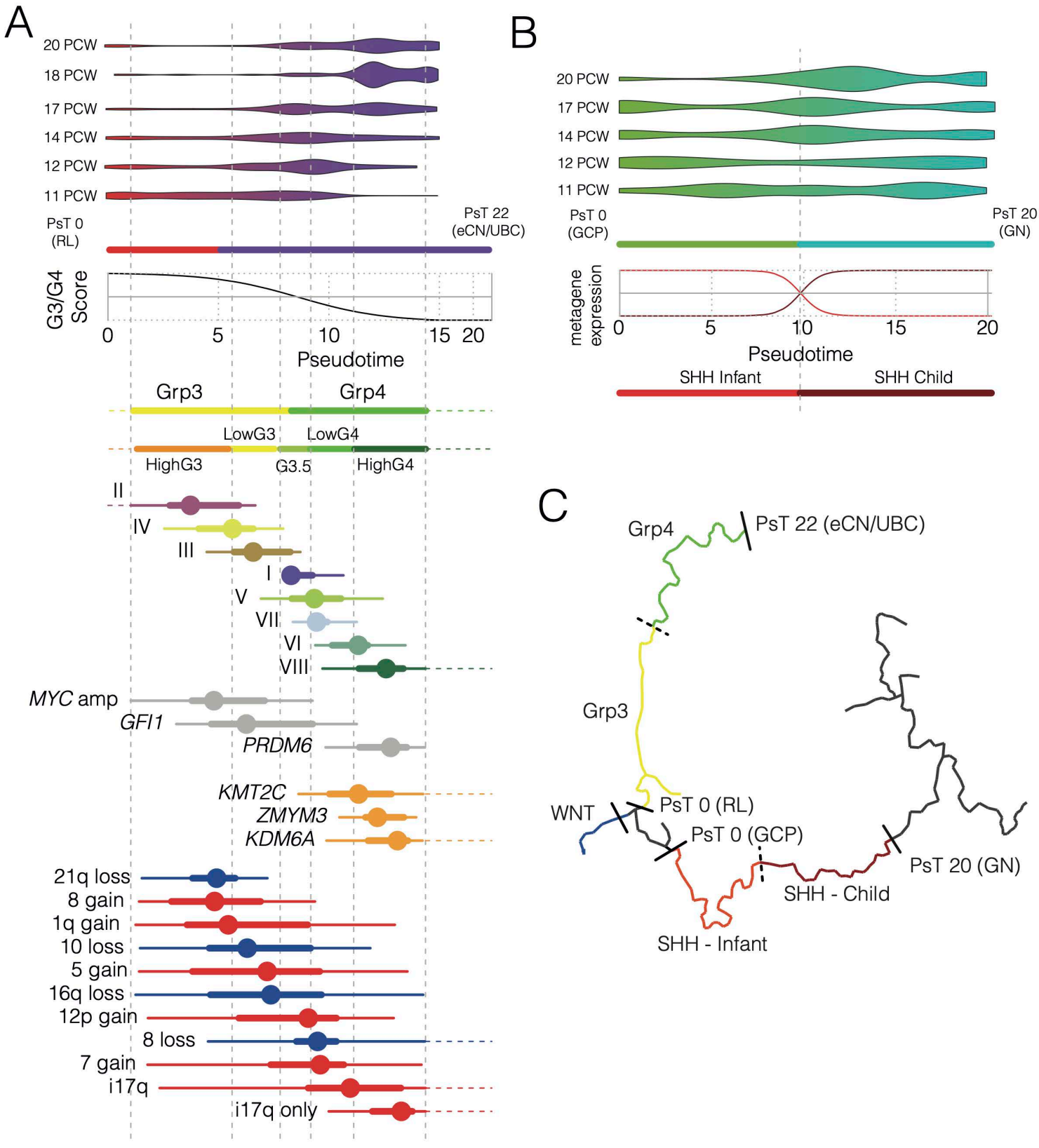
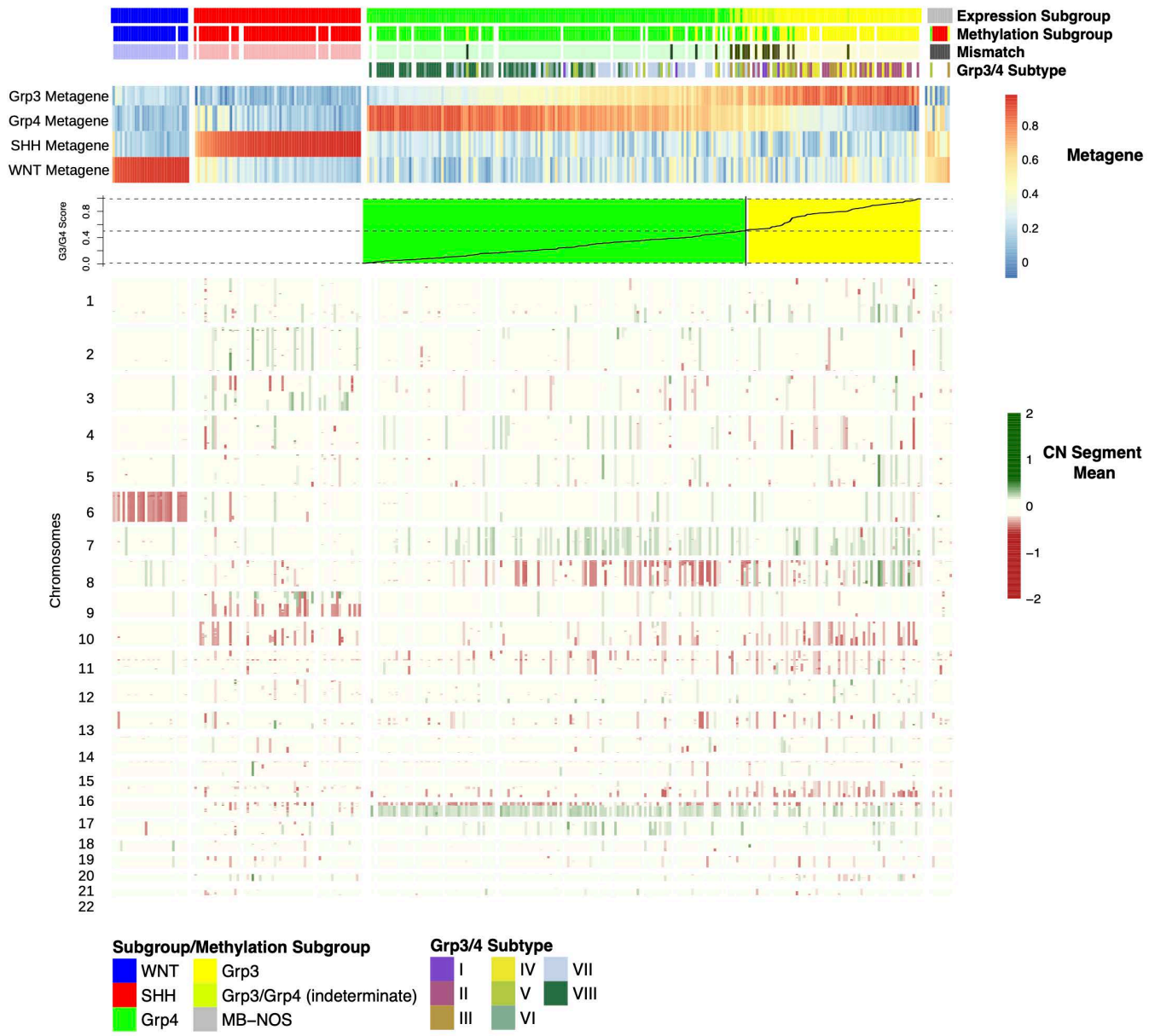


Figure 1S

A



B

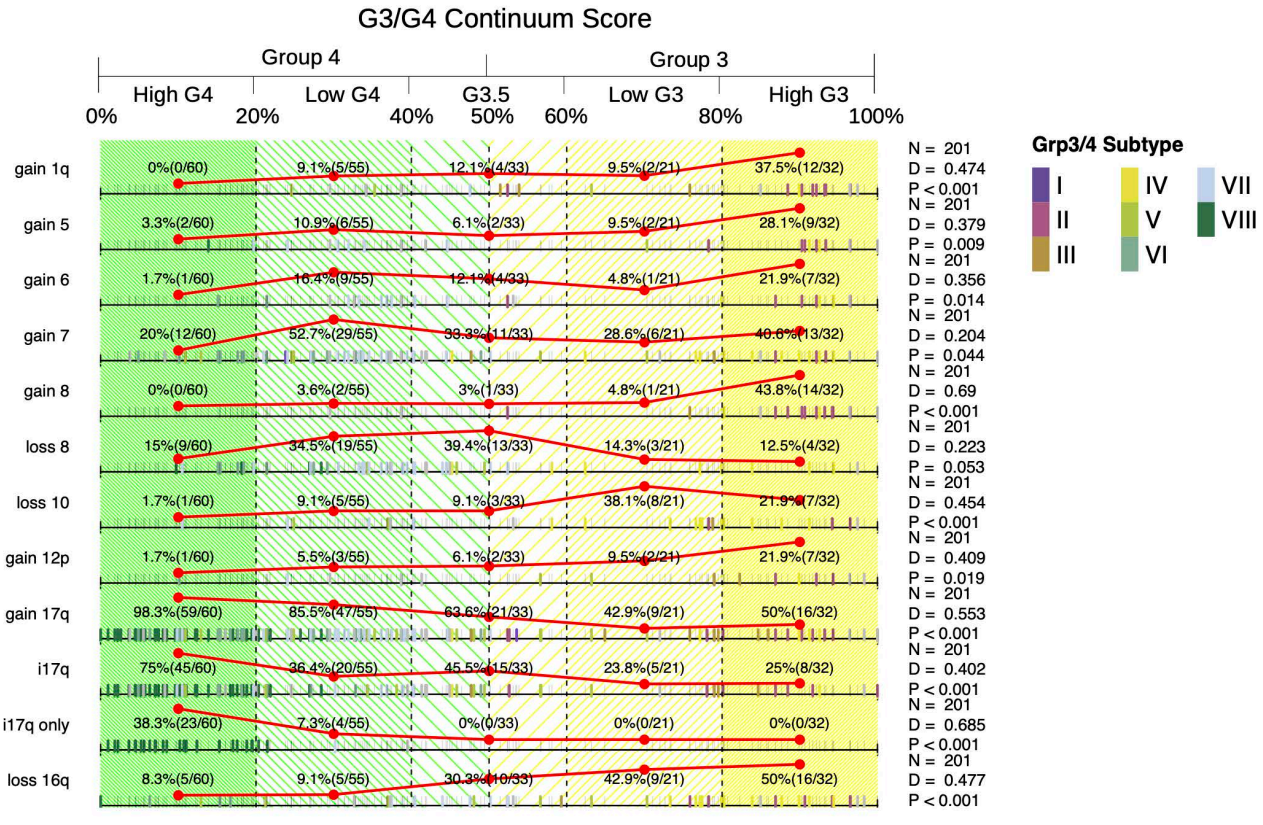
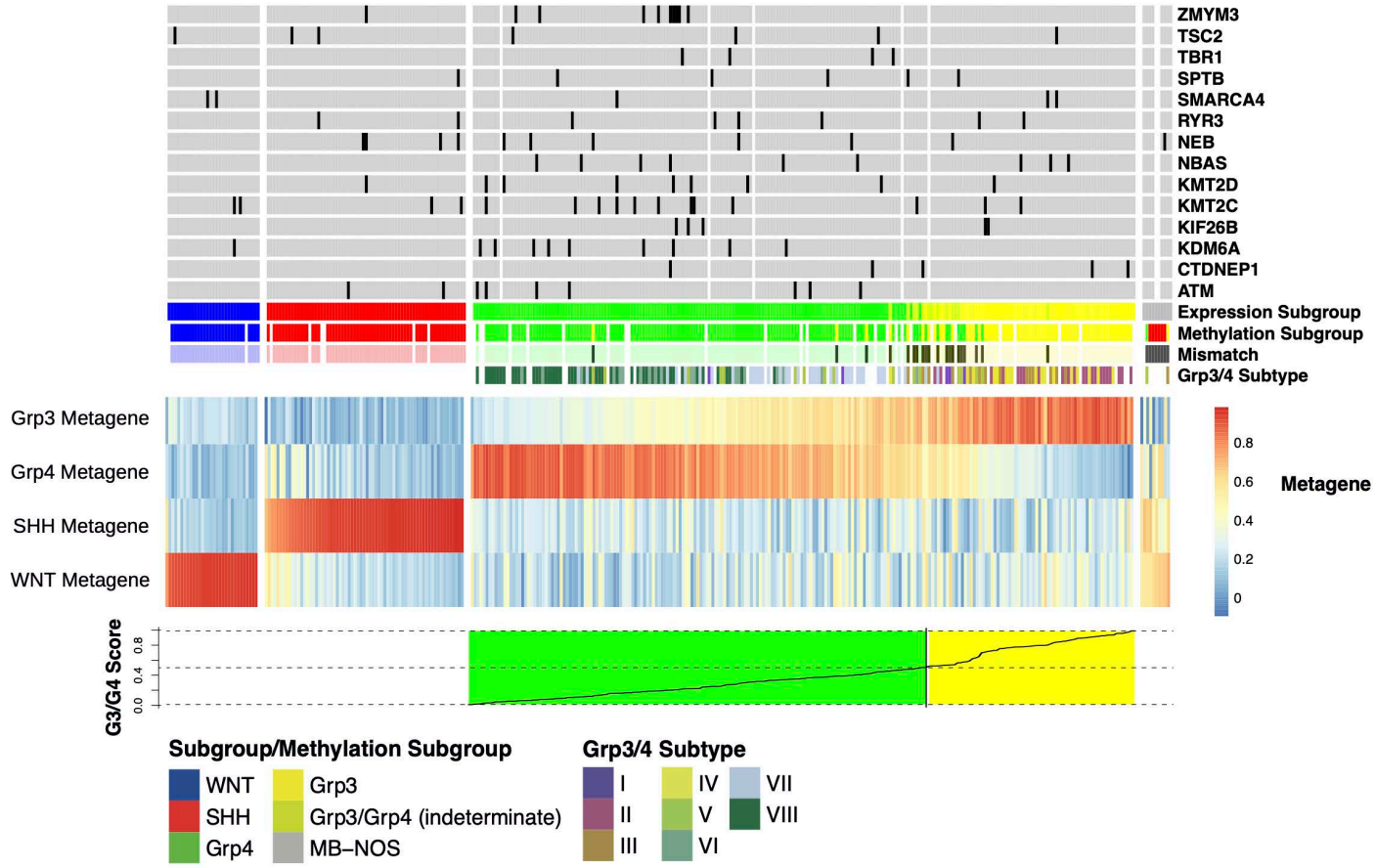


Figure 2S

A



B

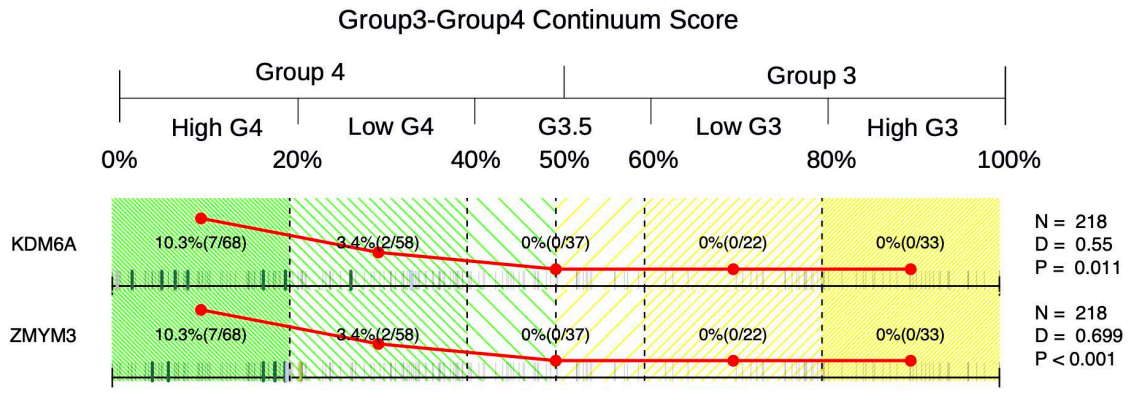


Figure 3S

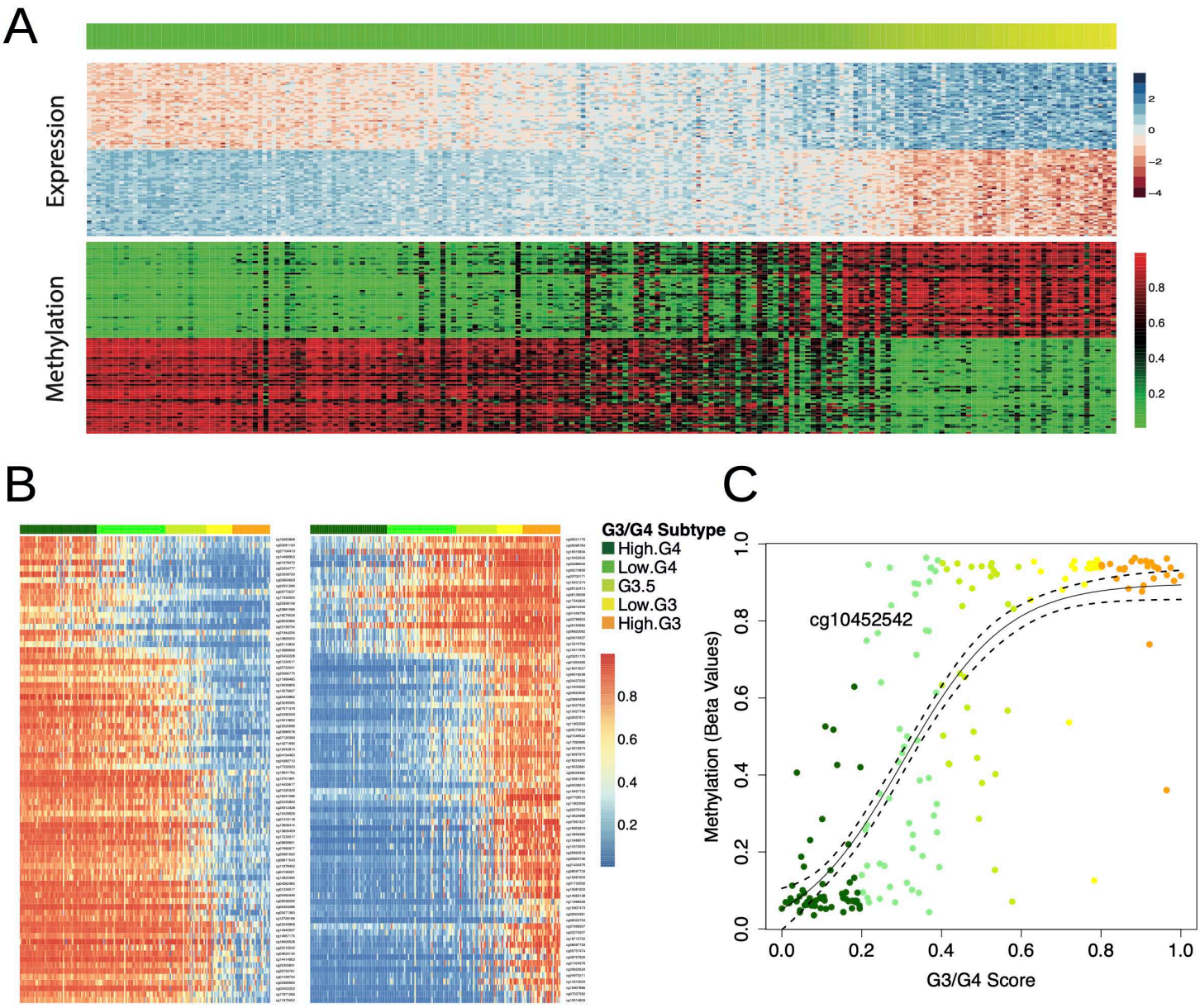


Figure 4S

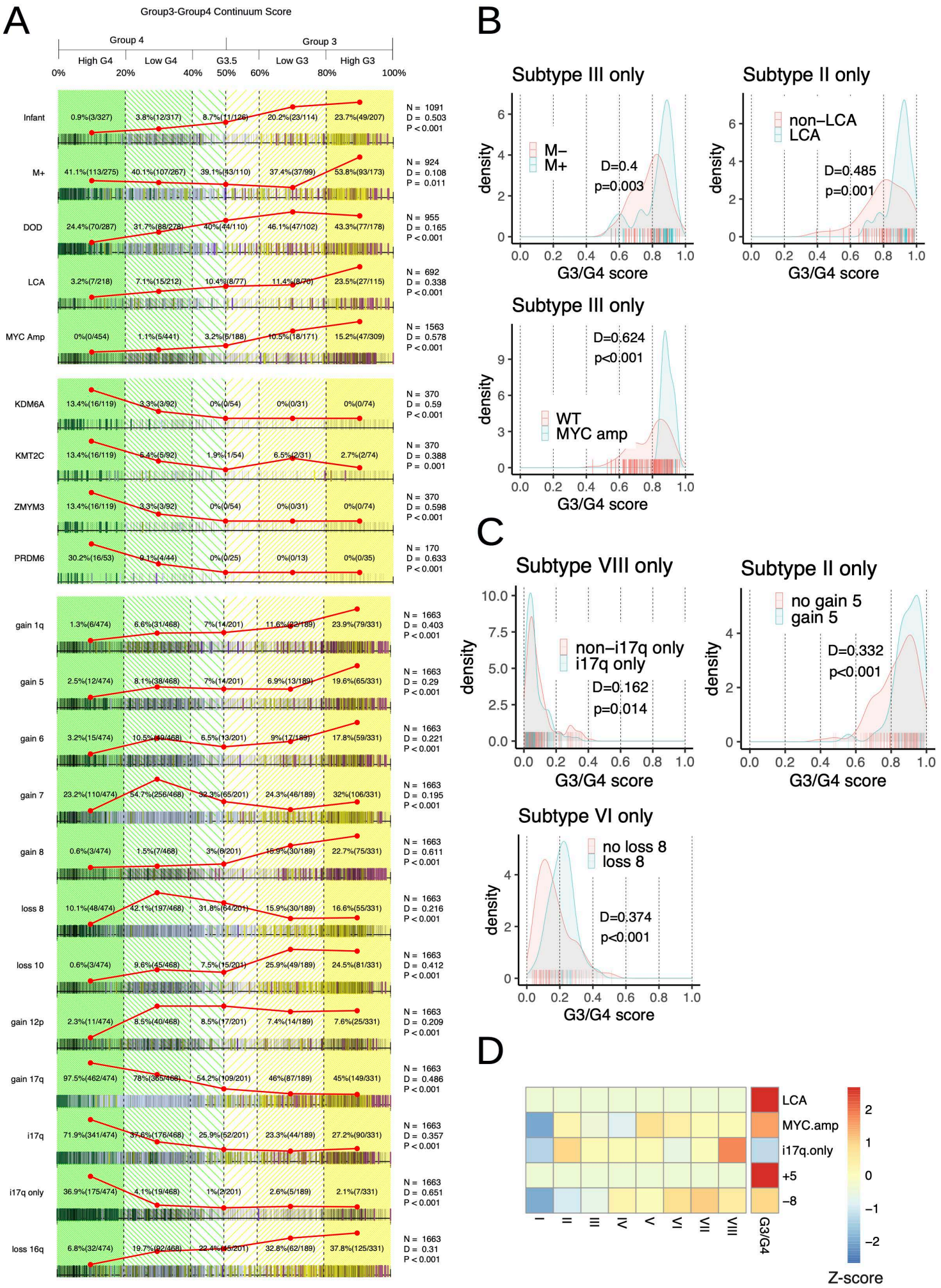
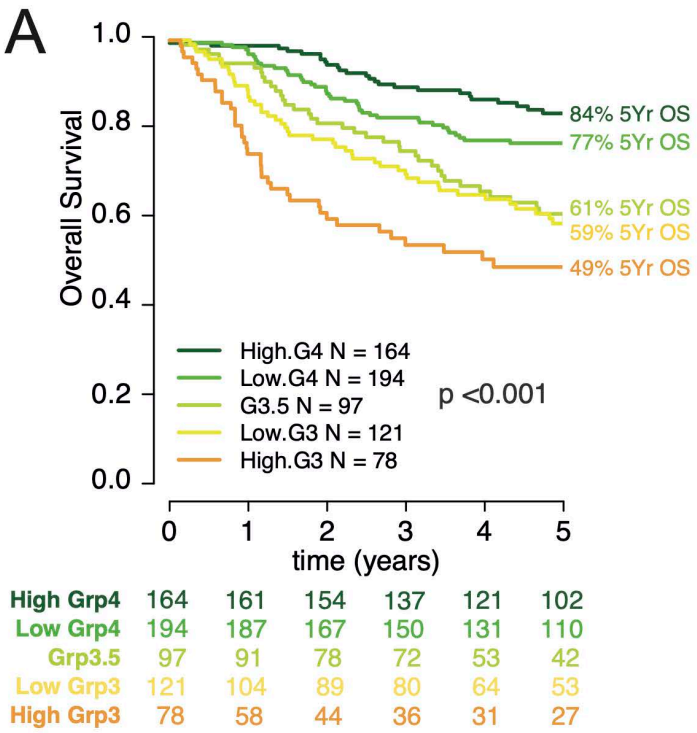
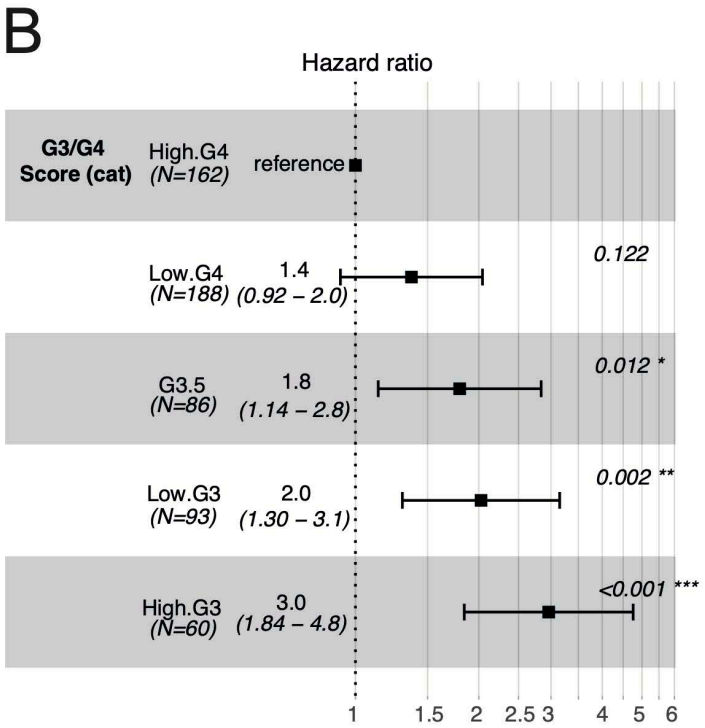


Figure 5S





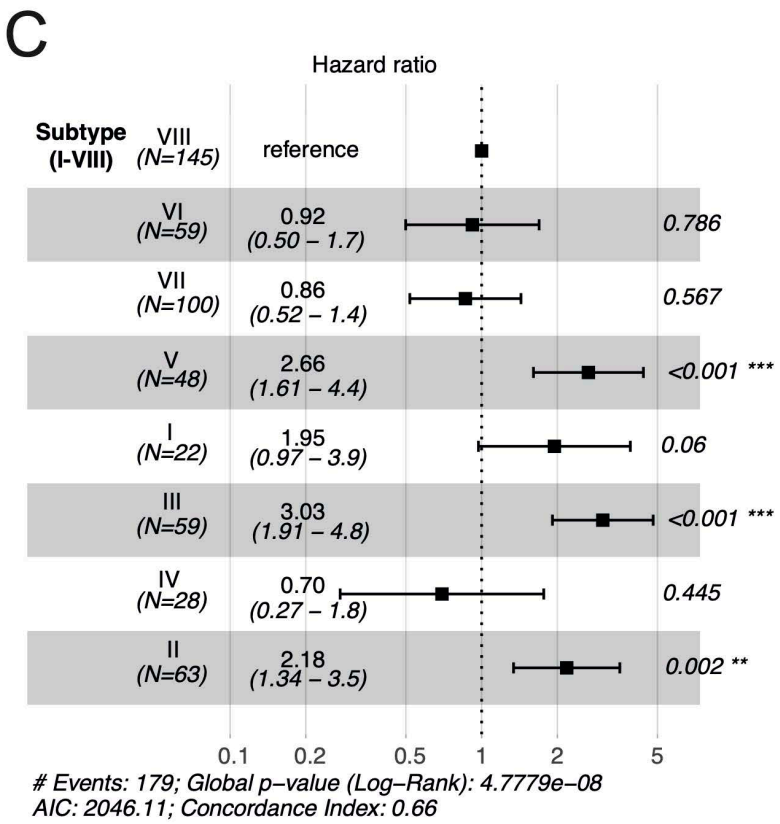


Figure 6S

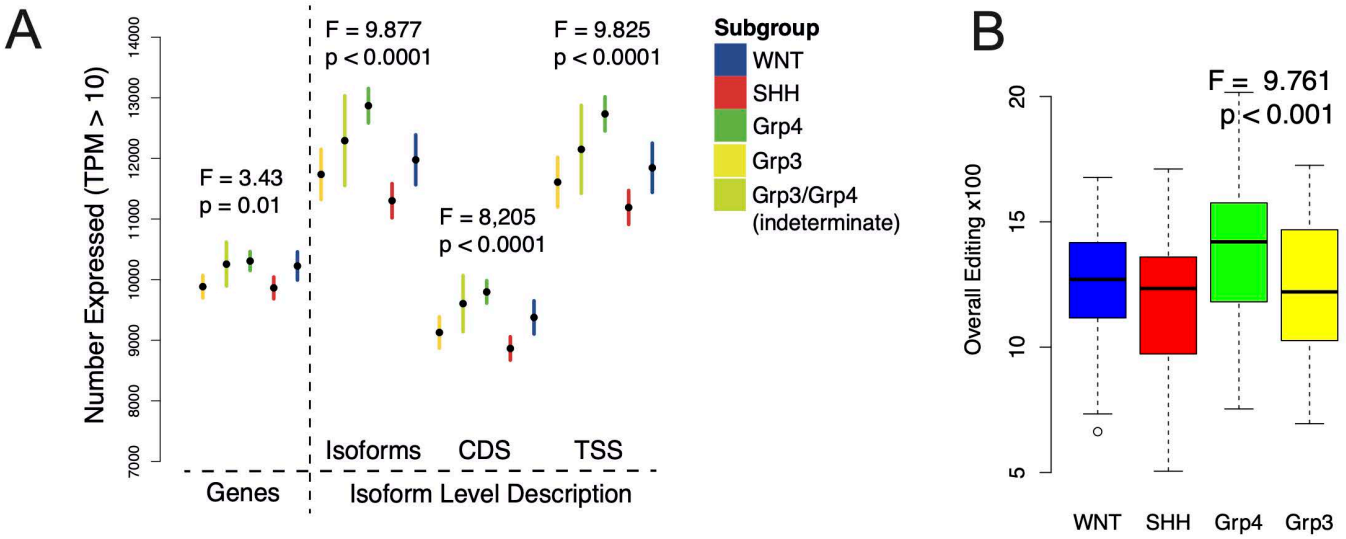
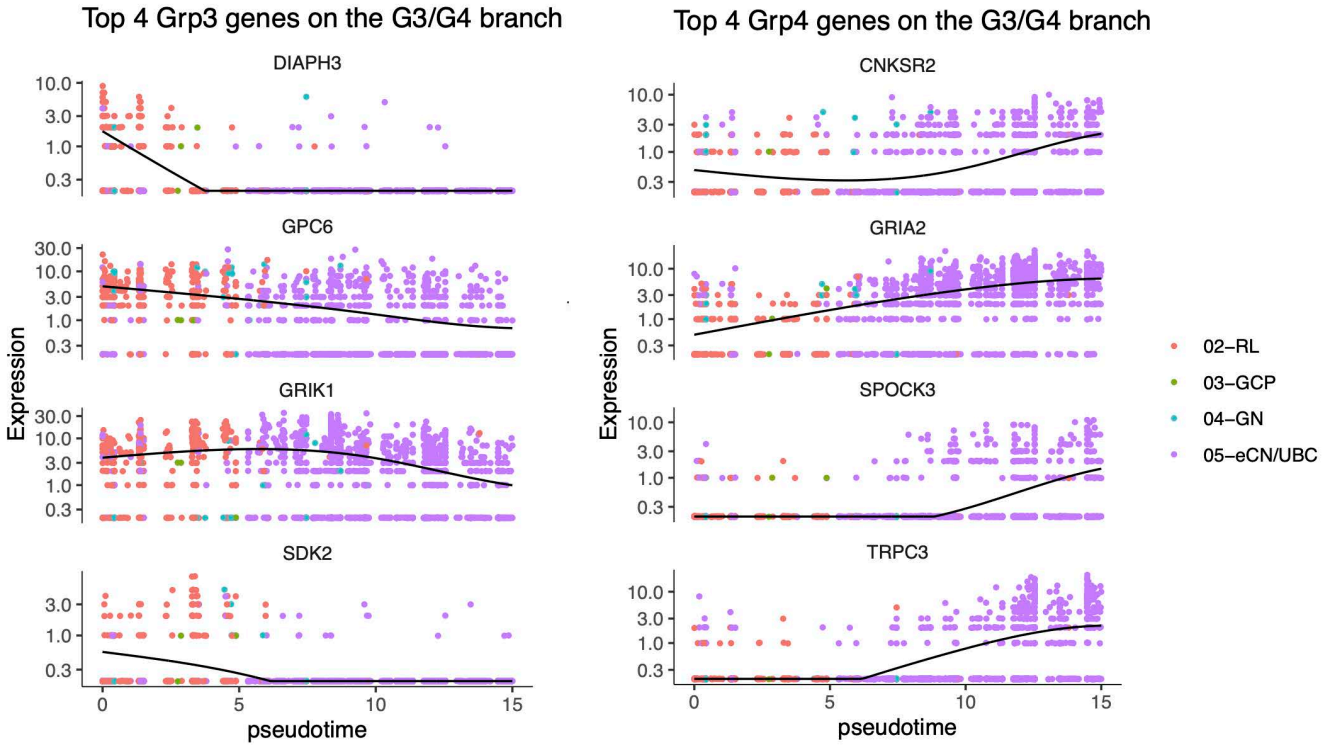


Figure 7S

A



B

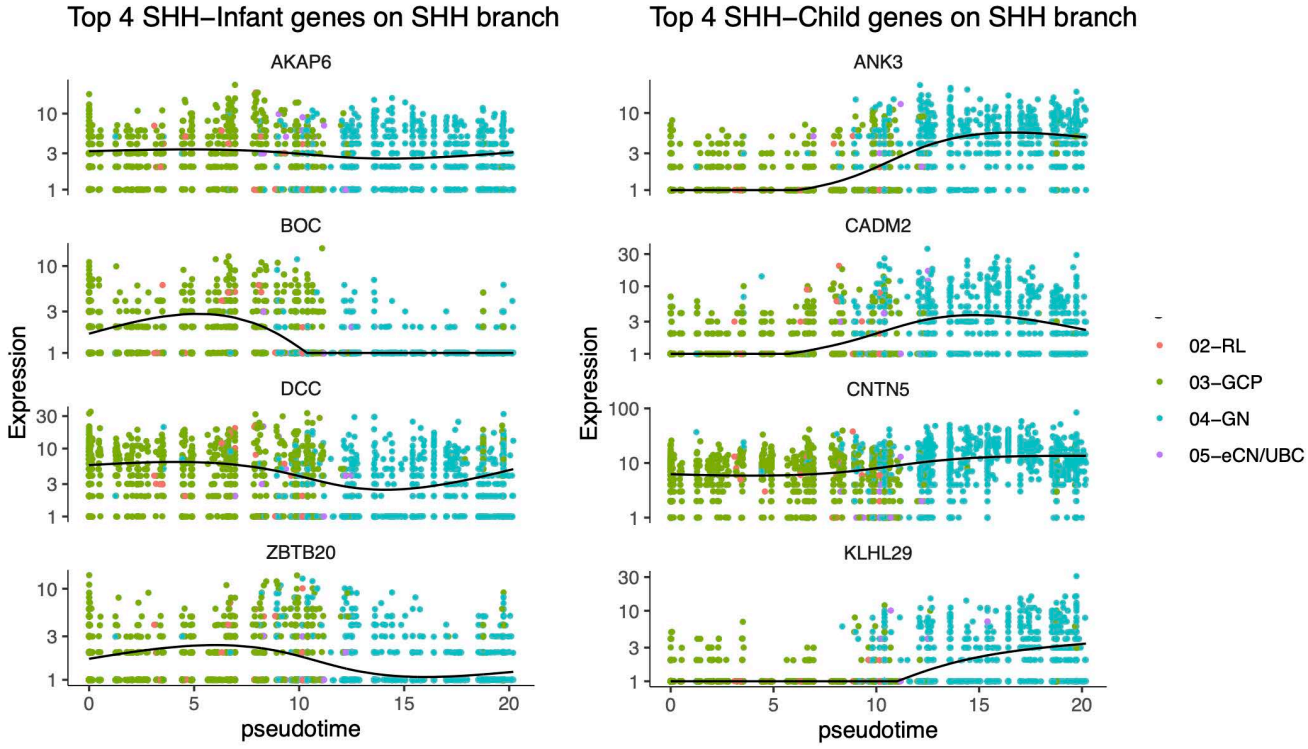
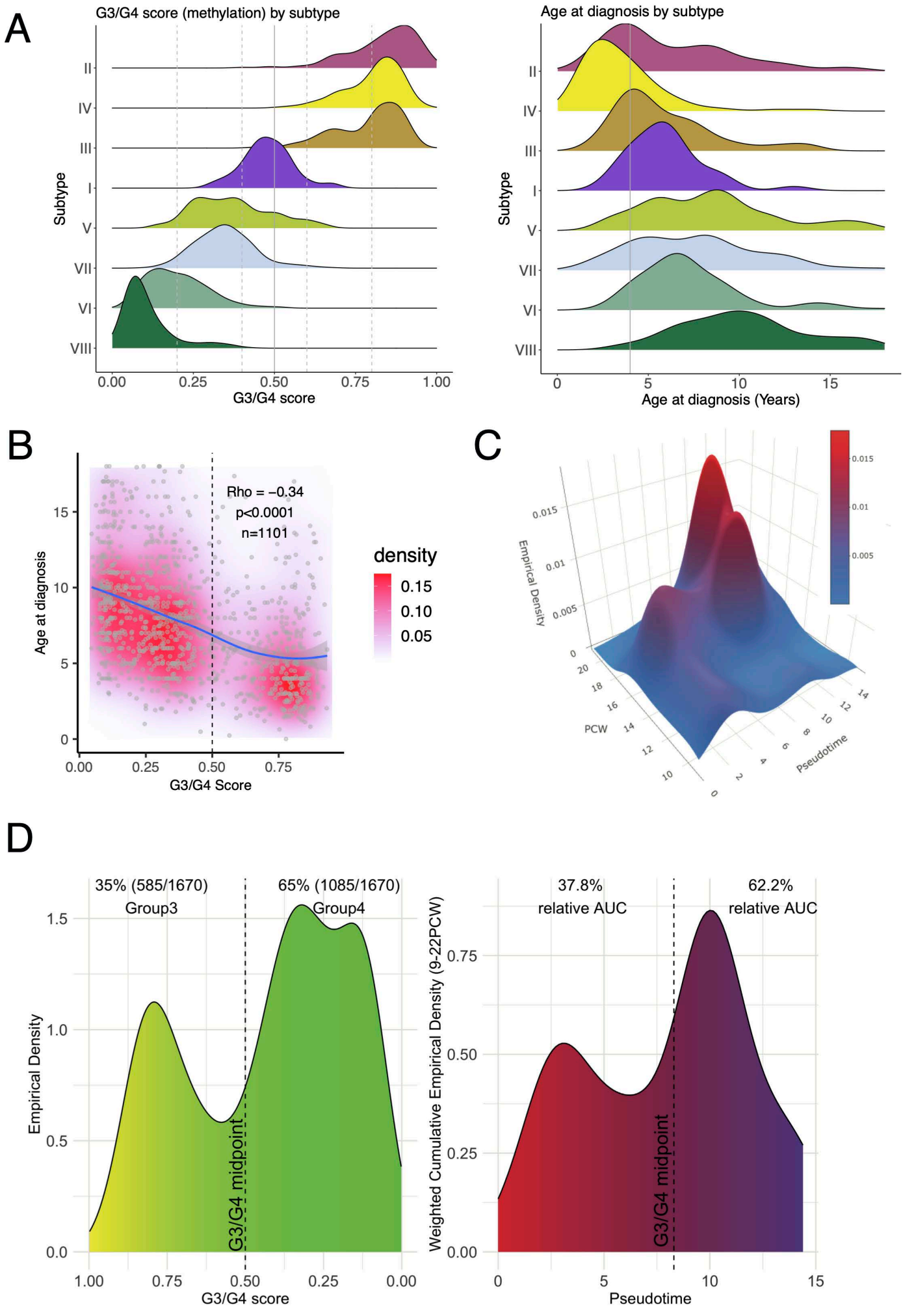


Figure 8S



STAR Methods

Resource Availability

Lead contact

Further information and request for resources and reagents should be directed to and will be fulfilled by the lead contact, Daniel Williamson (daniel.williamson@ncl.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data arising from this study has been deposited in E-MTAB-10754 and E-MTAB-10767 (ArrayExpress) and are publicly available as of the date of publication. Additionally, this study makes use of previously deposited data sets GSE130051, GSE93646, GSE119926 (GEO/NCBI). For scRNA-seq fetal cerebellar data, processed data are available through the Human Cell Atlas (https://www.covid19cellatlas.org/aldinger20) and the UCSC Cell Browser (https://cbl-dev.cells.ucsc.edu). Sequence data is available in the Database of Genotypes and Phenotypes, under accession number phs001908.v2.p1 (dbGAP/NCBI). Details are listed in the key resources table and method details section. No custom code was used in this study. Open-source algorithms were used as detailed in the method details section. Details on how these algorithms were used are available from the corresponding authors upon request.

Experimental Models

Human tissue samples

Snap frozen tumor samples from individuals with confirmed medulloblastoma diagnosis were used for RNA-seq analysis. These were provided as part of UK CCLG-approved biological study BS-2007–04 and/or with approval from Newcastle North Tyneside Research Ethics Committee (study reference 07/Q0905/71); informed, written consent was obtained from parents of all patients younger than 16 years.

Method Details

Patient samples and study cohort

331 tumor samples from individuals with confirmed medulloblastoma diagnosis were used for the RNA-seq analysis. Histopathological variants were defined according to the WHO 2016 guidelines (Louis et al., 2016). Metastatic status (M+) was defined as M>1 as per Chang's criteria (Chang et al., 1969). *MYC* and *MYCN* amplification status was assessed by fluorescence in situ hybridization and/or copy-number estimates from methylation array and *TP53*, *CTNNB1*, and *TERT* mutation status by Sanger sequencing. DNA was extracted using Qiagen DNeasy blood and tissue kit. Other mutations were assessed using next-generation sequencing. Whole-exome and targeted gene panel sequencing was performed using the Agilent SureSelect target enrichment platform and Illumina paired-end sequencing according to manufacturer's instructions. NGS datasets were analyzed for coding/exonic region variants using Genome Analysis Toolkit (GATK) version 3.7, according to Broad Institute's best practices (Burrows wheeler alignment, Haplotype Caller, Variant Quality Score Recalibration for exomes and Hard-filtering for panel) (Van der Auwera et al., 2013) and annotated using Ensembl Variant Effect Predictor (McLaren et al., 2016). Variants were predicted pathogenic if their consequence included coding or splice donor/acceptor mutations, max allele frequency was <0.01 in each of the large sequencing studies (ExAC, GnomAD/exomes, 1000 Genomes, ALFA) and predicted to be deleterious by both CAROL and FATHHM prediction tools (Lopes et al., 2012; Shihab et al., 2013). Variants called by targeted panel sequencing were called at a mean read depth of 278 (Standard Error of Mean =11). Exome studies were performed at mean depth of 40x. Pathogenic variants required a variant allele frequency ≥10%, a minimum read depth ≥10 and a minimum 2 variant forward reads and 2 variant reverse reads. Variants were further curated for obvious artefacts by visual inspection in Integrative Genomics Viewer (IGV) (J. T. Robinson et al., 2011). Chromosome-arm level copy-number estimates where derived from DNA methylation array data using conumee (R/Bioconductor). A larger previously published MB_{Grp3}/MB_{Grp4} cohort (Sharma et al., 2019) (Schwalbe et al., 2017) (GSE130051 & GSE93646) to which 166 novel profiles were added (E-MTAB-10754) (n=1670, exact samples used are detailed in Supplementary Table 2) was used for methylation-only analysis.

884 **RNA-seq analysis**

885 Total RNA was extracted from snap frozen tissue samples using trizol extraction followed by Qiagen RNeasy
886 Cleanup Kit and then subjected to transcriptome sequencing using Illumina TruSeq RNA Library Prep and
887 HiSeq 2500 platform achieving a ~90M paired end reads per sample. Following QC checks (fastqc/bamqc)
888 samples were aligned to genome hg19 using *RNA-star* (Dobin et al., 2013) in two-pass alignment mode and per
889 gene read counts generated using *ht-seq count* (Anders et al., 2015) and Gencode v25. Where isoform abundance
890 estimates were required these were generated using *kallisto* (Bray et al., 2016). For differential expression
891 analysis *DESeq2* (Love et al., 2014) (R/Bioconductor) was used for other analysis, clustering and visualization.
892 Read counts were first normalized and a variance stabilizing transform was first applied using the *vst* function
893 within *DESeq2* (R/Bioconductor). Additionally, a batch correction controlling for sequencing batch was applied
894 using the implementation of ComBat within the *sva* package (R/Bioconductor). Consensus NMF analysis was
895 performed as per the method described in Schwalbe *et al* (Schwalbe et al., 2017) and Sharma *et al* (Sharma et
896 al., 2019). Briefly, multi-run NMF is performed with n=250 iterations of 80% bootstrapping. Metagenes
897 calculated following each iteration are projected on to each removed sample and k-means clustering used to
898 predict the class of each removed sample based on the larger training set. A range of NMF metagene ranks (3-
899 10) and k-means clusters (3-10) are tested and cophenetic indices (a shorthand measure of the robustness of
900 sample clustering) used to evaluate the consistency of classification for each combination of metagenes.
901 Samples which were assigned to the same class with <90% consistency upon resampling were designated as
902 MB-NOS, except where they were alternately assigned as MB_{Grp3} or MB_{Grp4} with >90% consistency, in which
903 case they were classified as MB_{Grp3}/MB_{Grp4}.

904 Averaged and standardized metagene h-values from across the bootstraps were used as measures of metagene
905 expression. All NMF projections were performed using column-rank and post-projection normalization as per
906 the method described by Tamayo *et al* (Tamayo et al., 2007). t-SNE were used for visualization was performed
907 using the *Rtsne* package (R/CRAN).

908 G3/G4 score was calculated by applying a logistic transformation 1/(1+exp(-x)) to the MB_{Grp3} and MB_{Grp4}
909 metagenes (excluding two outliers). The G3/G4 score was calculated as the MB_{Grp3} proportion of the total
910 metagene scaled to between 0 and 1. For convenience of visualization, or where categorical comparison was

911 required, we referred to individuals >0 & ≤0.2 as “HighG4”, >0.2 & ≤0.4 as “LowG4”, >0.4 & ≤0.6 as “G3.5”,
912 >0.6 & ≤0.8 as “LowG3” and >0.8 & ≤1 as “HighG3”.

913 RNA editing was estimated using the QEdit/Reditools pipeline as previously described
914 (https://github.com/BioinfoUNIBA/QEdit) (Giudice et al., 2020). Differential RNA-editing was calculated
915 using a p-adjusted (Benjamini-Hochberg) Mann-Whitney U-test for two group analysis and Anova with
916 TukeyHSD (post-hoc) for multi-group analysis. Where unknown from DNA analysis *GFI1/GFI1B*, *PRDM6*
917 rearrangements were each inferred from RNA-seq data as per the method used originally by Northcott *et al*
918 (Northcott et al., 2017; 2014).

919 GSEA was performed using MsigDb library version 7.1 and the implementation of the original algorithm within
920 the package *fgsea* (R/Bioconductor) and ssGSEA using the implementation within *GSVA* (R/Bioconductor)
921 (Hänzelmann et al., 2013). The following gene sets were selected as reflective of the pathway categories given
922 in Figure 4C. MYC = "HALLMARK_MYC_TARGETS_V2", "MYC_UP.V1_UP",
923 "DANG_MYC_TARGETS_UP". Cell Cycle = "FISCHER_G1_S_CELL_CYCLE",
924 "GO_POSITIVE_REGULATION_OF_CELL_CYCLE", "GO_SIG
925 NAL_TRANSDUCTION_INVOLVED_IN_CELL_CYCLE_CHECKPOINT", TP53 = "CEBALL
926 OS_TARGETS_OF_TP53_AND_MYC_UP", "REACTOME_TRANSCRIPTIONAL_REGULATION
927 _BY_TP53”, “REACTOME_TP53_REGULATES", MTOR = "HALLMARK_MTORC1_SIGNALING",
928 "MTOR_UP.V1_UP", "MTOR_UP.N4.V1_UP", PHOTORECEPTOR = "GO_EYE_PHOTOREC
929 EPTOR_CELL_DIFFERENTIATION", "GO_CAMERA_TYPE_EYE_PHOTORECEPTOR_CELL_
930 DIFFERENTIATION", "GO_EYE_PHOTORECEPTOR_CELL_DEVELOPMENT", TGFB1 = "KARL
931 SSON_TGFB1_TARGETS_UP", "JAZAG_TGFB1_SIGNALING_VIA_SMAD4_UP", "KARAKAS_
932 TGFB1_SIGNALING" NOTCH = "GO_POSITIVE_REGULATION_OF_NOTCH_SIGNALING
933 _PATHWAY", "REACTOME_ACTIVATED_NOTCH1_TRANSMITS_SIGNAL_TO_THE_NUCLEUS",
934 "NGUYEN_NOTCH1_TARGETS_UP", Neuronal Diff = "GO_CENTRAL_NERVOUS_SYSTEM
935 _NEURON_DIFFERENTIATION" "LE_NEURONAL_DIFFERENTIATION_UP".

In analyzing association with G3/G4 score, the loss or gain of each non-acrocentric chromosome arm was considered as were the more frequent MB_{Grp3}/MB_{Grp4} mutations in genes *ATM*, *CTDNEP1*, *KDM6A*, *KIF26B*, *KMT2C*, *KMT2D*, *NBAS*, *NEB*, *RYR3*, *SMARCA4*, *SPTB*, *TBR1*, *TSC2*, and *ZMYM3*.

DNA methylation analysis

Beta/M-values were derived from HumanMethylation450 BeadChip (450k) and Infinium HumanMethylationEPIC (850k) arrays using the ssNOOB method within the package *minfi* (Aryee et al., 2014) excluding known SNPs and cross-hybridizing probes. In order to construct a random forest classifier which predicted G3/G4 score from DNA methylation data, we performed feature selection of CpGs using 192 MB_{Grp3}/MB_{Grp4} samples with both RNA-seq (i.e. known G3/G4 score) and Methylation array. We constructed using *limma* (R/Bioconductor) a number of bootstrapped (80% with 100 iterations) significance tests testing differential methylation between each of the categories HighG4, LowG4, G3.5, LowG3 and HighG3. We measured average performance for a range of numbers of features (10-100) on removed samples using a tuned support vector machine, however performance plateaued after a certain number of features, so it was decided to select the top 80 most frequently selected CpGs for each comparison. Thus n=400 CpG features were used to train a random forest classifier which was then subject to recursive feature elimination using 50x cross-validation and implemented using the *rfe/rfeControl* function within the *caret* package (R/CRAN). Where sigmoid curves are shown, these were fitted using the *fitmod* function within the *DoseFinding* package (R/Bioconductor). For visualization these were scaled to a minimum 0 and maximum 1.

Methylation subtype calling (Sharma et al., 2019) was obtained using an extension of the Heidelberg brain tumor classifier available at [https://www.molecularneuropathology.org/mnp]. A methylation classifier prediction score of >0.8 was used to assign subtype. Samples were excluded if not confirmed as MB by MNP.

Significantly differentially methylated regions (DMRs) distinguishing G4High, G4Low, G3.5, G3Low and G3High were calculated using *dmrcate* (R/Bioconductor) using settings lambda=1000, C=2. Regions were considered when the total number of CpGs ≥ 5, the minimum FDR < 0.05 and the mean Beta fold change $>$ 0.25. These were further filtered to identify DMRs which overlapped with the MB_{Grp3}/MB_{Grp4} specific enhancer/superenhancer regions identified by Lin *et al* (Lin et al., 2016).

scRNA-seq analysis

Previously published medulloblastoma scRNA-seq dataset (Hovestadt et al., 2019) GSE119926 was used. However, we used only the MB_{Grp3}/MB_{Grp4} primary patient samples (excluding the patient-derived xenografts) (n=4256 cells, n=15 samples) and excluded patients SJ970 and SJ723 due to the relatively few available cells. The pre-publication Human fetal cerebellar single cell reference data set, consisting of 69,174 cells, classified into 21 cell types and derived from 15 donors between 9 and 21 PCW, details can be found within https://www.biorxiv.org/content/10.1101/2020.06.30.174391v1 (Aldinger *et al* in press Nature Neuroscience). For the purposes of metagene projection *Seurat* (R/Bioconductor) (Butler et al., 2018) was used to select the 5000 most variable features using the "vst" method for both data sets and the resulting normalized matrices subject to NMF projection of the bulk metagenes and calculation of the G3/G4 score as per the bulk analysis described above. In this way, a per-cell metagene score and G3/G4 score was calculated.

Developmental trajectory analysis was performed using *monocle v3* (Qiu et al., 2017) (R/Bioconductor) using 12,243 cells classified as RL, GCP, GN or eCN/UBC which we defined broadly as the rhombic lip lineage as per Aldinger *et al*. Monocle v3 functions used were preprocess_cds, align_cds, reduce_dimension, cluster_cells, learn_graph, order_cells and plot_cells to visualize by UMAP. The relevant branches for the MB_{Grp3}/MB_{Grp4} and MB_{SHH} were divided as indicated (Figure 8) and the relationship between pseudotime and G3/G4 score/metagene was defined using a loess curve function. This enabled developmental and oncogenic events to be mapped onto a common scale (Figure 8). Genes whose expression varied significantly according to pseudotime were detected using Moran's test statistic as implemented by *monocle v3*. For analysis of the differences between $MB_{SHH\text{-}Infant}$ and $MB_{SHH\text{-}Child}$, a further metagene calculated using NMF rank = 2 only on MB_{SHH} (67/331 samples) was additionally projected onto the single cells in the same manner as the other metagenes. For calculating empirical density, the *density* function was used (R/Bioconductor) except where weighted two-dimensional estimation was needed in which case the *kde2d.weighted* function from the package *ggtern* (R/Bioconductor) was used. Weights were calculated as the number of cells at a given sampling point (9-21PCW) as a proportion of the total number of cells sampled.

Quantification and Statistical Analysis

Data analysis and visualization was carried out in R 3.5.3 except for the analysis of fetal cerebellar scRNA-seq which was performed using R 4.0.2. CRAN and Bioconductor packages used are given in the key resources table. To test significant association with time to death/progression free survival a log-rank test or cox-regression was used. Association analysis of individual clinico-pathological mutational and copy number features with G3/G4 score was performed using Kolmogorov-Smirnoff test. Where gene expression/pathway associations with G3/G4 score are assessed, these are assessed using Pearson's correlation coefficient. The test statistics and significant P-values ($p<0.05$) are stated in text and figures and were adjusted for multiple hypothesis testing using Benjamini-Hochberg where appropriate throughout. Where values of n are given, these generally pertain to number of sample/individual patients except where otherwise indicated. Boxplots, where used, show dispersion as per standard i.e. (center line = median, box = interquartile range, whisker= range minus outliers).

Data were excluded where samples were clearly indicated to be duplicated across multiple related datasets. Additional exclusions were carried out for samples where methylation array detection P value did not reach significance threshold in at least 90% of the array. Methylation samples were excluded from the analysis if not confirmed as medulloblastoma by MNP2.0. In our analysis of the scRNA-seq dataset GSE119926 we excluded patients SJ970 and SJ723 due to the relatively few available cells.

1005 **Key Resources Table**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Snap frozen medulloblastoma	CCLG Biobank/ Biological study and collaborating centres	See Supplementary Table S1
Chemicals, peptides, and recombinant proteins		
Trizol	Thermo Fisher	15596026
1 X Low TE Buffer (10mM Tris-HCL, ph 7.5-8.0, 0.1 mM EDTA)	Thermo Fisher	120900915
100% Ethanol, molecular biology grade	Sigma-Aldrich	E7023
Nuclease-free Water	Thermo Fisher	AM9930
Critical commercial assays		
RNeasy MinElute Cleanup Kit	Qiagen	74204
DNeasy Blood and Tissue Kit	Qiagen	69504
Agilent SureSelect XT2	Agilent	G9621A
Agilent SureSelect XTHS (Low Input)	Agilent	G9703A
Agilent SureSelect Custom DNA Target Enrichment Probes Tier 1 (500Kb)	Agilent	5190-4813
Afilent SureSelect XT Human All Exon v6 + UTR	Agilent	5190-8881
AMPPure XP Kit	Beckman Coulter	A63880
Herculase II Fusion DNA Polymerase	Agilent	600677
Dynabeads MyOne Streptavidin T1	Thermo Fisher	65601
Qubit dsDNA HS Assay Kit	Thermo Fisher	Q32851
Deposited data		
Medulloblastoma methylation array dataset E-MTAB-10754	This paper	Array Express: E-MTAB-10754
Medulloblastoma RNA-seq dataset E-MTAB-10767	This paper	Array Express: E-MTAB-10767
Medulloblastoma Methylation array dataset GSE130051	Sharma et al.	GEO accession: GSE130051
Medulloblastoma Methylation array dataset GSE93646	Schwalbe et al.	GEO accession: GSE93646
Medulloblastoma scRNA-seq dataset GSE119926	Hovestadt et al	GEO accession: GSE119926
Human fetal cerebellum scRNA-seq dataset	Human Cell Atlas (https://www.covid19cellatlas.org/aldinger20)	dbGAP accession: phs001908.v2.p1
Software and algorithms		
R v3.5.3 & v4.0.2 & R base packages	https://www.r-project.org	
Bioconductor	http://bioconductor.org	
Kallisto v0.46.0	https://pachterlab.github.io/kallisto/about	
RNA-STAR v2.7.0e	https://github.com/alexdobin/STAR	
HTSeq v0.9.1	https://htseq.readthedocs.io/en/master/	
SAMtools v1.9,	http://www.htslib.org	
Picard v2.2.4	https://broadinstitute.github.io/picard/	
QEdit	https://github.com/BioinfoUNIBA/QEdit	

Genome Analysis Toolkit (GATK) version 3.7	https://gatk.broadinstitute.org/hc/en-us	
Ensembl Variant Effect Predictor (VEP)	https://www.ensembl.org/info/docs/tools/vep/index.html	
DESeq2_1.22.2	bioconductor package: http://bioconductor.org	
minfi_1.28.4	bioconductor package: http://bioconductor.org	
NMF_0.23.0	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
limma_3.38.3	bioconductor package: http://bioconductor.org	
sva_3.36.0	bioconductor package: http://bioconductor.org	
tximport_1.10.1	bioconductor package: http://bioconductor.org	
tximportData_1.10.0	bioconductor package: http://bioconductor.org	
caret_6.0-86	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
DMRcate_1.18.0	bioconductor package: http://bioconductor.org	
Rtsne_0.15	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
biomaRt_2.38.0	bioconductor package: http://bioconductor.org	
ggplot2_3.3.2	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
SingleCellExperiment_1.4.1	bioconductor package: http://bioconductor.org	
Seurat_3.2.0	bioconductor package: http://bioconductor.org	
survival_3.2-7	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
tidyverse_1.3.0	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
mclust_5.4.6	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	

fgsea_1.8.0	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
vcdExtra_0.7-1	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
survminer_0.4.8	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
GSVA_1.30.0	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
Hmisc_4.4-1	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
lumi_2.34.0	bioconductor package: http://bioconductor.org	
e1071_1.7-3	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
mlbench_2.1-1	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
randomForest_4.6-14	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
DoseFinding_0.9-17	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
car_3.0-10	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
gplots_3.1.0	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	

RColorBrewer_1.1-2	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
pheatmap_1.0.12	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
ggridges_0.5.2	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
ggrepel_0.8.2	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
ggnewscale_0.4.3	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
Gviz_1.26.5	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
patchwork_1.0.1	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
SingleCellExperiment_1.10.1	bioconductor package: http://bioconductor.org	
Seurat_4.0.1	bioconductor package: http://bioconductor.org	
bestNormalize_1.7.0	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
ggnewscale_0.4.5	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
scales_1.1.1	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
ggpattern_0.1.3	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	

ggtern_3.3.0	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
MASS_7.3-51.6	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
ggridges_0.5.3	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
plotly_4.9.3	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
ggplot2_3.3.3	cran package: https://cran.r-project.org/web/packages/available_packages_by_name.html	
monocle3_0.2.2	bioconductor package: http://bioconductor.org	

Supplemental information

Supplementary Table 1. Demographic summary of patients.

Supplementary Table 2. Sentrix IDs of samples used in larger methylation cohort.