

Oncogene aberrations drive medulloblastoma progression, not initiation

<https://doi.org/10.1038/s41586-025-08973-5>

Received: 3 January 2024

Accepted: 3 April 2025

Published online: 07 May 2025

Open access

 Check for updates

Konstantin Okonechnikov^{1,2,3,27}, Piyush Joshi^{1,2,3,4,27}, Verena Körber^{5,6,27}, Anne Rademacher⁷, Michele Bortolomeazzi⁸, Jan-Philipp Mallm⁸, Jan Vaillant^{1,2,4,9,10}, Patricia Benites Goncalves da Silva^{1,2,4}, Britta Statz^{1,2,3}, Mari Sepp¹¹, Ioannis Sarropoulos¹¹, Tetsuya Yamada¹¹, Andrea Wittmann^{1,2,9}, Kathrin Schramm^{1,2,9}, Mirjam Blattner-Johnson^{1,2,9}, Petra Fiesel^{1,2,12}, Barbara Jones^{1,2,9,13}, Natalie Jäger^{1,2,3}, Till Milde^{1,2,13,14}, Kristian W. Pajtler^{1,2,3,13}, Cornelis M. van Tilburg^{1,2,13,14}, Olaf Witt^{1,2,13,14}, Konrad Bochennek¹⁵, Katharina Johanna Weber^{16,17,18,19}, Lisa Nonnenmacher²⁰, Christian Reimann²⁰, David R. Ghasemi^{1,2,3,21,22}, Ulrich Schüller^{21,22,23}, Martin Mynarek^{21,24}, Stefan Rutkowski²¹, David T. W. Jones^{1,2,9}, Andrey Korshunov^{1,2,12,25}, Karsten Rippe⁷, Frank Westermann^{1,2,26}, Supat Thongjuea^{1,2,3}, Thomas Höfer⁵, Henrik Kaessmann⁹, Lena M. Kutscher^{1,2,4,28} & Stefan M. Pfister^{1,2,3,13,28}

Despite recent advances in understanding disease biology, treatment of group 3/4 medulloblastoma remains a therapeutic challenge in paediatric neuro-oncology¹. Bulk-omics approaches have identified considerable intertumoural heterogeneity in group 3/4 medulloblastoma, including the presence of clear single-gene oncogenic drivers in only a subset of cases, whereas in most cases, large-scale copy number aberrations prevail^{2,3}. However, intratumoural heterogeneity, the role of oncogene aberrations, and broad copy number variation in tumour evolution and treatment resistance remain poorly understood. To dissect this interplay, we used single-cell technologies (single-nucleus RNA sequencing (snRNA-seq), single-nucleus assay for transposase-accessible chromatin with high-throughput sequencing (snATAC-seq) and spatial transcriptomics) on a cohort of group 3/4 medulloblastoma with known alterations in the oncogenes *MYC*, *MYCN* and *PRDM6*. We show that large-scale chromosomal aberrations are early tumour-initiating events, whereas the single-gene oncogenic events arise late and are typically subclonal, but *MYC* can become clonal upon disease progression to drive further tumour development and therapy resistance. Spatial transcriptomics shows that the subclones are mostly interspersed across tumour tissue, but clear segregation is also present. Using a population genetics model, we estimate medulloblastoma initiation in the cerebellar unipolar brush cell lineage starting from the first gestational trimester. Our findings demonstrate how single-cell technologies can be applied for early detection and diagnosis of this fatal disease.

Intratumoural heterogeneity, a hallmark of cancer, refers to the presence of diverse molecular and functional cell populations within a single tumour⁴. Intratumoural heterogeneity is driven by genetic mutations, transcriptomic or epigenomic plasticity and reprogramming of the microenvironment⁵. The malignant childhood tumour medulloblastoma is heterogeneous², especially in groups 3 and 4. This heterogeneity makes effective treatment of these tumours difficult and contributes to overall low survival rates⁶. Advanced DNA methylation profiling has classified group 3/4 tumours into eight distinct molecular subgroups⁷. In addition, single-cell transcriptomic profiling has unveiled the intricate regulatory activity of transcription factors and signalling pathways that orchestrate cellular diversity^{8,9}. Despite these advances, the role of oncogenes in shaping intratumour heterogeneity remains unknown.

A minority of group 3/4 medulloblastoma tumours harbour single-gene oncogenic drivers, including *MYC*¹⁰ and *MYCN*¹¹ amplifications

as well as *PRDM6* overexpression owing to enhancer hijacking by means of a tandem duplication of the adjacent *SNCAIP* gene². By contrast, most group 3/4 tumours display recurrent, large-scale copy number changes^{2,12,13}, including loss of chromosomes 8 and 11 and gain of chromosome 7 and isochromosome 17q. A fundamental question of which genetic events initiate and drive these tumours remains unanswered. Using single-cell multi-omics and spatial transcriptomic approaches, we determined the interplay between large-scale copy number variants (CNVs) and single-gene somatic events in driving medulloblastoma heterogeneity and evolution.

Driver oncogenic events are subclonal

To understand the clonal genetic events in tumour initiation, evolution and progression, we molecularly profiled a specific tumour cohort with a known amplification of *MYCN* or *MYC* or overexpression of *PRDM6*

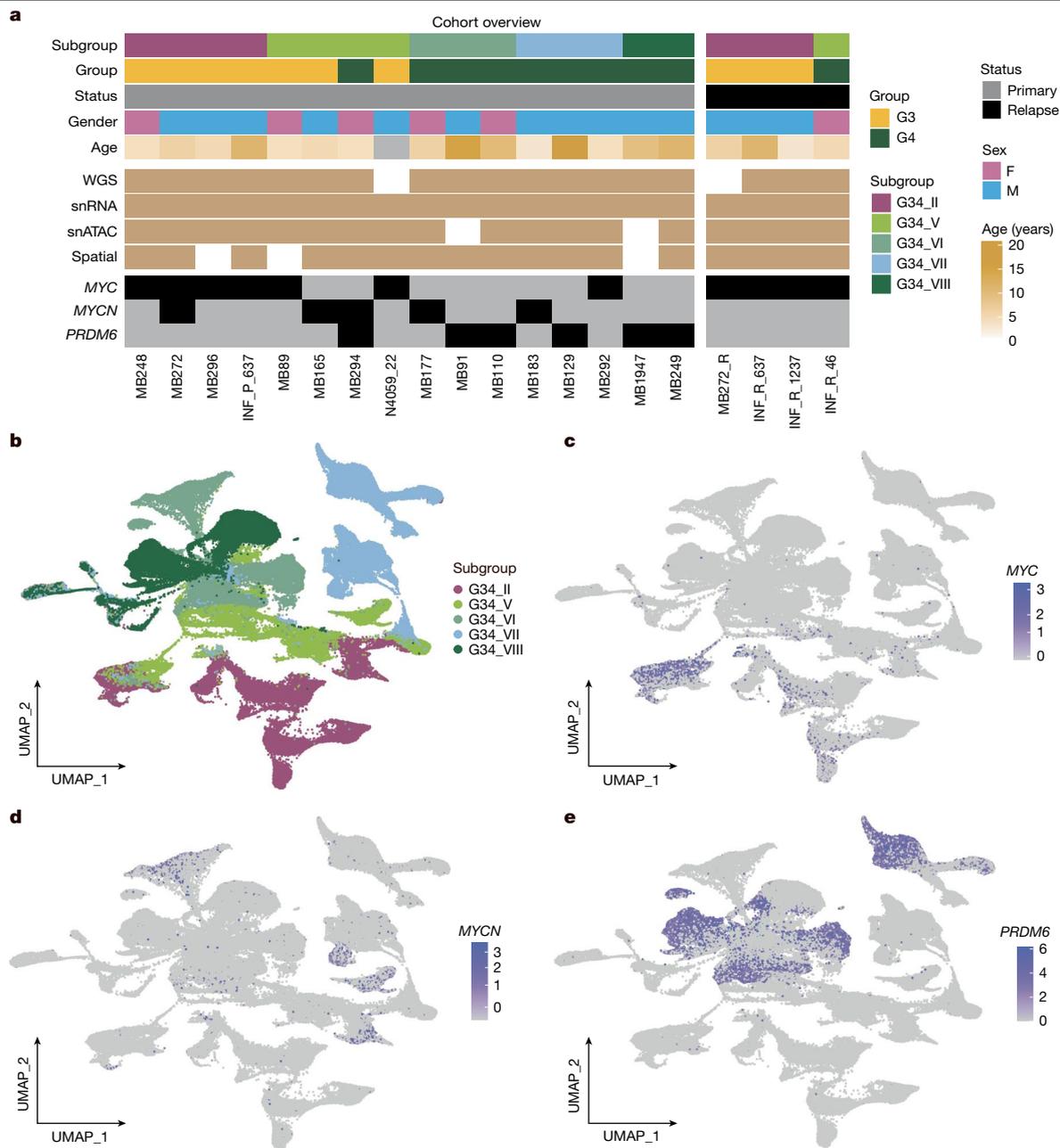


Fig. 1 | Single-nucleus transcriptional profiling of 16 oncogene-associated group 3/4 medulloblastoma primary tumour samples. a, Overview of target cohort with annotation. Two primary-relapse pairs (MB272/R, INF_P/R_637) are from the same patients. **b**, UMAP of snRNA-seq merged dataset, with

medulloblastoma subgroups annotated. **c–e**, Feature plots showing *MYC* (**c**), *MYCN* (**d**) and *PRDM6* (**e**) expression within the UMAP of the merged snRNA-seq dataset. G3, group 3; G4, group 4; G34, group 3/4 medulloblastoma; F, female; M, male.

($n = 16$ primary, $n = 4$ relapses; Fig. 1a and Supplementary Table 1). In larger datasets, amplification or activation of these oncogenes is present in approximately 30% of group 3/4 medulloblastoma cases², whereas approximately 70% of cases lack a single-gene somatic event. The presence of focal *MYC*/*MYCN* amplifications or the *SNCAIP* tandem duplication was verified using bulk molecular profiles or fluorescence in situ hybridization (FISH) for each sample in this cohort (Supplementary Table 1). We analysed single-nucleus profiles of our target cohort, examining both snRNA-seq ($n = 20$) and snATAC-seq ($n = 16$) in the same nuclei.

Uniform manifold approximation and projection (UMAP) visualization of the snRNA-seq data from primary tumours showed group 3/4 subgroup-specific clusters without batch effect adjustments (Fig. 1b);

mixed normal cell types arising from different samples clustered together as expected (Extended Data Fig. 1a). Expression of the oncogenes *MYC*, *MYCN* and *PRDM6* demonstrated clear sample specificity (Fig. 1c–e and Extended Data Fig. 1b). In addition, expression of known marker genes delineated non-tumour cell types, including *PTPRC* (microglia), *IGFBP7* (meningeal) and *AQP4* (astroglia) (Extended Data Fig. 1c–e). Non-tumour cell clusters were also verified from CNV profiling of the full combined snRNA-seq dataset (Extended Data Fig. 1f). This separation of structure was recapitulated with snATAC-seq, as visualized by means of UMAP (Extended Data Fig. 1g,h). For samples with multi-omics data, non-tumour cells in snATAC-seq were labelled on the basis of their associated non-tumour clusters from the snRNA-seq data (Extended Data Fig. 1g).

To investigate the clonal heterogeneity of *MYCN*-amplified tumour samples (subgroups V/VII), we adapted the inferCNV¹⁴ approach (Methods) to infer CNV profiles of cell clusters per sample, using both snRNA-seq and snATAC-seq data. To verify the single-cell CNV calls, we calculated the correlation between pseudobulk from single-cell and bulk DNA methylation CNV profiles (Supplementary Table 1) across all cases, as shown in Extended Data Fig. 2a–c. A full cohort cross-comparison demonstrated that most of the snRNA-seq pseudobulk CNV profiles matched well, with the exception of $n = 3$ false positive cases (Extended Data Fig. 2d). CNV profiles from snATAC-seq data showed the highest correlation to the correct control bulk profile for all samples (Extended Data Fig. 2e), demonstrating the benefit of using this data type for CNV calling.

From inspection of the CNV results per sample, in most cases, we observed clusters with discordant CNVs, which we labelled as subclones. For example, in all *MYCN*-amplified tumours ($n = 4$), we identified two distinct subclones: with (C1) and without (C2) *MYCN* amplification, respectively (Fig. 2a and Extended Data Fig. 3a). Notably, in all cases, reconstruction of the putative phylogenetic trees showed that *MYCN* amplification was not the initiating event for the tumour. Instead, large-scale CNVs, such as loss of chromosomes 8 and 10 or gain of chromosome 7 or 17q, were already present in the presumptive founder clone (C0) (Fig. 2b). Moreover, further unique CNVs were found only within *MYCN*- and non-*MYCN*-amplified subclones. Detailed visualizations of single-cell CNV profiles per sample are available through the interactive online web application (Methods).

Next, we examined the differentiation, proliferation and aggressive progenitor-like activity states of individual cells within each subclone using snRNA-seq expression of established reference gene lists for these defined medulloblastoma cell states⁸. We identified that both *MYCN*-amplified and non-amplified subclones maintained separate proliferating and differentiated compartments (Fig. 2c,d and Extended Data Fig. 3b,c). The *MYCN* subclone was also uniquely enriched with a progenitor-like gene expression signature (Fig. 2e). As cells differentiated, the oncogene itself showed lower expression within the *MYCN*-amplified subclone (Pearson correlation = -0.23 , $P < 2.2 \times 10^{-16}$; Extended Data Fig. 3d), demonstrating that *MYCN* is connected to the undifferentiated state¹⁵. Similar differentiation levels among subclones and a slight bias towards progenitor activity in *MYCN*-amplified subclones were observed in all four *MYCN*-amplified tumours (Extended Data Fig. 3e).

By inspecting differentially expressed genes specific for each subclone (Supplementary Table 2), we also identified a unique property of non-*MYCN* subclones with a stronger enrichment of genes expressed in unipolar brush cell progenitors ($P < 1.14 \times 10^{-6}$), the cell of origin of group 3/4 medulloblastoma^{16,17}, whereas *MYCN* subclone-associated genes were not enriched in these genes ($P > 0.05$). This observation was also confirmed using subclone-specific genes associated with *cis*-regulatory elements derived from the integration of snATAC-seq data (Supplementary Table 3).

Performing single-cell CNV analyses on *MYC*-amplified tumour samples (subgroups II/V), we identified a subclonal *MYC* amplification in six of seven samples (Fig. 2f and Extended Data Fig. 3f). Similar to *MYCN*-amplified tumours, the common and likely initiating events in the founder clone (C0) were large-scale chromosome 10 loss and/or chromosome 17q gain, with subclonal *MYC* amplification occurring later during tumour evolution. Remarkably, the clonal structure of *MYC*-amplified tumours was more complex ($n = 3$ of 7 cases), with the formation of three or more unique subclones (Fig. 2f, bottom). Typically, *MYC*-amplified subclones had their own proliferating and differentiating compartments (Fig. 2g,h and Extended Data Fig. 3g,h). Similar to the *MYCN*-amplified clones, only *MYC*-amplified clones demonstrated strong enrichment of progenitor-like activity compared with non-*MYC*-amplified compartments (Fig. 2i). Differentially expressed genes specific to *MYC*-amplified subclones were enriched

in known *MYC* target genes¹⁸ ($P < 1.11 \times 10^{-16}$; Supplementary Table 2) and *MYC* expression decreased as cells differentiated (Pearson correlation = -0.18 , $P < 2.2 \times 10^{-16}$; Extended Data Fig. 3i), whereas expressed genes in non-*MYC* subclones demonstrated enrichment in unipolar brush cell-related genes ($P < 1.54 \times 10^{-11}$; Supplementary Table 2). The *MYC* subclone-specific genes were not found to be enriched in any corresponding subclone-specific CNVs. Across six samples with subclonal *MYC* amplifications, the differentiation level was similar among subclones; however, progenitor-like activity was significantly enriched in *MYC*-amplified subclones (Extended Data Fig. 3j).

Last, we examined tumours with enhanced *PRDM6* expression (subgroups VII/VIII), in which *SNCAIP* gene duplication leads to aberrant activation of *PRDM6* by means of enhancer hijacking². In our cohort, we identified three of five samples in which *PRDM6* overexpression was subclonal (Fig. 2j and Extended Data Fig. 4a). Chromosome 17q gain was the most frequent CNV within the founder clone (C0). In contrast to *MYC* and *MYCN* clones, we could not identify a distinct proliferating compartment in *PRDM6*-specific clones (Fig. 2k,l). Instead, we found only an overall small proportion of cells (less than 5%) with the proliferation gene signature in *PRDM6* subclones (Extended Data Fig. 4b). We also did not identify enriched progenitor-like activity in the *PRDM6* subclones (Fig. 2m and Extended Data Fig. 4c), except in one specific case in which we detected a *MYCN*-amplified subclone that additionally harboured a *SNCAIP* duplication with associated *PRDM6* overexpression (Fig. 2b, bottom right).

To further verify the presence of subclones, we performed single-cell whole-genome DNA and RNA sequencing from the same cells, in a subset of *MYCN*- and *MYC*-amplified cases, $n = 6$ (Supplementary Table 1). In all tested samples, the presence of the corresponding specific subclones and their CNV profile matched between the snMulti-omic data (combined snRNA-seq and snATAC-seq data from the same cell), on the basis of single-cell RNA projection (Extended Data Fig. 5a) and whole-genome sequencing (WGS) CNV analysis (Extended Data Fig. 5b–h).

Despite the low mutational burden in medulloblastoma², we also inspected the somatic single-nucleotide variants (SNVs) to confirm the tumour phylogeny composition predicted by snRNA-seq and snATAC-seq data. Using WGS data, we examined mutations in CNV regions specific to the founder clone (C0) or not lying within CNVs. In 9 of 12 samples, we did not identify the presence of drivers or co-mutations (exceptions: two *MYC*-amplified cases, Fig. 2f, and one *PRDM6* case, Fig. 2j). We further investigated the mutational landscape at the single-cell level by increasing the sequencing coverage of the snATAC-seq data in three *MYC*-amplified cases. This approach allowed us to recover up to 40% (range, 20% to 60%) of somatic mutations per sample (Supplementary Table 4). The positive correlation (maximum P value: 7.4×10^{-6}) of their variant allele frequency (VAF) and the corresponding bulk WGS profiles verified the accuracy of the approach (Extended Data Fig. 4d–f). Importantly, in all three cases, it was possible to identify unique somatic mutations specific for the respective subclone, further supporting a common origin of these clones and secondary subclonal evolution (Extended Data Fig. 4g–i). Moreover, by comparing the extended somatic SNVs identified in the snATAC data with germline SNVs (see the Methods for details), the subclone-specific SNVs were observed to have approximately four times lower mean VAF in comparison with those SNVs that are common among subclones (Extended Data Fig. 4j–l).

Collectively, these findings nominate large-scale CNVs as likely tumour-initiating events in group 3/4 medulloblastoma, with focal oncogene aberrations occurring only during tumour evolution.

Tumour onset from first trimester onwards

To investigate clonal dynamics during the initiation of group 3/4 medulloblastomas, we analysed WGS data from the medulloblastoma

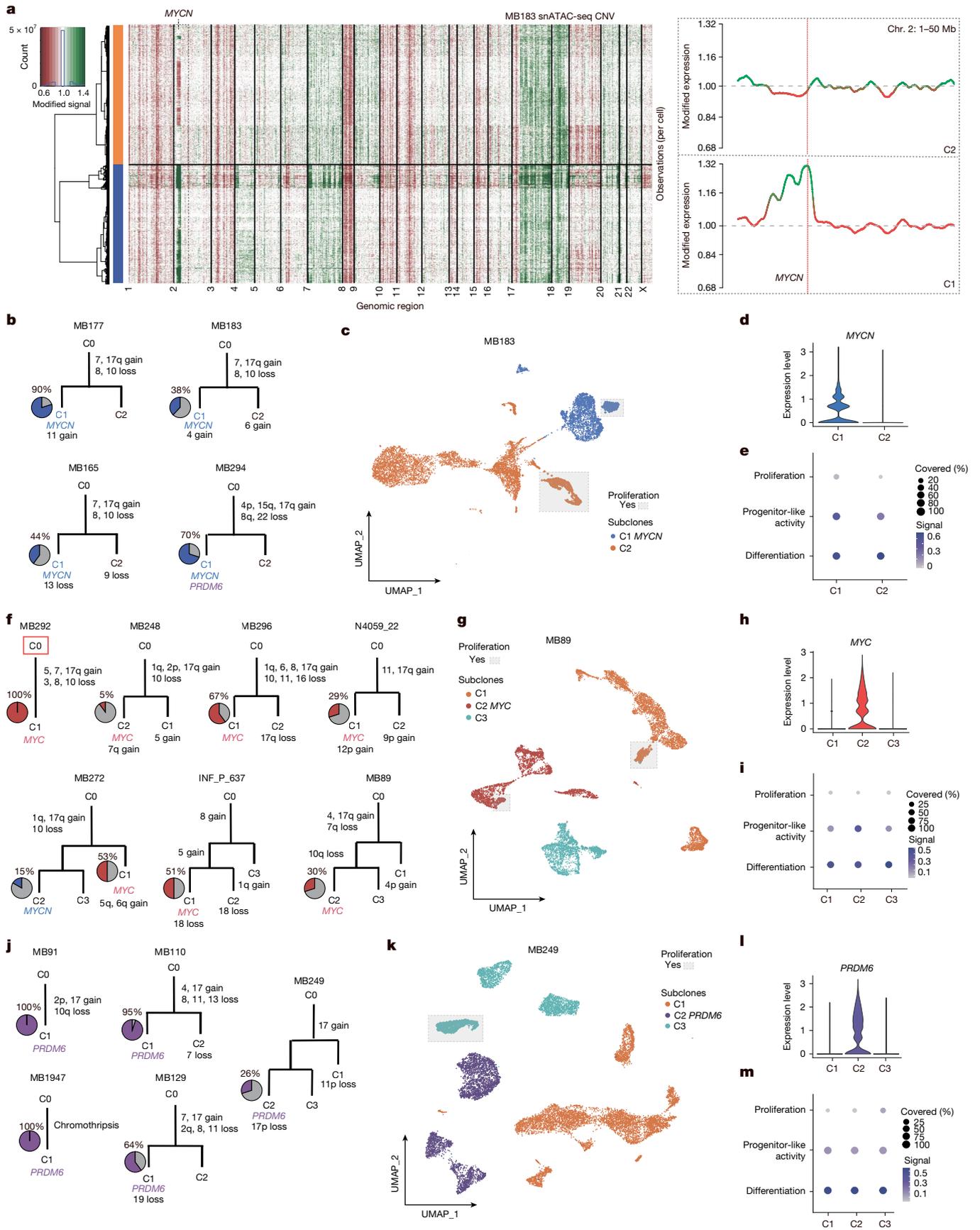


Fig. 2 | See next page for caption.

Fig. 2 | Clonal proliferation and differentiation gradients are independent of oncogene expression. **a**, Copy numbers derived from snATAC-seq data in *MYCN*-amplified sample MB183. Red, chromosome loss. Green, chromosome gain. Right side, inset of *MYCN* chromosome region. **b**, Somatic phylogeny trees for *MYCN* samples. Blue, proportion of *MYCN*-expressing cells. **c**, snRNA-seq UMAP of single *MYCN* sample MB183. Grey boxes, proliferating cell clusters with strong proliferation enrichment. Blue, C1 clone. Orange, C2 clone. **d**, *MYCN* expression in C1 and C2 clones. **e**, Per cell gene set variance analysis (GSVA) enrichments of proliferation, progenitor-like activity and differentiation in single sample shown in **c**. **f**, Somatic phylogeny trees for *MYC* samples. Red, proportion of *MYC*-expressing cells. Red square, cases MB292 and MB248 with somatic mutations in C0. **g**, snRNA-seq UMAP of single *MYC*

sample MB89. Grey boxes, proliferating cell clusters with strong proliferation enrichment. Red, *MYC*-expressing C2 clone. Orange, C1 clone. Aquamarine, C3 clone. **h**, *MYC* expression in C1, C2 and C3 clones. **i**, Per cell GSVA enrichments of proliferation, progenitor-like activity and differentiation in single sample shown in **g**. **j**, Somatic phylogeny trees for *PRDM6* samples. Purple, proportion of *PRDM6*-expressing cells. Red square, case MB249 with somatic mutations outside CNV regions. **k**, snRNA-seq UMAP of single *PRDM6* sample MB249. Grey box, proliferating cell cluster with strong proliferation enrichment. Purple, *PRDM6*-expressing C2 clone. Orange, C1 clone. Aquamarine, differentiation signal enrichment in C3 clone. **l**, *PRDM6* expression in C1, C2 and C3 clones. **m**, Per cell GSVA enrichments of proliferation, progenitor-like activity and differentiation in single sample shown in **k**.

International Cancer Genome Consortium (ICGC) cohort² (Supplementary Table 5). Somatic tissues accumulate SNVs continuously over time^{19–21}, and hence SNV density in the tumour cell of origin (in population genetics, ‘most recent common ancestor’ (MRCA)) can be interpreted as a measure for the patient’s age at tumour initiation^{22,23}. To time the developmental origin of medulloblastoma with this approach, we quantified clonal SNV densities from the allele frequency distribution of somatic variants in 181 primary medulloblastomas of all subgroups (Extended Data Fig. 6a,b; comprising 108 group 3/4 medulloblastomas, 21 infant Sonic Hedgehog (SHH)-medulloblastomas, 35 childhood/adulthood SHH-medulloblastomas and 17 WNT-medulloblastomas). Overall, the clonal SNV densities across subgroups recapitulated the age-incidence distribution of the disease, with infant SHH-medulloblastoma having the lowest densities (0.02 ± 0.01 SNVs per megabase (Mb)), followed by group 3/4 medulloblastoma (0.1 ± 0.08 SNVs per Mb), WNT-medulloblastoma (0.28 ± 0.47 SNVs per Mb) and adult SHH-medulloblastoma (0.41 ± 0.46 SNVs per Mb; Extended Data Fig. 6c). Clonal SNV densities were also correlated with age at diagnosis (Spearman’s $\rho = 0.73$, $P < 2.2 \times 10^{-16}$; Extended Data Fig. 6d), collectively supporting our approach to infer the evolutionary dynamics at medulloblastoma onset from somatic SNVs.

To estimate age of tumour initiation in group 3/4 medulloblastomas, we analysed 109 tumour samples of this subgroup in more detail (Fig. 3a). As with the entire cohort, clonal SNV densities were likewise correlated with the age at diagnosis among group 3/4 medulloblastoma (Spearman’s $\rho = 0.51$, $P = 3.959 \times 10^{-8}$; Extended Data Fig. 6e). However, contrary to the clear temporal order in tumour initiation of the major medulloblastoma groups, clonal SNV densities were statistically indistinguishable between group 3/4 medulloblastoma subgroups I–VIII (Wilcoxon rank sum test, all adjusted P values > 0.05), indicating that growth of the final tumour mass begins around the same developmental time window in all group 3/4 medulloblastoma subgroups (Fig. 3b).

To refine our analysis, we timed the acquisition of clonal CNVs (copy number gains or loss of heterozygosity (LOH)) relative to the tumour’s MRCA. To this end, we compared densities of clonal SNVs acquired before a chromosomal gain, and hence present on multiple copies of a chromosomal region, with the density of clonal SNVs overall (Methods). Similar to neuroblastoma²³ and other tumour entities²², 34 of 109 group 3/4 medulloblastomas showed evidence of having acquired at least one copy number gain in an early common ancestor (ECA), antecedent to the tumour’s MRCA (Fig. 3c,d). The number of such early CNVs varied between 1 and 16 per tumour (mean, 5.3), with no significant difference in SNV density between early CNVs within a tumour (Extended Data Fig. 6f). Hence, in tumours with early CNVs, all early CNVs probably arose in an ECA during a confined time window before the onset of tumour growth. In the remaining cases, where we identified no early CNVs, clonal chromosomal gains probably occurred concomitantly with, or shortly before, the onset of tumour growth. To corroborate these observations, we contrasted our approach²³ with an alternative computational tool, MutationTimeR²², which yielded similar results (Extended Data Fig. 6g). Hence, our data suggest that

at least some CNVs in group 3/4 medulloblastoma arise before the onset of tumour growth, in line with multiple rounds of mutation and selection at tumour initiation.

To date these events in actual time, we calibrated a population genetics model of mutation and selection during tumour initiation²³ with the measured SNV densities at ECA and MRCA, along with the patient age at diagnosis (see the Methods for details). Briefly, the model assumes that medulloblastoma initiation is driven by clonal selection for two consecutive drivers in the transient cell population of (differentiating) unipolar brush progenitor cells from the rhombic lip^{16,17,24} (Extended Data Fig. 6h). Acquisition of the first driver defines a pre-malignant state, arising before a tumour’s MRCA. Although in principle the first driver can be any type of mutation, small driver mutations are overall rare in group 3/4 medulloblastoma², and thus are probably not the major driving force of early tumour evolution. Hence, we focused on cases in which early CNVs defined an ECA, and associated the time point at which the first driver mutation was acquired with the mutation density in the ECA. We assumed that the second driver emerges subsequently in the pre-malignant clone, spawned by the ECA. Hence, we associated the mutation density in the MRCA with the acquisition of the second driver and the onset of tumour growth. Finally, upon malignant transformation of the tumour’s MRCA, we assumed exponential growth to a tumour size of 10^9 cells (corresponding to a few cubic centimetres) at the age of diagnosis²⁵ (see Methods for details).

Using the bespoke model, we estimated driver mutation rates and associated selective advantages from the clonal SNV densities at ECA and MRCA measured across group 3/4 medulloblastomas. Simultaneously, we estimated per-tumour doubling times using age and the sub-clonal VAFs of 35 tumours with sufficient data quality and information on age at diagnosis (Supplementary Table 5). Consistent with a higher activity of S-phase genes and *MYC* target genes and poorer overall survival²⁶ (Extended Data Fig. 6i–k), we estimated shorter tumour doubling times in medulloblastomas at the group 3 pole as compared with tumours at the group 4 pole (Extended Data Fig. 6j), confirming our modelling approach. We then used the per-tumour doubling times to translate SNV densities at ECA and MRCA into real-time, finding that the first oncogenic event (that is, the ECA) occurs within the first gestational trimester in 24% of cases, during late gestation in around 35% of cases and within the first year of life in 26% of cases (Fig. 3e and Extended Data Fig. 6l). The onset of tumour growth from its MRCA is placed considerably later, within the first decade of life (Fig. 3f), suggesting a long latency phase between pre-malignancy and the detection of a symptomatic tumour. Overall, the inferred dynamics of tumour initiation are consistent with a tumour origin in (differentiating) unipolar brush progenitor cells^{24,27}, sustaining a pre-malignant clone that outlives the cell state of origin for several years (Fig. 3g).

Early acquisition of large-scale CNVs

To gain mechanistic insight into group 3/4 medulloblastoma initiation, we asked whether particular mutations occur predominantly early

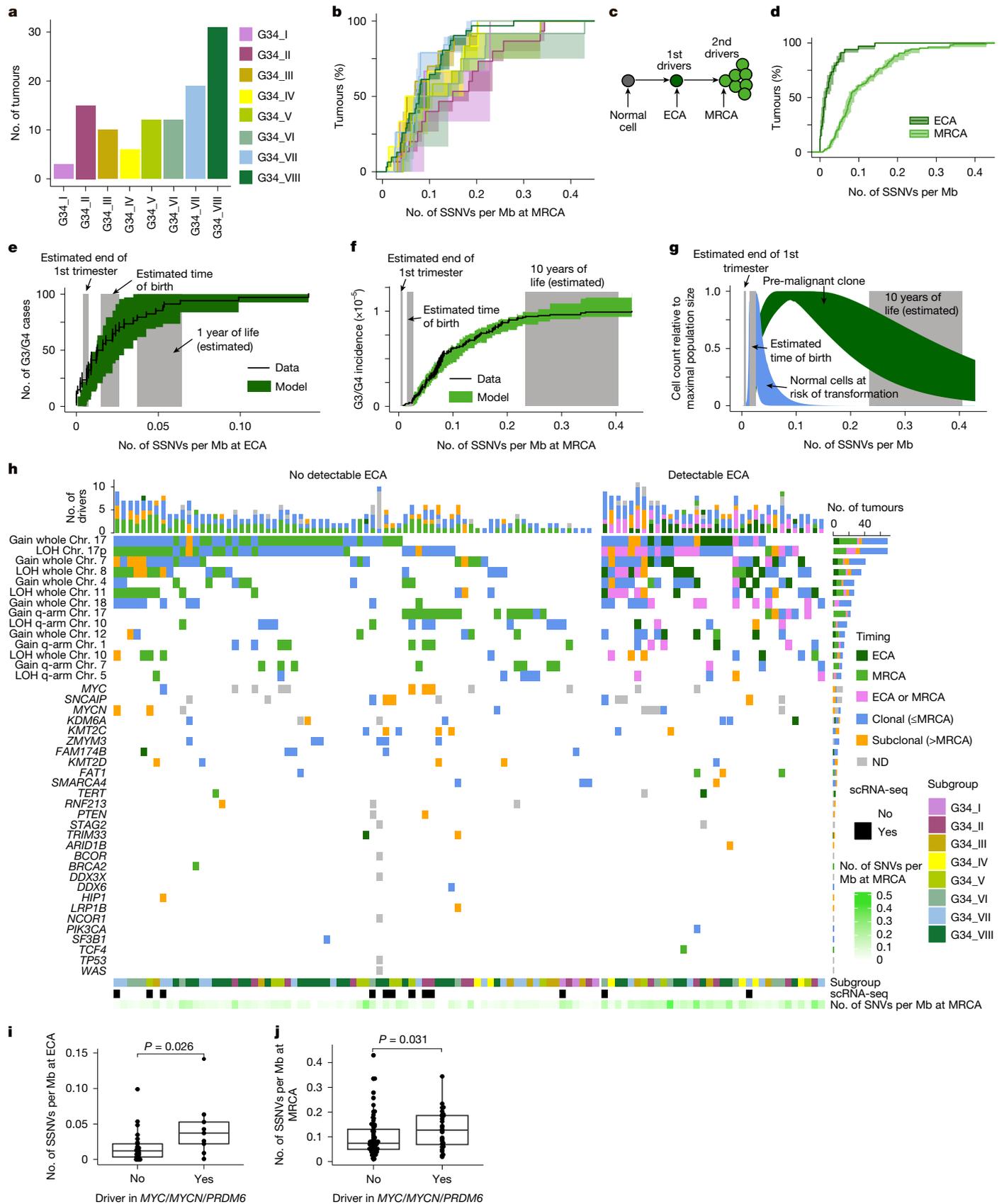


Fig. 3 | See next page for caption.

Fig. 3 | Somatic mutation profiles and association with cell of origin. **a**, Group 3/4 medulloblastoma subgroups analysed by bulk WGS. **b**, SNV densities at MRCA per group 3/4 medulloblastoma subgroup (I, $n = 3$; II, $n = 15$; III, $n = 10$; IV, $n = 6$; V, $n = 12$; VI, $n = 12$; VII, $n = 19$; VIII, $n = 31$). Shown are mean and 95% CI (estimated by bootstrapping the genomic segments 1,000 times). **c**, Early medulloblastoma evolution. Driver mutation in an ECA spawns a pre-malignant lesion. Malignant transformation occurs upon further drivers in the tumour's MRCA. **d**, SNV densities at ECA and MRCA for group 3/4 medulloblastoma ($n = 108$). Mean and 95% CI, estimated by bootstrapping the genomic segments 1,000 times. **e**, Model fit to SNV densities at ECA. Line, mean and standard deviation (estimated by bootstrapping the genomic segments 1,000 times) of the measured SNV densities; green and grey areas, 95% credible interval of the model fit, and of key time points. **f**, As in **e**, but for SNV densities at MRCA.

g, 95% credible intervals of modelled tissue of origin (blue) and pre-malignant clone (green). Grey areas as in **f**. **h**, Mutation spectrum with timing information ('ECA', CNV uniquely timed to ECA; 'MRCA', CNV uniquely timed to MRCA; 'ECA or MRCA', CNV in agreement with both ECA and MRCA; 'clonal', CNV/small mutation was clonal, no further mapping to ECA/MRCA possible; 'subclonal', CNV/small mutation was subclonal; ND, no data). Subclonality information for amplification of *MYC/MYCN* and duplication of *SNCAIP* from single-cell data. **i**, SNV density at ECA in group 3/4 medulloblastoma with and without driver in *MYC/MYCN/PRDM6*. *P* value, unpaired Wilcoxon rank sum test ($n = 80$ without, $n = 28$ with driver). **j**, As in **i**, but for SNV density at MRCA. 95% CI, 95% confidence interval; scRNA-seq, single-cell RNA-seq; SSNV, somatic single-nucleotide variants.

or late. To address this question, we first focused on CNVs that were found more frequently than expected by chance, and hence are likely drivers of malignancy. Combining the enrichment results obtained in our cohort (Extended Data Fig. 7a and Methods) with published data²⁸, we classified gains of chromosomes 1q, 4, 7, 12, 17/17q and 18, as well as LOH on chromosomes 5q, 8, 10/10q, 11 and 17p, as putative drivers of group 3/4 medulloblastoma initiation. Except for four cases, in which no ECA was identifiable, all group 3/4 medulloblastomas harboured at least one of these CNVs clonally (Fig. 3h). To search for putative drivers located on these regions, we analysed the expression of genes lying in commonly gained or lost regions. For this purpose, we contrasted gene expression between tumours with and without particular CNVs in a bulk RNA-seq cohort of group 3/4 medulloblastoma³ (Supplementary Table 6a) and inspected differentially expressed genes specific for group 3/4 medulloblastoma in a global central nervous system tumour cohort²⁴ (Supplementary Table 6b). Among these identified genes, up to 10% have been described as known somatic drivers, and thus associated with medulloblastoma evolution²⁹.

Overall, gains of chromosome 17 or 17q were the most frequent aberrations in group 3/4 medulloblastoma, followed by gain of whole chromosome 7 and LOH of whole chromosome 8. The SNV densities at chromosomal gains were often smaller than the SNV densities at the tumour's MRCA, in particular for gains of chromosomes 4, 7, 12 and 17 and LOH of chromosomes 8 and 11 (Extended Data Fig. 7b). Although we cannot rule out that small mutations or chromosomal losses preceded these gains, their consistent early timing suggests that chromosomal gains or losses might be among the earliest events during group 3/4 medulloblastoma initiation. In contrast to the high abundance of CNVs, focal events in known driver genes (SNVs, indels, focal amplifications/deletions or structural rearrangement) were overall rare (Fig. 3h). Among these, amplification of *MYC* or *MYCN* and duplication of *SNCAIP* leading to *PRDM6* overexpression were the most frequent alterations (Fig. 3h). However, the single-cell analysis (compare with Fig. 2b,f,j) showed that these mutations were mostly subclonal. Interestingly, group 3/4 medulloblastomas with amplification of *MYC* or *MYCN* or duplicated *SNCAIP* had significantly higher SNV densities at both ECA (Fig. 3i) and MRCA (Fig. 3j) than the remaining tumours, suggesting that later onset of (pre-)malignancy may predispose to the subsequent acquisition of these drivers. In general, the mutational landscape in group 3/4 medulloblastoma suggests a fundamental role of CNVs during tumour initiation, although further mutations acquired during disease progression seem to drive subclonal evolution in a subset of tumours only.

Subclonal spatial heterogeneity

To better understand the spatial relationship of the tumour subclones and disclose further insight into their evolution, we performed spatial transcriptomics on samples with available material ($n = 13$ primary, $n = 4$ relapse). We used technology that applies multiplexed in situ

hybridization of a selected gene set to achieve single-cell spatial resolution of a tumour sample (Supplementary Tables 7 and 8). UMAP visualization of merged spatial data from primary tumour samples reflected combined snMultiomic profiling, allowing us to distinguish subgroup-specific properties and identify non-tumour cell types (Extended Data Fig. 8a–e). The spatial locations of tumour cells were determined by the expression of *MYC*, *MYCN* and *PRDM6* (Fig. 4a), along with other genes associated with proliferation (for example, *MKI67*; Extended Data Fig. 8f). The tumour microenvironment, including glial, immune and meningeal cells, was characterized using cell type-specific markers (Extended Data Fig. 8f).

To determine the spatial distribution of the identified subclones, we projected the snRNA-seq data onto the spatial data (Fig. 4a, last row). Overall, we distinguished two major spatial localization patterns: interspersed, in which independent subclones mixed throughout the tumour sample, and segregated, in which a clear boundary between independent subclones could be delineated. In most cases, the subclones were interspersed, as observed from the spatial distribution of the corresponding marker gene expression.

In *MYCN*-amplified tumours, the observed clonal architecture derived from snRNA-seq data was also present in the spatial data (Fig. 4b). The subclones exhibited an interspersed spatial pattern (Fig. 4c), with pockets of *MYCN* and non-*MYCN* clones highlighted through neighbourhood enrichment within the tumour sample (Fig. 4c, inset). The proliferating compartments within these subclones were also interspersed across the tumour tissue, observed by *MKI67* expression (Fig. 4e and Extended Data Fig. 8f). On a smaller scale, however, neighbourhood enrichment analysis showed that proliferating cells of subclones clustered together, away from the differentiating cells (Fig. 4d). The normal cells were mostly isolated from the tumour subcompartments.

A similarly interspersed spatial pattern of *MYC* and non-*MYC* subclones was present in *MYC*-amplified tumours (Fig. 4f,g), yet islands of segregated non-*MYC* subclones were also observed (Fig. 4g, inset). Proliferating cells (*MKI67*⁺) were interspersed throughout the tumour tissue (Fig. 4i and Extended Data Fig. 8f). The differentiated tumour cell compartment within the *MYC* subclone (C1-diff) was in closer proximity to the proliferating cell compartment within the same subclone (C1-prolif; Fig. 4h). Normal cells were largely isolated from the *MYC*-amplified subclone compartments.

We observed a segregated spatial separation of subclones in the *PRDM6* sample (Fig. 4j,k,m). This spatial segregation of clones was confirmed in another region of the same tumour specimen (Extended Data Fig. 8g,h). Although the subclones were segregated, the proliferating cell compartments within the subclones were interspersed within the spatial block (Fig. 4n and Extended Data Fig. 8i). As expected, the neighbourhood enrichment analysis in this sample showed the separation of the *PRDM6* clone from other compartments (Fig. 4i). In another sample we also confirmed the dual-oncogene, the *PRDM6-MYCN* subclone (Fig. 2b, bottom right), in which cells with activity in both genes reside in the same spatial regions (Extended Data Fig. 8j,k).

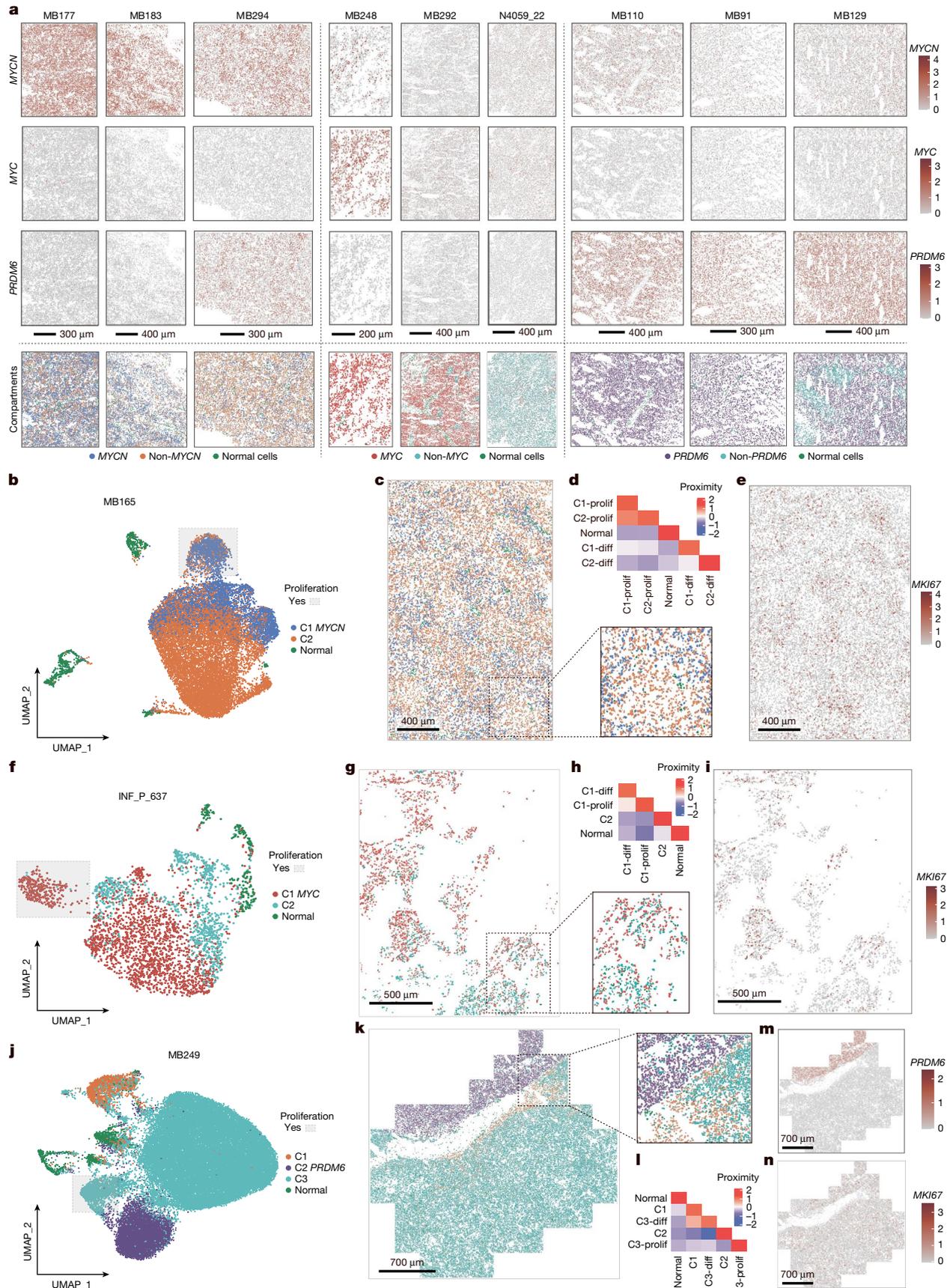


Fig. 4 | See next page for caption.

Fig. 4 | Spatial heterogeneity across oncogene-associated group 3/4 medulloblastoma samples. **a**, Spatial gene expression of *MYC*, *MYCN* and *PRDM6*. Last row, projection of clones derived from snRNA-seq. **b**, Spatial data UMAP of representative *MYCN* sample. **c**, Spatial visualization of clones of sample in **b**. Enlarged view of a fragment in the bottom right. **d**, Proximity of each compartment to each other of sample in **b**. **e**, *MKI67* spatial expression of sample in **b**. **f**, Spatial data UMAP of representative *MYC* sample. **g**, Spatial

visualization of clones of sample in **f**. Magnification of specific region in bottom right. **h**, Proximity of each compartment to each other of sample in **f**. **i**, *MKI67* spatial expression of sample in **f**. **j**, Spatial data UMAP of representative *PRDM6* sample. **k**, Spatial visualization of clones of sample in **j**. Magnification of specific region in top right. **l**, Proximity of each compartment to each other of sample in **j**. **m, n**, *PRDM6* (**m**) and *MKI67* (**n**) spatial expression of sample in **j**. diff, differentiated; prolif, proliferating.

Together, the observed spatial patterns suggest that clonal evolution does not lead to spatial compartmentalization within tumours. Instead, cellular migration may dictate communication among the clones, which can then drive competing or collaborative interactions among the co-existing tumour populations.

MYC subclones take over at relapse

We identified a primary tumour sample with two distinct subclones (MB272), harbouring *MYC* (C3) and *MYCN* (C2) amplifications simultaneously (Figs. 2f and 5a–c). This subgroup II sample was originally characterized as *MYC*-amplified only on the basis of bulk methylation profiling (Extended Data Fig. 9a), whereas the *MYCN*-amplified subclones were observed in the single-cell multiome and spatial transcriptomics analysis. This discrepancy most probably results from examining different fragments of the tumour tissue for each analysis. The *MYC* and *MYCN* subclones in this tumour sample had proliferating and differentiating compartments (Fig. 5b,c). The *MYC* subclone was strongly enriched with progenitor-like activity (Extended Data Fig. 9b). Remarkably, a clear spatial separation was observed between *MYC*- and *MYCN*-expressing cells (Fig. 5d and Extended Data Fig. 9c,d). This spatial segregation (Fig. 5e,f) reflected the phylogenetic tree of tumour evolution projected from the snRNA-seq CNV annotation (Fig. 2f). Further sets of subclone-specific genes also showed explicit spatial specificity (Extended Data Fig. 9e,f). As expected, low contact proximity was identified between *MYC* and *MYCN* subclones (Extended Data Fig. 9g).

According to current knowledge, *MYC* and *MYCN* amplifications are considered mutually exclusive events³⁰ in medulloblastoma (Extended Data Fig. 9h) and other tumour types³¹. Because of the unexpected occurrence of both amplifications in this case, we conducted a systematic analysis across a larger medulloblastoma cohort to identify further cases in which these oncogenes may co-occur. We identified six putative cases on the basis of DNA methylation CNV profiles (Extended Data Fig. 9i). Using immunohistochemistry, we identified another case in which both *MYC* and *MYCN* staining were seen in the same tumour sample (Extended Data Fig. 6j). During the preparation of this manuscript, a case study reported a primary tumour sample in which both *MYC*- and *MYCN*-amplified cells were present³². Together, these independent cases suggest the possibility that *MYC* and *MYCN* amplifications co-occur within the same tumour more frequently than originally thought. Nevertheless, it is noteworthy that individual cells within the tumour express only one of these oncogenes and are spatially segregated.

The presence of *MYC* and *MYCN* subclones cannot be distinguished using bulk profile techniques owing to the potential low presence of cells of a particular subclone in the obtained data. Therefore, we generated unique signatures of *MYC* and *MYCN* subclones derived from single-cell data to identify further samples harbouring two oncogene amplifications. We performed a deconvolution analysis of bulk transcriptome profiles, using the *MYC/MYCN* case as the reference control. Using this method, we detected further samples in which *MYC* and *MYCN* subclones may co-occur in the same sample (Extended Data Fig. 9k). We validated this finding using FISH on an identified sample with available material (Extended Data Fig. 6l).

We next checked whether this information could be exploited for diagnostic purposes. Therefore, we investigated whether the relative

presence of *MYC* or *MYCN* subclones derived from the deconvolution analysis predicted patient outcomes. In subgroup V, 4 of 41 cases harboured a known *MYC* amplification, on the basis of CNV profiles, and correlated with a low probability of survival (Extended Data Fig. 9m). We identified 14 potential cases with an occurrence of an *MYC*-amplified subclone on the basis of deconvolution. These patients had a lower overall survival (Extended Data Fig. 9n). Therefore, the poor outcomes of subgroup V patients may be explained by an undiagnosed *MYC* subclone that potentially outcompetes other subclones to drive relapse.

To further test this possibility, we performed single-nucleus molecular profiling on four relapse *MYC*-amplified cases. In all relapse cases, new subclones arose, but all tumour cells harboured the *MYC* amplification (Fig. 5g). For example, in the matched relapse *MYC/MYCN* case, the *MYCN* subclone was lost at relapse (Fig. 5h). This loss of *MYCN* expression was confirmed using spatial transcriptomics (Fig. 5i and Extended Data Fig. 9o,p). These results suggest that the *MYC* subclone outcompetes other subclone(s) during tumour progression and hence the presence of subclonal *MYC* amplification at diagnosis may predict the probability of relapse.

Discussion

Despite advances in understanding the cellular origin of group 3/4 medulloblastoma, the tumour-initiating and driving mechanisms remain elusive. In our study, we use single-cell multi-omics data analysis combined with spatial profiling to identify the somatic subclone properties within these tumours and demonstrate that the genetic aberrations that lead to overexpression of oncogenes are not the likely initiating events in group 3/4 medulloblastoma. Therefore, *MYC* and *MYCN* are probably not the primary ‘drivers’, but instead are acquired after malignant transformation and probably accelerate tumour growth. Instead, our results suggest that the initiating or ‘driving’ events in group 3/4 medulloblastoma are large-scale CNVs. This finding is in line with the hypothesis that tetraploidization is a frequent early event in medulloblastoma¹² and is associated with intermediate survival rates and high risk of relapse³³. Using mutational clocks and mathematical modelling, we find that medulloblastoma initiation is probably a multi-step process. Specifically, our model suggests that early CNVs, acquired as early as in fetal development, drive a pre-malignant clone, in which malignant transformation occurs within the first decade of life. Although this process is similar to observations in mouse models, in which an initial hyperplastic state precedes medulloblastoma growth^{34,35}, our modelling approach has some limitations. We here assume that unipolar brush cell (UBC) progenitors, the likely cell of origin of group 3/4 medulloblastomas^{16,17}, divide at a fairly constant rate and that tumour growth can be approximated with exponential growth. Both assumptions may oversimplify the true biology, which may influence our timing estimates. Reassuringly, however, a disease origin that dates back to fetal development, or the first year of life, agrees with the detection of UBC progenitors in human brain samples from this time span²⁴. Intriguingly, another paediatric tumour, neuroblastoma, has a similar order of genetic events: CNVs are the initiating event and occur early in the first trimester of pregnancy²³. How exactly large-scale CNVs drive early tumourigenesis in different cell types and whether this knowledge can potentially be exploited for early cancer detection remain to be explored.

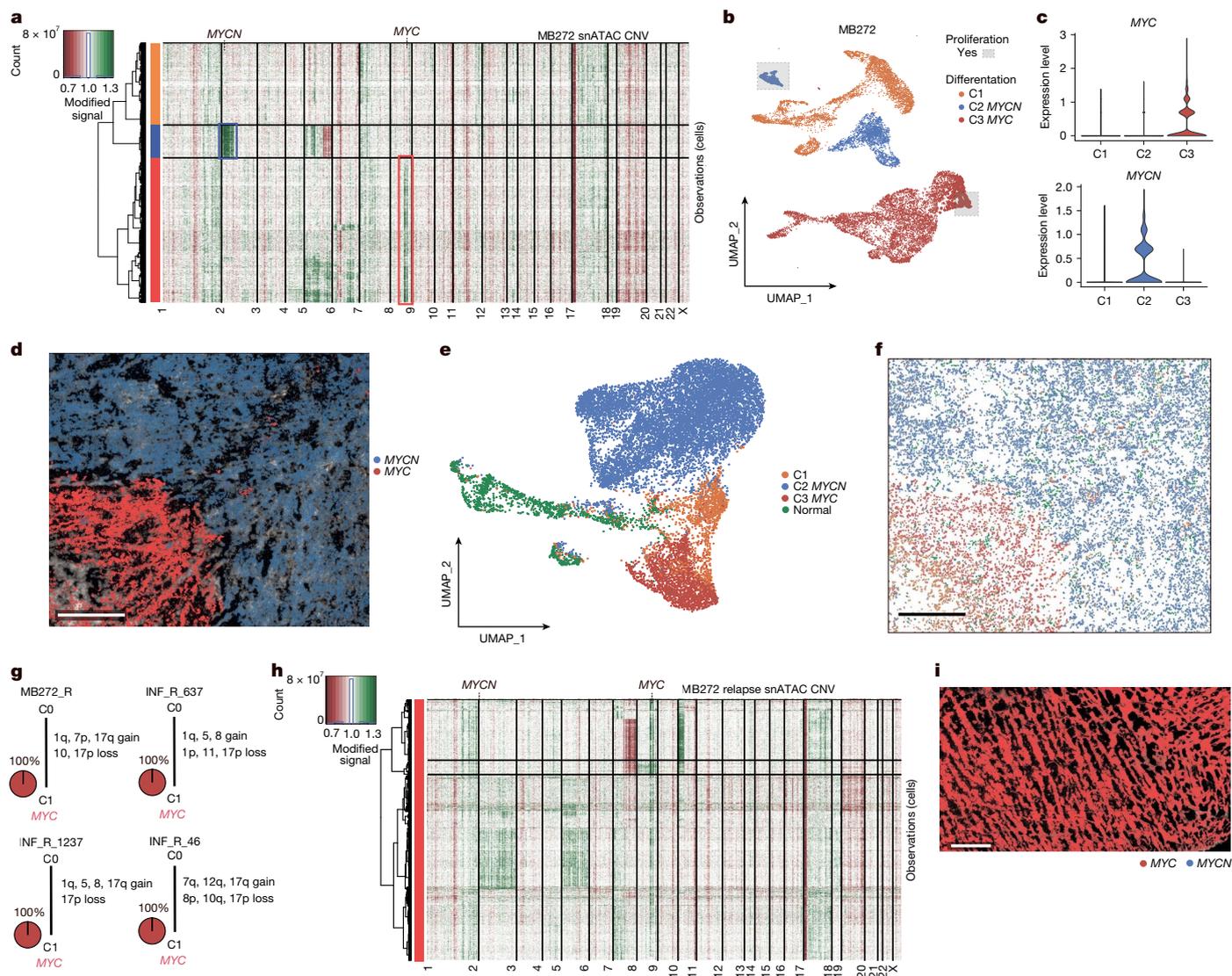


Fig. 5 | Independent oncogene subclones may co-occur in one tumour, but subclones are lost at relapse. a, Copy number profiles of snATAC-seq data from *MYC-MYCN* sample MB272. Red, chromosome loss. Green, chromosome gain. **b**, snRNA-seq UMAP of sample shown in **a**. Grey boxes, proliferating cell clusters with strong proliferation enrichment. Blue, *MYCN*-expressing C2 clone. Red, *MYC*-expressing C3 clone. Orange, C1 clone. **c**, *MYC* and *MYCN* expression in C1, C2 and C3 clones. **d**, Spatial gene expression of *MYC* (red) and

MYCN (blue) from original signals. **e**, Spatial data UMAP of sample shown in **d**. **f**, Spatial visualization of clones of sample in **d**. **g**, Somatic phylogeny trees for *MYC* relapse samples. **h**, Copy number profiles of snATAC-seq data of relapse sample arising from primary sample shown in **a–f**. **i**, Spatial gene expression of *MYC* (red) and *MYCN* (blue) in spatial transcriptomic relapse sample of case shown in **a–f**. Scale bars, 400 μ m (**d**, **f**), 300 μ m (**i**).

Group 3/4 medulloblastoma with *MYC*, *MYCN* or *PRDM6* alterations have complex subclonal structures, with each subclone having unique properties. Therefore, our results strongly argue against the dogmatic ‘cancer stem cell’ hierarchy for group 3/4 medulloblastoma, as these tumours maintain distinct subclones with separate proliferating and differentiating compartments, irrespective of the presence of recognized oncogenes. Although *MYC* or *MYCN* amplifications could arise from the formation of circular extrachromosomal DNA³⁶, our data show that all subclones with *MYC*/*MYCN* amplification demonstrated other large, specific CNV gains and losses within the oncogene-amplified subclone, suggesting these clones are not the founder clone.

Single-cell spatial data from group 3/4 tumours allowed us to inspect the composition of the subclones across tumour tissue fragments; however, owing to the fixed number of genes ($n = 100$) and limited image size (maximum 2 mm), some other spatial properties, such as

formation of blood vessels, were not covered in our study. The application of new spatial techniques that overcome such limitations will be an important research direction in the future.

In addition, undetected *MYC*/*MYCN* subclones challenge the bulk analysis approach in the diagnostic space. These findings, along with support from others³⁷, suggest that single-cell analyses may be an important diagnostic tool in the future, especially for group 4 and subgroup V tumours. In addition, our data challenge the cutoffs used for a tumour to be called *MYC*/*MYCN*-amplified by FISH, as even the smallest *MYC* subclones, which initially have low abundance, have the potential to expand into the dominant clone during relapse. *MYC* amplification may drive disease progression and contribute to therapy resistance and relapse. Such a pattern of *MYC* dominance in subclonal evolution has been observed in other tumours including gliomas³⁸, suggesting that our results may be relevant also to other tumour entities associated with *MYC* oncogenesis.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-025-08973-5>.

1. Bailey, S. et al. Clinical trials in high-risk medulloblastoma: evolution of the SIOP-Europe HR-MB trial. *Cancers* **14**, 374 (2022).
2. Northcott, P. A. et al. The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
3. Cavalli, F. M. et al. Intertumoral heterogeneity within medulloblastoma subgroups. *Cancer Cell* **31**, 737–754.e736 (2017).
4. Hanahan, D. Hallmarks of cancer: new dimensions. *Cancer Discov.* **12**, 31–46 (2022).
5. Hinohara, K. & Polyak, K. Intratumoral heterogeneity: more than just mutations. *Trends Cell Biol.* **29**, 569–579 (2019).
6. Liu, Y., Xiao, B., Li, S. & Liu, J. Risk factors for survival in patients with medulloblastoma: a systematic review and meta-analysis. *Front. Oncol.* **12**, 827054 (2022).
7. Sharma, T. et al. Second-generation molecular subgrouping of medulloblastoma: an international meta-analysis of Group 3 and Group 4 subtypes. *Acta Neuropathol.* **138**, 309–326 (2019).
8. Hovestadt, V. et al. Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* **572**, 74–79 (2019).
9. Riemondy, K. A. et al. Neoplastic and immune single-cell transcriptomics define subgroup-specific intra-tumoral heterogeneity of childhood medulloblastoma. *Neuro. Oncol.* **24**, 273–286 (2022).
10. Roussel, M. F. & Robinson, G. W. Role of MYC in medulloblastoma. *Cold Spring Harb. Perspect. Med.* **3**, a014308 (2013).
11. Shrestha, S., Morcavallo, A., Gorrini, C. & Chesler, L. Biological role of MYCN in medulloblastoma: novel therapeutic opportunities and challenges ahead. *Front. Oncol.* **11**, 694320 (2021).
12. Jones, D. T. et al. Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).
13. Korshunov, A. et al. Gene expression profiling of Group 3 medulloblastomas defines a clinically tractable stratification based on KIRREL2 expression. *Acta Neuropathol.* **144**, 339–352 (2022).
14. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
15. He, S., Liu, Z., Oh, D.-Y. & Thiele, C. J. MYCN and the epigenome. *Front. Oncol.* **3**, 1 (2013).
16. Hendrikse, L. D. et al. Failure of human rhombic lip differentiation underlies medulloblastoma formation. *Nature* **609**, 1021–1028 (2022).
17. Smith, K. S. et al. Unified rhombic lip origins of group 3 and group 4 medulloblastoma. *Nature* **609**, 1012–1020 (2022).
18. Zeller, K. I., Jegga, A. G., Aronow, B. J., O'Donnell, K. A. & Dang, C. V. An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol.* **4**, R69 (2003).
19. Cagan, A. et al. Somatic mutation rates scale with lifespan across mammals. *Nature* **604**, 517–524 (2022).
20. Ganz, J. et al. Contrasting somatic mutation patterns in aging human neurons and oligodendrocytes. *Cell* **187**, 1955–1970.e23 (2024).
21. Mitchell, E. et al. Clonal dynamics of haematopoiesis across the human lifespan. *Nature* **606**, 343–350 (2022).
22. Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).
23. Körber, V. et al. Neuroblastoma arises in early fetal development and its evolutionary duration predicts outcome. *Nat. Genet.* **55**, 619–630 (2023).
24. Okonechnikov, K. et al. Mapping pediatric brain tumors to their origins in the developing cerebellum. *Neuro. Oncol.* **25**, 1895–1909 (2023).
25. Frangioni, J. V. New technologies for human cancer imaging. *J. Clin. Oncol.* **26**, 4012–4021 (2008).
26. Williamson, D. et al. Medulloblastoma group 3 and 4 tumors comprise a clinically and biologically significant expression continuum reflecting human cerebellar development. *Cell Rep.* **40**, 111162 (2022).
27. Sepp, M. et al. Cellular development and evolution of the mammalian cerebellum. *Nature* <https://doi.org/10.1038/s41586-023-06884-x> (2023).
28. Doussouki, M. E., Gajjar, A. & Chamdine, O. Molecular genetics of medulloblastoma in children: diagnostic, therapeutic and prognostic implications. *Future Neurol.* **14**, FN18 (2019).
29. Sondka, Z. et al. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).
30. Bourdeaut, F. et al. MYC and MYCN amplification can be reliably assessed by aCGH in medulloblastoma. *Cancer Genet.* **206**, 124–129 (2013).
31. Rickman, D. S., Schulte, J. H. & Eilers, M. The expanding world of N-MYC-driven tumors. *Cancer Discov.* **8**, 150–163 (2018).
32. Minasi, S. et al. Case report of a pediatric medulloblastoma with concurrent MYC and MYCN subclonal amplification in distinct populations of neoplastic cells. *Virchows Arch.* **485.1**, 153–158 (2023).
33. Mynarek, M. et al. Identification of low and very high-risk patients with non-WNT/non-SHH medulloblastoma by improved clinico-molecular stratification of the HIT2000 and I-HIT-MED cohorts. *Acta Neuropathol.* **145**, 97–112 (2023).
34. Pei, Y. et al. An animal model of MYC-driven medulloblastoma. *Cancer Cell* **21**, 155–167 (2012).
35. Gibson, P. et al. Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468**, 1095–1099 (2010).
36. Chapman, O. S. et al. Circular extrachromosomal DNA promotes tumor heterogeneity in high-risk medulloblastoma. *Nat. Genet.* **55**, 2189–2199 (2023).
37. Danilenko, M. et al. Single-cell DNA sequencing identifies risk-associated clonal complexity and evolutionary trajectories in childhood medulloblastoma development. *Acta Neuropathol.* **144**, 565–578 (2022).
38. Ceresa, D. et al. Early clonal extinction in glioblastoma progression revealed by genetic barcoding. *Cancer Cell* **41**, 1466–1479. e1469 (2023).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

¹Hopp Children's Cancer Center (KiTZ), Heidelberg, Germany. ²National Center for Tumor Diseases (NCT) Heidelberg, a partnership between DKFZ and Heidelberg University Hospital, Heidelberg, Germany. ³Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Heidelberg, Germany. ⁴Developmental Origins of Pediatric Cancer Junior Research Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁵Division of Theoretical Systems Biology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁶MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. ⁷Division of Chromatin Networks, German Cancer Research Center (DKFZ) and Bioquant, Heidelberg, Germany. ⁸Single-cell Open Lab, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁹Division of Pediatric Glioma Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁰Faculty of Biosciences, Heidelberg University, Heidelberg, Germany. ¹¹Center for Molecular Biology of Heidelberg University (ZMBH), DKFZ-ZMBH Alliance, Heidelberg, Germany. ¹²CCU Neuropathology, German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Heidelberg, Germany. ¹³Department of Pediatric Oncology, Hematology & Immunology, Heidelberg University Hospital, Heidelberg, Germany. ¹⁴CCU Pediatric Oncology, German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Heidelberg, Germany. ¹⁵Frankfurt University Hospital, Goethe University, Frankfurt, Germany. ¹⁶Goethe University Frankfurt, University Hospital, Neurological Institute (Edinger Institute), Frankfurt am Main, Germany. ¹⁷German Cancer Consortium (DKTK), Heidelberg, Germany and German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁸Frankfurt Cancer Institute (FCI), Frankfurt am Main, Germany. ¹⁹Goethe University Frankfurt, University Hospital, University Cancer Center (UCT) Frankfurt, Frankfurt am Main, Germany. ²⁰University Children's Hospital, Ulm, Germany. ²¹Department of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ²²Research Institute Children's Cancer Center Hamburg, Hamburg, Germany. ²³Department of Neuropathology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ²⁴Mildred Scheel Cancer Career Center HaTriCS4, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ²⁵Department of Neuropathology, Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany. ²⁶Division of Neuroblastoma Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany. ²⁷These authors contributed equally: Konstantin Okonechnikov, Piyush Joshi, Verena Körber. ²⁸These authors jointly supervised this work: Lena M. Kutscher, Stefan M. Pfister. [✉]e-mail: l.kutscher@kitz-heidelberg.de; s.pfister@kitz-heidelberg.de

Methods

Target cohort selection and verification

Target tumour tissue samples were collected from global medulloblastoma published materials (ICGC², fresh-frozen paraffin-embedded (FFPE)¹³ and Individualized Therapy For Relapsed Malignancies in Childhood (INFORM)³⁹ cohorts). For each selected case, the copy number/structural variant profiles from methylation and/or WGS data were used to identify *MYC/MYCN* amplification and *SNCAIP* structural variant presence. Bulk gene expression RNA-seq profiles from these samples were used to inspect *MYC/MYCN/SNCAIP/PRDM6* expression as well. For some cases with sufficiently available FFPE material, further FISH experiments were performed to verify the selection (details in Supplementary Table 1). No statistical methods were used to predetermine sample size.

Single-nucleus multi-omics sequencing

Flash-frozen tumour samples were processed to extract nuclei as described earlier²⁷. Extracted nuclei were processed using Chromium Single Cell Multiome ATAC Gene expression kit and Chromium Controller instrument (10x Genomics) as per manufacturer's recommendations. One sample, MB248, was processed with Chromium Next GEM Single Cell 3' reagent kit as per the manufacturer's recommendation. In total, 15,000–20,000 nuclei were loaded per channel along with the multiome gel bead. Libraries were quantified using Qubit Fluorometer (Thermo Fisher Scientific) and profiled using Fragment Analyzer. snRNA-seq and snATAC-seq libraries were sequenced using a Next-Seq2000 to the recommended lengths. If the snATAC-seq library was not of good quality, we still used the obtained snRNA-seq library if that was found to be appropriate on the basis of quality control parameters. snRNA-seq and snATAC-seq datasets were further analysed separately.

snRNA-seq data analysis

De-multiplexed reads were aligned to human genome assembly GRCh38 (v. p13, release 37, gencodegenes.org). Comprehensive gene annotation (PRI) was customized by filtering to transcripts with the following biotype: protein coding, lncRNA, lIG and TR gene and pseudogene as recommended for cellranger mkgtf wrapper. Reads were aligned using STARsolo with parameters: `--soloType CB_UMI_Simple --soloFeatures GeneGeneFull --soloUMIfiltering MultiGeneUMI --soloCBmatchWLtype IMM_multi_pseudocounts --soloCellFilter None --outSAMmultNmax 1 --limitSjdbInsertNsj 1500000`. For overlapping genes for which intronic alignment recovered low counts, exonic alignment counts were used. Cells were separated from debris using the diem pipeline⁴⁰. Cells with mitochondria fraction above 2%, number of detected genes above 6,600 and an intronic fraction (number of reads aligned to intron/total number of reads aligned to exon + intron) less than 25% were also filtered out. Filtered cells were corrected for background signature using the SoupX pipeline⁴¹. Finally, the scrublet⁴² tool was used to remove putative doublets. Further, the gene expression matrices from all samples were merged together in the full matrix and processed by means of the Seurat package⁴³ to normalize, compute top principal components ($n = 30$), find most highly variable genes ($n = 2,500$) and visualize by means of UMAP. After distinguishing non-tumour cells on the basis of corresponding markers and combined UMAPs, per sample processing was performed using the Seurat toolkit using the same settings combined with cell clustering. The enrichment of proliferation, differentiation and progenitor-like activity of medulloblastoma-specific markers per cell was performed using the single sample function from the GSVA R package⁴⁴ using two independent reference datasets^{8,9}. Cell clusters enriched with proliferation signals were selected and marked on the basis of maximum GSVA signal enrichment from manual inspection per sample.

snATAC-seq data analysis

ATAC-seq reads were aligned to GRCh38 using the Cellranger arc wrapper. The selected cells were processed using the Signac R package⁴⁵

to filter the doublets/outliers on the basis of signal per cell distribution analysis and to inspect the cell compartments by means of UMAP visualization after normalization and identification of the most highly variable regions. snRNA-seq data information was used to annotate the cells from corresponding processed data.

To identify differentially enriched *cis*-regulatory elements per sample per annotation, peaks were first called on the merged snATAC-seq data using the ArchR R package⁴⁶ by first creating pseudobulk replicates using `addGroupCoverages` (`minCells=2000`, `maxCells=5000`, `minReplicates=2`, `maxReplicates=5`, `groupBy="Sample"`, `maxFragments=100 * 10^6`) and calling reproducible peaks using `addReproduciblePeakSet` (`groupBy="Sample"`, `maxPeaks=150000`). Obtained peaks were further filtered on the basis of presence in at least 5% of cells in any sample. Marker peaks were then obtained using `getMarkerFeatures()` on the basis of subclonal annotation and filtering on the basis of area under the curve > 0.52 and false discovery rate (FDR) < 0.01 .

Genes correlated to the peaks were identified using `addPeak2GeneLinks()` using the subset of cells with dual snRNA-seq and snATAC-seq profiles in the merged data. Correlated genes (more than 0.1) were intersected with peaks associated with genes identified on the basis of GREAT⁴⁷ annotation to identify the robust pairs of peaks to gene links.

Single-cell CNV phylogeny reconstruction

Initially, CNV analysis was performed on a subset of samples ($n = 2$ for each oncogene) using the InferCNV tool¹⁴ on the raw gene expression matrix with droplet protocol adjusted parameters (average read counts cutoff 0.5, smooth method runmeans, denoise active) and hierarchical clustering by means of the ward.D2 method to derive the clonal phylogeny. The main subclones obtained from this phylogeny demonstrated a close match to the initial Seurat clustering results (mean purity evaluation metrics across samples: 0.921). To further improve the visualization and increase computational efficiency, the CNV analysis was performed on the full cohort by transferring single cells into meta-cells, on the basis of the established method⁴⁸. For this purpose, we computed the sum of gene expression counts across $n = 5$ cells combined within the clusters derived from Seurat processing. The meta-cell InferCNV calling was performed for each sample separately with read counts cutoff 0.5, and the phylogeny clustering results were visualized in UMAPs with k as number of clusters, varying from 2 to 5. Differentially expressed genes for identified subclones per sample were computed by means of a Wilcoxon rank sum test. Significant *MYC/MYCN/PRDM6* differential expression and progenitor-like activity enrichment values were computed per cell to finalize the derived phylogeny cut limit for each case after manual inspection. The selected cut limit for the number of subclones in the phylogeny was verified by using random tree subcluster partition in the phylogeny reconstruction with a minimum $P < 0.05$.

All tumour samples ($n = 16$) were also merged together and CNV profiling was performed to verify the status of non-tumour cells. The annotation of verified normal cells was further used as reference control for each sample.

For snATAC-seq data, a matrix with all genomic regions as raw and read counts per column per sample was used to adjust for InferCNV input format. Further meta-cell formation and the same CNV calling procedure as for snRNA-seq were performed on the derived matrices. The subclone annotation derived from snRNA-seq data was used to assign corresponding cluster phylogeny per sample.

Interactive CNV visualization for the results from snRNA-seq and snATAC-seq data per sample is available through ShinyApp: <http://kokonech.shinyapps.io/mbOncoAberrations>.

Mutation calling from snATAC-seq data

For $n = 3$ *MYC* cases the number of reads in snATAC-seq data was increased up to 3×10^9 per sample to have sufficient coverage for

mutation calling. The alignment of novel reads was performed using the same strategy as described above. Initial subclone annotation was used to classify the cells. Afterwards, the mutation calling was performed using the SComatic method⁴⁹ on the BAM files extracted for each subclone. Initial mutation filtering was performed on the basis of inspection of somatic mutations derived from WGS data² as the main control. Results were also confirmed using the bcftools method⁵⁰ on the full merged pseudobulk snATAC-seq data with standard settings. The VAF from the full merged pseudobulk mutation calling was used to compare with the VAF from bulk control results. The visualization of somatic mutation presence across subclones per sample was performed by means of the R package ComplexHeatmap. For more relaxed filtering control, blood bulk WGS control was applied to exclude non-somatic mutations, while mutation filtering limits were strengthened: min coverage 20× and minimum support 5×. Annotation of identified mutations was performed using Annovar toolkit⁵¹ using gencode v.38 materials.

Single-cell DNA and RNA sequencing procedure

Flash-frozen tumour samples were processed to extract nuclei as described previously⁵². Single nuclei were sorted into the wells of DNA LoBind 96-well plates using fluorescence activated cell sorting. Each plate was used for downstream genome and transcriptome isolation, reverse transcription of the transcriptomes and primary template-directed amplification using the ResolveOME Multiomic kit from Bioskryb, according to the manufacturer's protocol. Selected wells were used for quality control before library preparation to assess the quantity and quality of the extracted genome or transcriptome DNA, using the 4200 TapeStation System. Sequencing libraries were prepared using Illumina-compatible unique dual index adaptors, according to the manufacturer's protocol as provided in the ResolveOME kit. After amplification, barcoded libraries were pooled and purified using Ampure Beads to select 250–1,000-base pair (bp) fragments with an average peak size of 400–500 bp. Multiplex libraries were sequenced using one lane of the NovaSeq 6000 System each, with 100-bp paired-end sequencing for the genome libraries and 100-bp single-read sequencing for the transcriptome libraries.

Single-cell DNA and RNA sequencing data analysis

DNA sequencing reads per cell were initially processed (quality control, alignment) by means of the BJ-DNA-QC Nextflow-based pipeline from BioSkryb using hg38 as the main reference. CNV profiling was performed with the Ginkgo tool⁵³ on the basis of resolution 100 kilobases (kb). Outlier signal adjustment was performed on the basis of cut for mean plus 2 s.d. Heatmap visualization was performed by means of the ComplexHeatmap R package using the clustering method ward.D2 with Euclidean distance.

RNA-seq reads per cell were aligned to the hg38 reference by means of the STAR tool⁵⁴. Gene expression counts were computed by means of the FeatureCounts option from the Subread toolkit⁵⁵ using gencode v.38 reference. Combined gene expression matrix analysis was performed by means of the Seurat toolkit⁴³. Non-tumour cells as well as subclone specificity were identified on the basis of the inspection of known gene markers and projection into snRNA-seq profiles using the transfer Seurat R toolkit function.

Molecular cartography

The specific gene set ($n = 100$) covering group 3/4 known driver genes alongside marker genes of the developing cerebellum non-malignant cell types was selected for the protocol (Supplementary Table 7). The gene selection was on the basis of specific properties. First, the main selection was split into two blocks: tumour-specific genes (60%) and normal cell markers (40%). The tumour-specific genes had several blocks of selection: (1) known target markers of group 3/4 tumours

including *MYC*, *MYCN*, *SNCAIP* and *PRDM6*; (2) proliferation, differentiation and cell-cycle activity markers; (3) cells-of-origin-associated markers. The normal cell type markers were selected on the basis of knowledge about normal cell types of the cerebellum that could also be present in the tumour. For each cell type, at least two markers were selected. All selected genes were also verified on available snRNA-seq data as well as bulk profiles.

Optimal cutting temperature compound (OCT)-embedded samples were cryo-sectioned into 10- μ m sections onto a molecular cartography slide. Fixation, permeabilization, hybridization and automated fluorescence microscopy imaging were performed according to the manufacturer's protocol (Molecular preparation of human brain (beta), Molecular colouring, workflow setup) as described previously⁵².

Spatial data analysis

The detection of cell boundaries was performed with CellPose⁵⁶. Afterwards, gene expression counts were computed per cell and extracted using further custom Python scripts. Initial cell filtering was performed by assigning the minimum number of counts/genes per cell and size of the cells. Afterwards, the analysis of the formed gene expression matrix, including clustering and UMAP visualization, was executed using the Seurat toolkit⁴³. Annotation of cell states and types was achieved through direct projection with the snRNA-seq data by means of transfer function and verified by visual inspection of marker genes. Spatial-specific analysis, the detection of closest cell connections, was conducted using the Giotto toolkit⁵⁷.

Deconvolution analysis of *MYC*/*MYCN* cases

The assigned *MYC*/*MYCN* single-cell dataset (MB272) with annotation of compartments was used as reference control for the CIBERSORT method⁵⁸ to perform deconvolution on a set of bulk FFPE medulloblastoma RNA-seq profiles from *MYC*/*MYCN* samples¹³. For each case, *MYC*/*MYCN* status was derived from methylation copy number profile. The deconvolution values obtained from CIBERSORT results were visualized by using the ComplexHeatmap R package in order to demonstrate compartment enrichments per sample.

Survival analyses on the basis of expression of *MYC* as well as the computed deconvolution of *MYC* compartment proportion of multiple genes were performed using the Kaplan–Meyer algorithm with applied Bonferroni correction for multiple testing. The resulting plots were generated by means of the R2 Genomics Analysis and Visualization Platform (<http://r2.amc.nl>).

Bulk RNA-seq data analysis

Bulk RNA-seq data from the ICGC cohort overlapping with WGS data were available for 85 group 3/4 medulloblastomas of our cohort². We performed GSEA on reads per kilobase of transcript per million mapped read scores using the R package GSEABase. We used the following gene sets: *MYC* target genes: 'HALLMARK_MYC_TARGETS_V2', 'MYC_UP.V1_UP', 'DANG_MYC_TARGETS_UP'²⁶; S-phase genes: 'REACTOME_S_PHASE', 'SA_G1_AND_S_PHASES'.

Extended gene expression and associated CNV profile metadata for 405 group 3/4 medulloblastoma samples were integrated for inspection of common CNV gains/losses³. The chromosomal arm gains or losses as provided in this corresponding study were binarized. DESeq2 was used to compute differentially expressed genes using lfcShrink($\text{type} = \text{"apeglm"}$) and filtered using adjusted $P < 0.001$ and log fold change (lfc) > 0.5 criteria to obtain a list of statistically significant altered gene expression. We then used msigdb and clusterProfiler R packages to identify chromosomal loci of the differentially expressed genes.

Additionally evident differentially expressed genes (adjusted $P < 0.05$) among the central nervous system tumours provided in the corresponding study²⁴ were inspected to identify those that are specific

Article

for common group 3/4 gain (overexpressed, $lfc > 0.5$) and loss (low expressed, $lfc < 0.5$) regions.

WGS data

Mutation calls (SNVs, indels, CNVs and structural variants (SVs)) of previously published WGS data from medulloblastomas of all subtypes were taken from the ICGC dataset². Only samples from primary tumours with clear subtype annotation and clear ploidy status were included; see Supplementary Table 5 for an overview on these samples and associated clinical data.

Driver mutations (SNVs and CNVs)

Non-synonymous SNVs, small indels and small structural rearrangements (amplifications, defined as copy number gains ≥ 10 , homozygous deletions with less than 0.9 copy numbers and translocations with a minimal event score of 5) were classified as driver mutations if they targeted a splice-site or an exonic region of *PRDM6*, *MYC* or a gene listed as a putative driver of medulloblastoma in the cancer driver database intogen⁵⁹ (release date 31 May 2023). Moreover, we included *TERT* promoter mutations at hg19 positions 1295228 and 1295250 as drivers. High-level amplifications affecting *MYC* or *MYCN* (identified from methylation/WGS copy number profiles) and duplications of *SNCAIP*, leading to overexpression of *PRDM6* (identified from WGS SV calling), were additionally integrated from a previous global data analysis².

Large-scale CNVs were defined as CNVs spanning at least 1 Mb and with a coverage ratio less than 0.9 or a coverage ratio greater than 1.1, according to the output by ACESeq. Retained CNVs with a size of at least 25% the size of the p arm of a respective chromosome were further classified as affecting both arms if the CNV spanned the centromere, or else as affecting the p arm or the q arm. Among these CNVs, we tested for positive enrichment of particular chromosomes in the cohort using a binomial test with success probability 1/24 (that is, assuming that each chromosome has equal probability to be affected by the CNV). Chromosomes with an adjusted $P < 0.05$ (Holm's correction) were classified as likely drivers of medulloblastoma. This analysis was separately performed for gained and lost chromosomes. Among group 3/4 medulloblastomas, we identified gains of chromosomes 4, 7/7q, 12 and 17/17q and losses of chromosomes 8, 10/10q, 11 and 17p as significant. We augmented this list by gains of chromosomes 18 and 1q and loss of 5q, as was reported previously⁶⁰.

Timing of CNVs, ECA and MRCA

Quantification of mutation densities at copy number gains was performed using the R package NBevolution v.0.0.0.9000, which is described in detail in the corresponding study²³. In brief, we counted clonal mutations separately on each autosome, stratified by copy number state using the function `count.clonal.mutations()` with `max.CN=4`, excluding chromosomal segments with length less than 10^7 bp. `count.clonal.mutations()` fits a binomial mixture model with success probabilities according to the expected mean values of the clonal VAF peaks, which, for an impure sample with tumour cell content ρ , are given by

$$\text{VAFs} \in \left\{ \frac{\rho}{\zeta}, \frac{(\text{CN} - b)\rho}{\zeta}, \frac{b\rho}{\zeta} \right\}, \quad (1)$$

where CN denotes the copy number of a given segment, b denotes the copy number of the minor allele (that is, the allele with the smallest number of copies) on this segment and

$$\zeta = \rho \text{CN} + 2(1 - \rho) \quad (2)$$

is the average copy number of a given locus in the sample. Mutation densities (SSNVs per bp) at MRCA and ECA, denoted by \bar{m}_{MRCA} and \bar{m}_{ECA}

respectively, were computed using the function `MRCA.ECA.quantification()`. In brief, `MRCA.ECA.quantification()` first estimates \bar{m}_{MRCA} from the number of all clonal mutations and the total size of the analysed genome, $g = \sum_l g_l$, where the index l labels individual segments contributing to the analysis, yielding

$$\bar{m}_{\text{MRCA}} = \sum_l \frac{n_{k,l} + n_{\text{CN}-b,l}(\text{CN}_l - b_l) + n_{b,l}b_l}{g \text{CN}_l}, \quad (3)$$

where $n_{k,l}$ denotes the number of clonal mutations present on k copies of the l -th segment. Note that \bar{m}_{MRCA} normalizes mutation densities per copy and hence can be interpreted as molecular time. The function `MRCA.ECA.quantification()` also computes lower and upper 95% confidence bounds for \bar{m}_{MRCA} by bootstrapping the genomic segments 1,000 times. In the next step, `MRCA.ECA.quantification()` asks for evidence for an earlier common ancestor in the data. If a chromosomal gain occurs concomitantly with the onset of tumour growth, the density of amplified clonal mutations (that is, present on multiple copies of a gained allele) will, on average, be equal to the density of clonal mutations on non-amplified chromosomes or chromosomal regions. By contrast, if a chromosomal gain occurs earlier, the density of amplified clonal mutations on the gained allele will be smaller than the density of clonal mutations on non-amplified chromosomes or chromosomal regions. To distinguish these two cases, `MRCA.ECA.quantification()` tests for each gained segment whether or not the density of amplified clonal mutations agrees with the mutation density at MRCA, on the basis of a negative binomial distribution. If the mutation density of amplified clonal mutations is significantly smaller than the mutation density at MRCA, the segment is assigned to an earlier time point or else to the MRCA. `MRCA.ECA.quantification()` then asks whether segments assigned to earlier time points emerged in the same time window or during different time windows. We define as the null hypothesis that all CNVs emerged in an ECA. To test the null hypothesis, `MRCA.ECA.quantification()` computes the mutation densities at the ECA from the number of clonal mutations on the amplified minor allele, $n_{b,l}$ (if $b > 1$), and from the number of mutations on the amplified major allele, $n_{\text{CN}-b,l}$ as

$$\bar{m}_{\text{ECA}} = \frac{\sum_{l, P_{\text{adj},l} \leq 0.01} n_{b,l} + n_{\text{CN}-b,l}}{\sum_{l, P_{\text{adj},l} \leq 0.01} g_{l,b} + g_{l,\text{CN}-b}}, \quad (4)$$

where $P_{\text{adj},l}$ is the adjusted P value of segment l belonging to the MRCA, and $g_{l,b}$ and $g_{l,\text{CN}-b}$ are the length of segment l , contributed by the amplified minor and major alleles, respectively. In analogy to \bar{m}_{MRCA} , `MRCA.ECA.quantification()` also estimates lower and upper 95% confidence bounds by bootstrapping. On the basis of a negative binomial distribution (and an FDR of 0.01), `MRCA.ECA.quantification()` then tests for each early segment whether or not its mutation density indeed conforms to a joined ECA, as defined by \bar{m}_{ECA} . Only segments with mutation densities conforming to \bar{m}_{ECA} are assigned to the ECA, whereas all other segments are reported as conforming neither to the ECA nor to the MRCA.

Upon timing MRCA and ECA for each sample, we translated mutation densities into weeks post conception (p.c.) by inferring SSNV rates per diploid genome and embryonic day ($\mu\lambda$), using the measured VAF distributions and age at diagnosis as outlined below in section 'Real-time estimate of cell division rate'. As mutation calling was performed by comparing tumours against a matched blood control, mutation densities correlate with the time after gastrulation (at approximately 2 weeks after conception). Thus, the mutation density per haploid genome (3.3×10^9 bp), \bar{m} , relates to the time p.c. according to $\bar{m}(t) = \frac{\mu\lambda}{\text{day } 3.3 \times 10^9} \frac{1}{t - 14}$ d). The estimated time of birth was taken as 38 weeks after gastrulation (40 weeks p.c.).

Timing of SNVs and small indels

We classified SNVs and small indels as subclonal or clonal on the basis of the number of variant reads, n_{var} , the number of reference reads, n_{ref} , tumour purity ρ and copy number k . Specifically, mutations were classified as subclonal if the probability to sample at most n_{var} variant reads out of $n_{\text{var}} + n_{\text{ref}}$ total reads according to a binomial distribution with success probability $\frac{\rho}{\rho k + 2(1-\rho)}$ was smaller than 5%. If a mutation was classified as clonal and fell on a region with $k=3$, we moreover classified the mutation as early clonal (that is, acquired before the chromosomal gain on the gained chromosome and hence present on at least two copies) if the probability to sample at most n_{var} variant reads out of $n_{\text{var}} + n_{\text{ref}}$ total reads was at least 5% according to a binomial distribution with success probability $\frac{2\rho}{\rho k + 2(1-\rho)}$, or else as late clonal.

Modelling medulloblastoma initiation

We modelled medulloblastoma initiation and growth with a population genetics model originally developed for neuroblastoma, as described previously²³. In brief, the model assumes that disease initiation is driven by two consecutive drivers in a transiently expanding tissue of origin, which for group 3/4 medulloblastoma is probably the population of unipolar brush cell progenitors (UBCPs^{16,17,24}). The two driver events are associated with the ECA and the MRCA of the tumour, and spawn, respectively, a pre-malignant and the malignant tumour clone. We assumed that both drivers occur with small probabilities μ_1 and μ_2 during cell divisions, and confer a selective advantage (r and s , respectively) that acts by reducing cell differentiation. Moreover, we assumed that UBPCs acquire on average μ neutral somatic variants per cell division, which we modelled with a Poisson process. The population of UBPCs has been experimentally described from week 9 p.c. until the time of birth^{24,27}. To capture this trend, we modelled an initial phase of exponential growth at rate $\lambda_1 - \delta_1$, $\lambda_1 > \delta_1$ until time T , where λ_1 and δ_1 denote the division and differentiation rate, respectively, and a subsequent phase of exponential decline at rate $\lambda_2 - \delta_2$, $\lambda_2 < \delta_2$.

Following refs. 23,61, we calculated the probability of the MRCA to occur at time t according to

$$P_{\text{MRCA}} = \begin{cases} P_{\text{MRCA,I}}, & t < T \\ 1 - (1 - P_{\text{MRCA,I}})(1 - P_{\text{MRCA,II}})(1 - P_{\text{MRCA,III}}), & t \geq T' \end{cases} \quad (5)$$

with⁴⁰

$$P_{\text{MRCA,I}}(t) = \sum_{x=1}^{N(t)-1} e^{-\mu_1(x-1)}(1 - e^{-\mu_1}) \left(1 - \exp \left\{ -\frac{\mu_1 \mu_2 \lambda_1 T N(t) F}{1 - \frac{\delta_1}{\lambda_1}} \right\} \right),$$

$$P_{\text{MRCA,II}}(t) = 1 - \exp \left(-\frac{\mu_1 \mu_2 \lambda_1 \lambda_2 v_{2,D} T}{\lambda_2 - \delta_2 / r} N(T) \{ e^{(\lambda_2 - \delta_2 / r)(t-T)} - 1 \} \right),$$

$P_{\text{MRCA,III}}(t)$

$$= 1 - \exp \left(-\frac{\mu_1 \mu_2 \lambda_2^2 v_{2,D} N(T)}{\delta_2 \left(\frac{1}{r} - 1 \right)} \left\{ \frac{e^{(\lambda_2 - \delta_2)(t-T)} - 1}{\lambda_2 - \delta_2} - \frac{e^{(\lambda_2 - \frac{\delta_2}{r})(t-T)} - 1}{\lambda_2 - \frac{\delta_2}{r}} \right\} \right),$$

where $v_{2,D} = 1 - \frac{\delta_2}{s\lambda_2}$ is the survival probability of a cell undergoing the second oncogenic event while the population decays, $F = \int_0^1 v_{2,E} / (v_{2,E} z^\alpha) dz$, $v_{2,E} = 1 - \frac{\delta_1}{s\lambda_1}$ and $\alpha = \frac{\delta_1 - s\lambda_1}{\delta_1 - \lambda_1}$. Moreover, we calculated the probability of the ECA to occur at t_1 , conditioned on the MRCA occurring at t_2 , as described previously²³

$$P(t_1|t_2) = \frac{t_1}{t_2}; t_1 < t_2 \leq T,$$

and

$$P(t_1|t_2) = \frac{\Theta(T - t_1) \lambda_1 \delta_2 \left(1 - \frac{1}{r} \right) t_1 + \Theta(t_1 - T) [\lambda_2 + \lambda_1 T \delta_2 (1 - 1/r) - \lambda_2 e^{\delta_2(1-1/r)(T-t_1)}]}{\lambda_2 + \lambda_1 T \delta_2 (1 - 1/r) - \lambda_2 e^{\delta_2(1-1/r)(T-t_2)}}; \quad (6)$$

$$t_2 > T,$$

where $\Theta(\cdot)$ is the Heavyside step function and $0 \leq t_1 < t_2$.

To estimate the model parameters from the WGS data, we contrasted the probability of acquiring the first and second drivers with the measured distribution of SNV densities at ECA and MRCA in group 3/4 medulloblastomas using approximate Bayesian computation with sequential Monte-Carlo sampling (ABC-SMC), as implemented in pyABC⁶². We used a population size of 1,000 parameter sets and 25 SMC generations or $\varepsilon \leq 0.05$ as termination criteria. The model fit was performed in analogy to Körber et al.²³ (code and pseudo-code are available at <https://github.com/kokonech/mbOncoABerrations>). 95% posterior-probability bounds for the model fits were estimated by simulating the model with each sampled parameter set and cutting off 2.5% at each end of the simulated distribution.

Modelling medulloblastoma growth

We modelled medulloblastoma growth from the MRCA as exponential growth with rate $\lambda_T - \delta_T$, where λ_T denotes the division rate and δ_T the loss rate (owing to differentiation or death) in the tumour. Denoting the number of tumour cells with $N_T(t)$ and assuming that neutral mutations are on average acquired at a rate $\mu \lambda_T N_T(t)$ per haploid genome during tumour growth, the site frequency spectrum of neutral variants at t_{end} is, on average⁶³,

$$S_k(i, \mu) = \int_0^{t_{\text{end}}} P_{1,i}(\lambda_T, \delta_T, t_{\text{end}} - t) \mu k \lambda_T N_T(t) dt, \quad (7)$$

where k is the chromosomal copy number, and $P_{1,i}(\lambda_T, \delta_T, t_{\text{end}} - t)$ is the probability to grow from a single cell to a clone of size i within a time span $t_{\text{end}} - t$, according to a supercritical linear birth-death process (for example, see ref. 64).

To estimate μ and δ_T/λ_T from the WGS data, we followed the strategy described previously²³ to compare $M_k(a, \mu)$, the cumulative allele frequency histogram of variants present in at least a cells in regions with copy number k , given a mutation rate μ , between model and data. Here,

$$M_k(a, \mu) = \sum_{i=a}^b S_k(i, \mu) \approx \int_0^{t_{\text{end}}} \mu k \lambda_T N_T(t) \frac{P_{1,b}(\lambda_T, \delta_T, t_{\text{end}} - t) - P_{1,a}(\lambda_T, \delta_T, t_{\text{end}} - t)}{\log \beta(t_{\text{end}} - t)} dt, \quad (8)$$

$$\beta(t) = \frac{\lambda_T (e^{(\lambda_T - \delta_T)t} - 1)}{\lambda_T e^{(\lambda_T - \delta_T)t} - \delta_T}.$$

To this end, we used ABC-SMC⁶² with a population size of 1,000 parameter sets and 25 generations or $\varepsilon \leq 0.05$ as termination criteria. To learn the dynamics of tumour growth with confidence, we included tumours with well-defined subclonal tails and no evidence for subclonal selection. Tumours were selected (Supplementary Table 5) on the basis of visual inspection of the VAF histograms, to remove cases with poor subclonal resolution. In addition, we removed cases without age information and those with evidence for subclonal selection as suggested by the evolutionary model implemented in Mobster⁶⁵ (setting auto-setup = 'FAST'), which we ran on autosomes and upon computing pseudo-heterozygous VAFs, $\sqrt{\text{VAF}}$, defined as 50% of the mutant sample fraction, SF (hence, $\sqrt{\text{VAF}} = \frac{c}{2k} \text{VAF}$, where k is the number of alleles carrying the mutation and VAF is the actual VAF). For the 35 retained group 3/4 medulloblastomas, we followed the strategy outlined by

Article

Körber et al.²³ to estimate the model parameters from the measured VAF distribution (code and pseudo-code are available at <https://github.com/kokonech/mBOncoAberrations>).

Real-time estimate of cell division rate

From the model fits of medulloblastoma initiation and growth to WGS data, we estimated differentiation/loss rates and mutation rates relative to the rate of cell divisions. To convert these estimates to real-time, we used the age distribution at diagnosis for calibration. In a first step, we estimated the cell division rate of UBCPs, λ , from the number of generations between gastrulation and MRCA plus the number of generations between MRCA and diagnosis (t_D), which can be inferred from the mutational burden in the tumour²³. Recall the selective advantage of the growing tumour, s . With $\lambda_T = s\lambda$, this yields

$$\lambda = \frac{1}{t_D} \left(\frac{\bar{m}_{MRCA}}{\mu} + \frac{\log N_T(t_D)}{1 - \frac{\delta_T}{s\lambda}} \right) = \frac{1}{t_D} \left(\frac{\bar{m}_{MRCA}}{\mu} + \frac{\log N_T(t_D)}{\mu} \mu_{\text{eff}} \right), \quad (9)$$

where we used the estimate for μ from the parameter inference for medulloblastoma initiation and the estimate for the effective mutation rate, $\mu_{\text{eff}} = \frac{\mu}{1 - \frac{\delta_T}{s\lambda}}$, from the parameter inference for medulloblastoma growth. Assuming a tumour mass in the order of a few cubic centimetres and hence $N_T(t_D) = 10^9$ cells, and defining t_D as the age at diagnosis, A , plus, on average, 250 d of embryogenesis after gastrulation, we obtained for each tumour (labelled with index i) an estimate for the division rate with mean,

$\langle \lambda_i \rangle = \frac{1}{2\langle \mu \rangle (A_i + 250 \text{ d})} (\langle 2\bar{m}_{MRCA,i} \rangle + \log 10^9 \langle \mu_{\text{eff},i} \rangle)$, and standard deviation,

$\sigma(\lambda_i) = \frac{1}{2\langle \mu \rangle (A_i + 250 \text{ d})} \left(\frac{2\langle \bar{m}_{MRCA,i} \rangle + \log 10^9 \langle \mu_{\text{eff},i} \rangle}{2\langle \mu \rangle} \sigma(2\mu) + \sigma(2\bar{m}_{MRCA,i}) + \sigma(\mu_{\text{eff},i}) \right)$, in

actual time (where the factor 2 accounts for the fact that $\langle \mu \rangle$ and $\bar{m}_{MRCA,i}$ measure the mutation rate and the mutation density, respectively, per haploid genome).

Finally, we computed the mutation rate per day during tumour initiation, by computing $\mu \lambda_{T,i}$ with associated uncertainty $\mu \Delta \lambda_{T,i} + \lambda_{T,i} \sigma(\mu)$, which relates the molecular clock to real-time. For this purpose, we averaged across the inferences from all tumours that went into the analysis.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The DNA whole-genome sequencing mutation results were integrated from the corresponding medulloblastoma molecular landscape study² deposited at the European Genome-Phenome Archive under accession number EGAS00001001953. Single-nucleus RNA and ATAC data are available at the GEO database under accession numbers GSE253557 and GSE253573. All raw images and processed data after cell segmentation from spatial transcriptomics experiments are available at the GEO database and can be accessed under accession number GSE252090. Variant calls from WGS data are available from Mendeley Data (<https://doi.org/10.17632/g4r22w9pp8.1>). scRNA/ATAC and WGS data analysis results can be visually inspected on the interactive web application <http://kokonech.shinyapps.io/mBOncoAberrations>.

Code availability

All custom Python and R scripts as well as details about the external software environment applied during the data analysis are available

at GitHub (<https://github.com/kokonech/mBOncoAberrations>) and at Zenodo (<https://doi.org/10.5281/zenodo.15083518>)⁶⁶.

- Van Tilburg, C. M. et al. The pediatric precision oncology INFORM registry: clinical outcome and benefit for patients with very high-evidence targets. *Cancer Discov.* **11**, 2764–2779 (2021).
- Alvarez, M. et al. Enhancing droplet-based single-nucleus RNA-seq resolution using the semi-supervised machine learning classifier DIEM. *Sci. Rep.* **10**, 11019 (2020).
- Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, gaa151 (2020).
- Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291. e289 (2019).
- Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902. e1821 (2019).
- Hänzelmann, S., Castelo, R. & Guinney, J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
- Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
- Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
- Tanigawa, Y., Dyer, E. S. & Bejerano, G. WhichTF is functionally important in your open chromatin data? *PLoS Comput. Biol.* **18**, e1010378 (2022).
- Blanco-Carmona, E. et al. Tumor heterogeneity and tumor-microglia interactions in primary and recurrent IDH1-mutant gliomas. *Cell Rep. Med.* **4**, 101249 (2023).
- Muyas, F. et al. De novo detection of somatic mutations in high-throughput single-cell profiling data sets. *Nat. Biotechnol.* **42**, 758–767 (2024).
- Danecek, P. & McCarthy, S. A. BCFTools/csq: haplotype-aware variant consequences. *Bioinformatics* **33**, 2037–2039 (2017).
- Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Res.* **38**, e164 (2021).
- Ghasemi, D. R. et al. Compartments in medulloblastoma with extensive nodularity are connected through differentiation along the granular precursor lineage. *Nat. Commun.* **15**, 269 (2024).
- Garvin, T. et al. Interactive analysis and assessment of single-cell copy-number variations. *Nat. Methods* **12**, 1058–1060 (2015).
- Dobin, A. & Gingeras, T. R. in *Data Mining Techniques for the Life Sciences* (eds Carugo, O. & Eisenhaber, F.) 245–262 (Springer, 2016).
- Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
- Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
- Dries, R. et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 78 (2021).
- Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* **1711**, 243–259 (2018).
- Martínez-Jiménez, F. et al. A compendium of mutational cancer driver genes. *Nat. Rev. Cancer* **20**, 555–572 (2020).
- Juraschka, J. & Taylor, M. D. Medulloblastoma in the age of molecular subgroups: a review. *J. Neurosurg. Pediatr.* **24**, 353–363 (2019).
- Haeno, H., Iwasa, Y. & Michor, F. The evolution of two mutations during clonal expansion. *Genetics* **177**, 2209–2221 (2007).
- Klinger, E., Rickert, D. & Hasenauer, J. pyABC: distributed, likelihood-free inference. *Bioinformatics* **34**, 3591–3593 (2018).
- Ohtsuki, H. & Innan, H. Forward and backward evolutionary processes and allele frequency spectrum in a cancer cell population. *Theor. Popul. Biol.* **117**, 43–50 (2017).
- Bailey, N. T. *The Elements of Stochastic Processes with Applications to the Natural Sciences* Vol. 25 (Wiley, 1991).
- Caravagna, G., Sanguinetti, G., Graham, T. A. & Sottoriva, A. The MOBSTER R package for tumour subclonal deconvolution from bulk DNA whole-genome sequencing data. *BMC Bioinformatics* **21**, 531 (2020).
- Körber, V. & Okonechnikov, K. kokonech/mBOncoAberrations: first main release. Zenodo <https://doi.org/10.5281/zenodo.15083518> (2025).

Acknowledgements This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 819894) and Seventh Framework Programme (FP7-2007-2013) (grant agreement no. 615253). The INFORM programme is financially supported by the German Cancer Research Center (DKFZ), several German health insurance companies, the German Cancer Consortium (DKTK), the German Federal Ministry of Education and Research (BMBF), the German Federal Ministry of Health (BMG), the Ministry of Science, Research and the Arts of the State of Baden-Württemberg (MWK BW), the German Cancer Aid (DKH), the German Childhood Cancer Foundation (DKS), RTL television and the aid organization BILD hilft e.V. (Ein Herz für Kinder), and a generous private donation of the Scheu family. The study was in parts funded by the European Horizon 2020 Programme—iPC - individualized Paediatric Cure (grant no. 826121) to S.M.P. and by The Medulloblastoma Initiative to L.M.K. and S.M.P. V.K. acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project no. 526169089). K.J.W. was funded by the Mildred Scheel Career Center Frankfurt (Deutsche Krebshilfe). We also thank C. Maus, E. Wang, L. Weiser and G. Warsaw for their highly dedicated support in data management and processing, and R. Autry, G. Balasubramanian, C. Previti and R. Kabbe for their sincere and dedicated contribution to the bioinformatics analyses.

Author contributions K.O., P.J. and V.K. performed the main data processing and multi-level bioinformatics analysis. A.R., M.B. and J.-P.M. contributed with spatial data generation and analysis. P.B.G.S., B.S., J.V., D.R.G., A.W. and L.M.K. prepared tumour materials. M.S., I.S., T.Y. and H.K. contributed to the project from an evolutionary biology perspective. K.S., M.B.-J., P.F., B.J., N.J., T.M., K.W.P., C.M.v.T., O.W., K.B., K.J.W., L.N., C.R., U.S., M.M., S.R. and D.T.W.J. provided

tumour tissue samples and feedback for the manuscript. A.K., K.R., F.W., S.T. and T.H. contributed to the design of the study. K.O., P.J., V.K., L.M.K. and S.M.P. prepared the figures and wrote the manuscript on the basis of feedback from all authors. L.M.K. and S.M.P. co-led the study.

Funding Open access funding provided by Deutsches Krebsforschungszentrum (DKFZ).

Competing interests C.M.v.T. participates on advisory boards for Alexion, Bayer and Novartis. The other authors declare no competing interests.

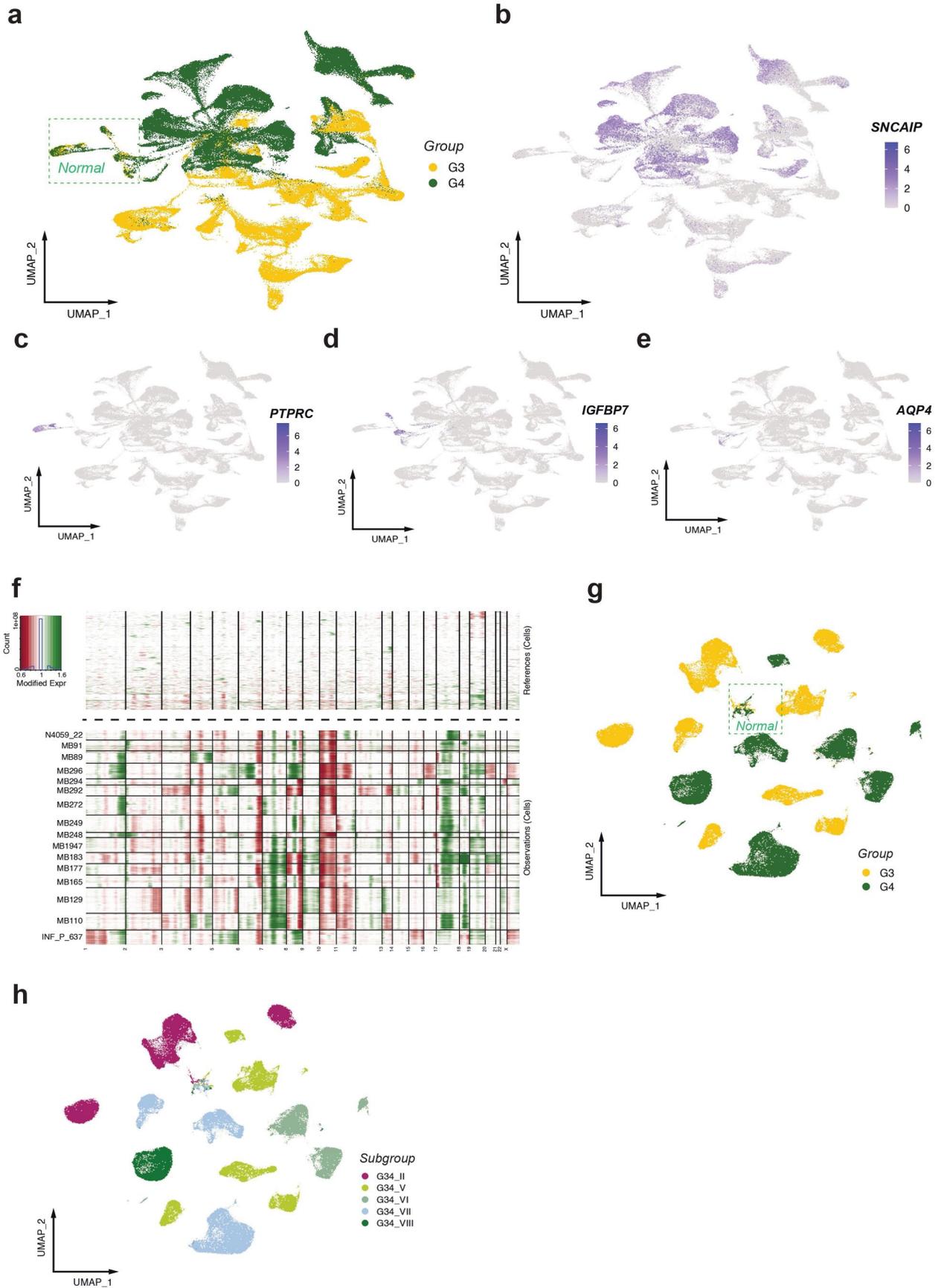
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-08973-5>.

Correspondence and requests for materials should be addressed to Lena M. Kutscher or Stefan M. Pfister.

Peer review information *Nature* thanks Richard Gilbertson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

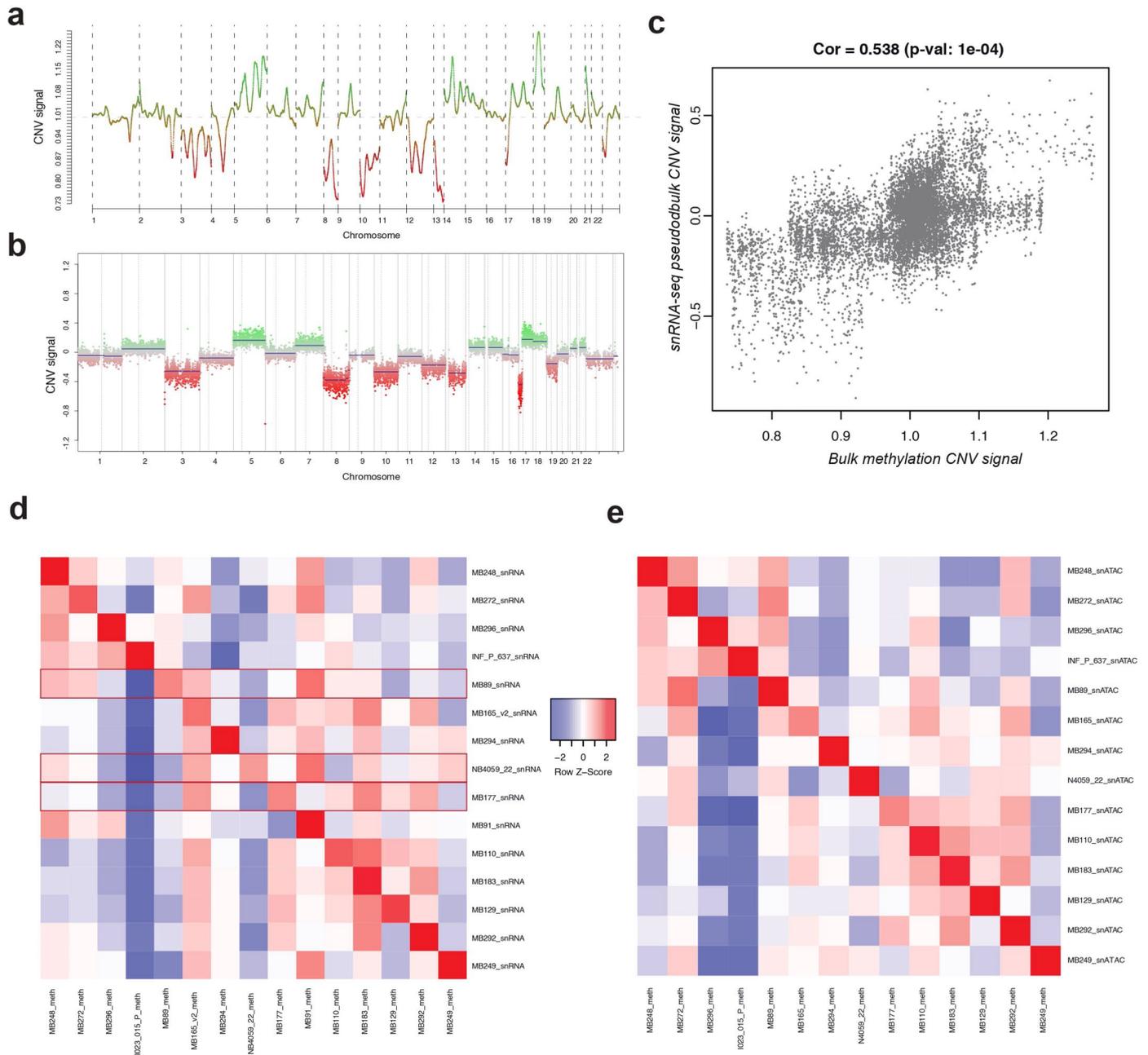
Reprints and permissions information is available at <http://www.nature.com/reprints>.



Extended Data Fig. 1 | See next page for caption.

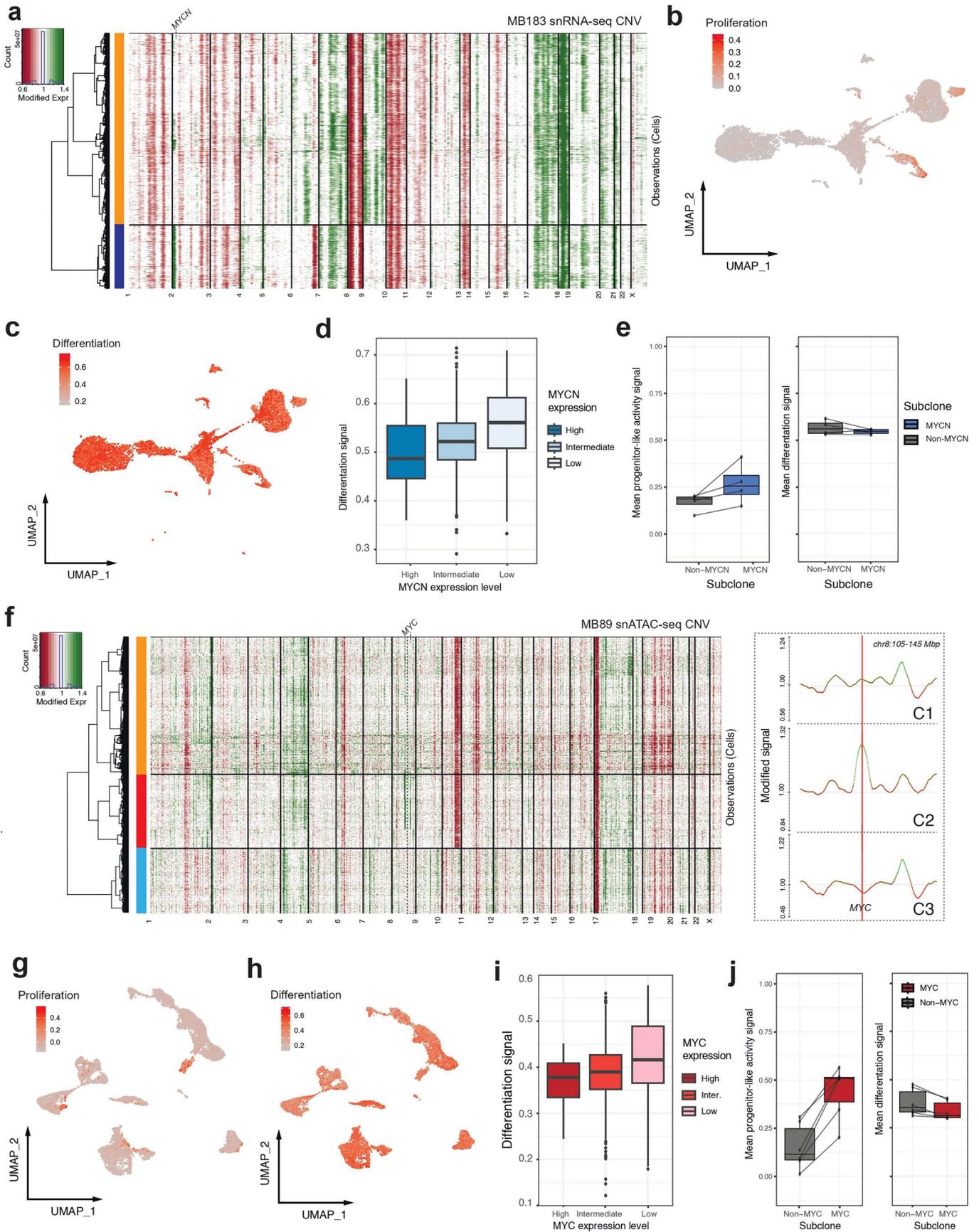
Extended Data Fig. 1 | Group 3/4 medulloblastoma single-nuclei RNA and ATAC data properties. a) UMAP of snRNA-seq merged dataset with medulloblastoma groups annotated. Non-tumor cells marked by dotted box. Feature plots showing b) *SNCAIP*, c) *PTPRC*, d) *IGFBP7* and e) *AQP4* expression within UMAP of merged snRNA-seq dataset. f) Global CNV profiles derived

from snRNA-seq data. Top fragment: non-tumor cells. G) UMAP of snATAC-seq merged dataset with medulloblastoma groups annotated. Green box, normal cells. h) UMAP of snATAC-seq merged dataset with medulloblastoma subgroups annotated.



Extended Data Fig. 2 | Validation of copy number profiling using single-nuclei RNA and ATAC data. a) Merged pseudo-bulk CNV profile of snRNA-seq data from sample MB292 b) Methylation data-derived CNV profiles from sample MB292. c) Correlation plot of CNV values across 500 Kbp bins between

snRNA-seq pseudo-bulk and methylation bulk profiles from sample MB292. d) Cross-comparison of snRNA-seq CNV profiles against bulk profiles. Red boxes, 3 cases where the highest correlation does not correspond to the same sample. e) Cross-comparison of snATAC-seq CNV profiles against bulk profiles.

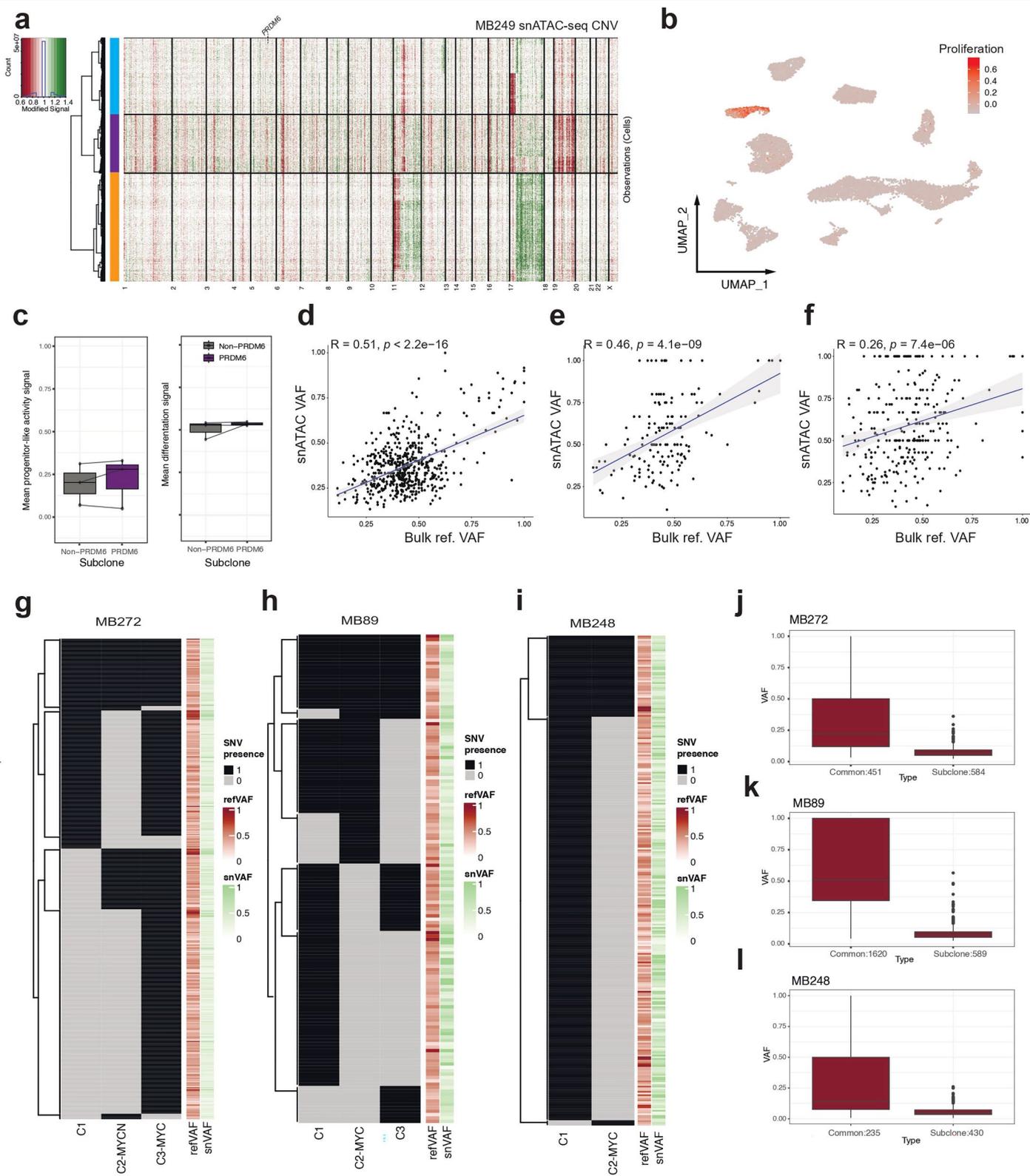


Extended Data Fig. 3 | See next page for caption.

Article

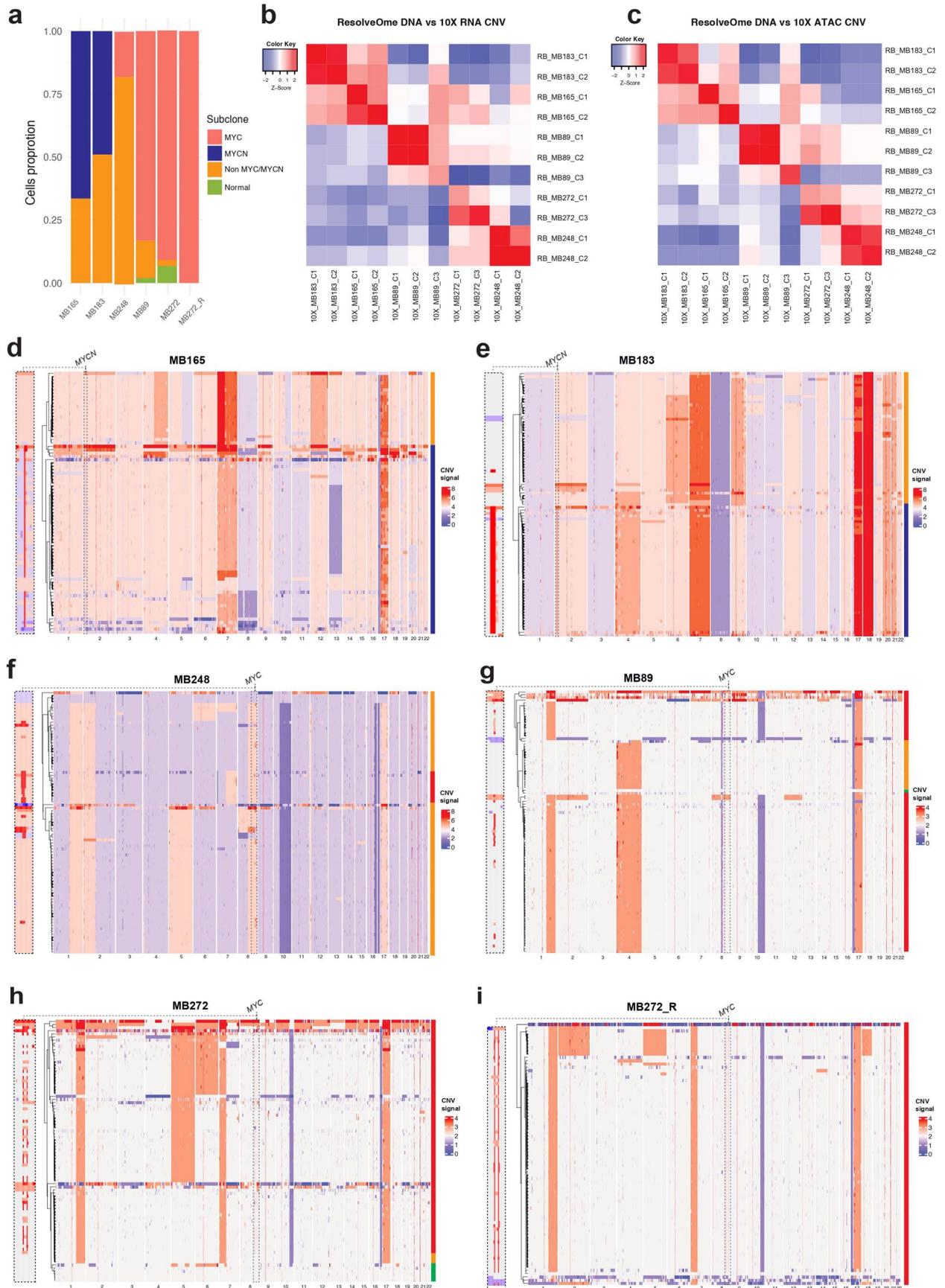
Extended Data Fig. 3 | Copy number profiling of single-nuclei profiles from Group 3/4 *MYC*- and *MYCN*-amplified samples. a) Copy number profile of snRNA-seq data from *MYCN* samples MB183. b,c) Per cell GSVA enrichment of proliferation (b) and differentiation (c) markers within UMAP of MB183 *MYCN*-amplified sample. d) Differentiation signal compared to ranked *MYCN* expression within *MYCN*-amplified subclone in sample MB183. *MYCN* normalized expression cutoffs: low = zero, intermediate > 0 and <2, high > 2. e) Boxplots showing difference in mean signal of progenitor-like activity (left) and differentiation (right) between *MYCN*-amplified and non-*MYCN*-amplified

subclones in $n = 4$ tumor cases. f) Copy number profiles of snATAC-seq data from *MYC* sample MB89. Right side: zoom-in on *MYC* region. g,h) Per cell GSVA enrichment of proliferation (g) and differentiation (h) markers within UMAP of MB89 *MYC*-amplified sample. i) Differentiation signal compared to ranked *MYC* expression within *MYC*-specific subclone in sample MB89. *MYC* normalized expression cutoffs: low = zero, intermediate > 0 and <2, high > 2. j) Boxplots showing difference in mean signal of progenitor-like activity (left side, t-test p-val: 0.003) and differentiation (right side) between *MYC*-amplified and non-*MYC*-amplified subclones in $n = 6$ tumor cases.



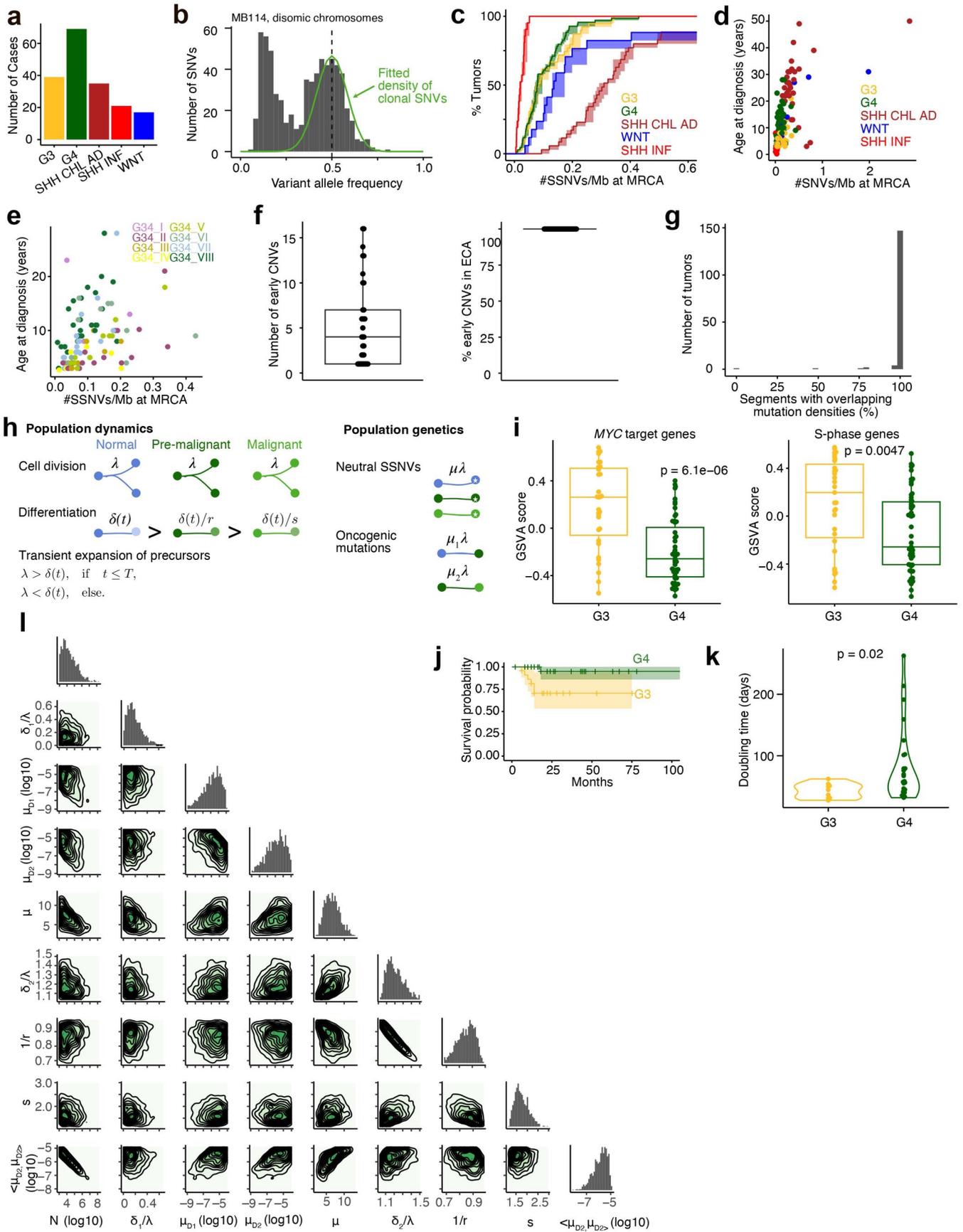
Extended Data Fig. 4 | Copy number profiling of single nuclei profiles from Group 3/4 PRDM6 samples. a) Copy number profiles of snATAC-seq from *SCNAIP-PRDM6* sample MB249. b) Per cell GSVA enrichment of proliferation markers within UMAP of MB249 PRDM6 sample. c) Boxplots showing the difference in the mean signal of progenitor-like activity (left) and differentiation (right) between PRDM6- and non-PRDM6 subclones in $n = 3$ tumor cases. d-e) Correlation of VAF between mutations called from snATAC and bulk WGS

data from $n = 3$ MYC cases: MB272 (d), MB89 (e), MB248 (f). g-i) Mutation heatmaps obtained via snATAC-seq data across subclones from $n = 3$ MYC cases: MB272 (g), MB89(h), MB248 (i). j-l) Boxplots of comparison for number of subclone-specific vs. common mutations across snATAC profiles from $n = 3$ MYC cases: MB272 (j), MB89(k), MB248 (l) SNVs were filtered using bulk germline control and additional filtering parameters (see Methods for details).



Extended Data Fig. 5 | Single-nuclei DNA and RNA data from Group 3/4 samples confirm CNV subclones. a) Projection of RNA data into annotation of subclones. b-c) Cross-comparison of snRNA-seq (b) and snATAC-seq (c) subclonal CNV pseudo-bulk profiles against scDNA subclonal CNV profiles.

d-i) Single-cell DNA copy number profiles from *MYCN* samples MB165 (d), MB183 (e), and *MYC* samples MB248 (f), MB89 (g), MB272 primary (h) and MB272 relapse (i). Zoom-in for target region (*MYC* or *MYCN*) is provided on the right side.



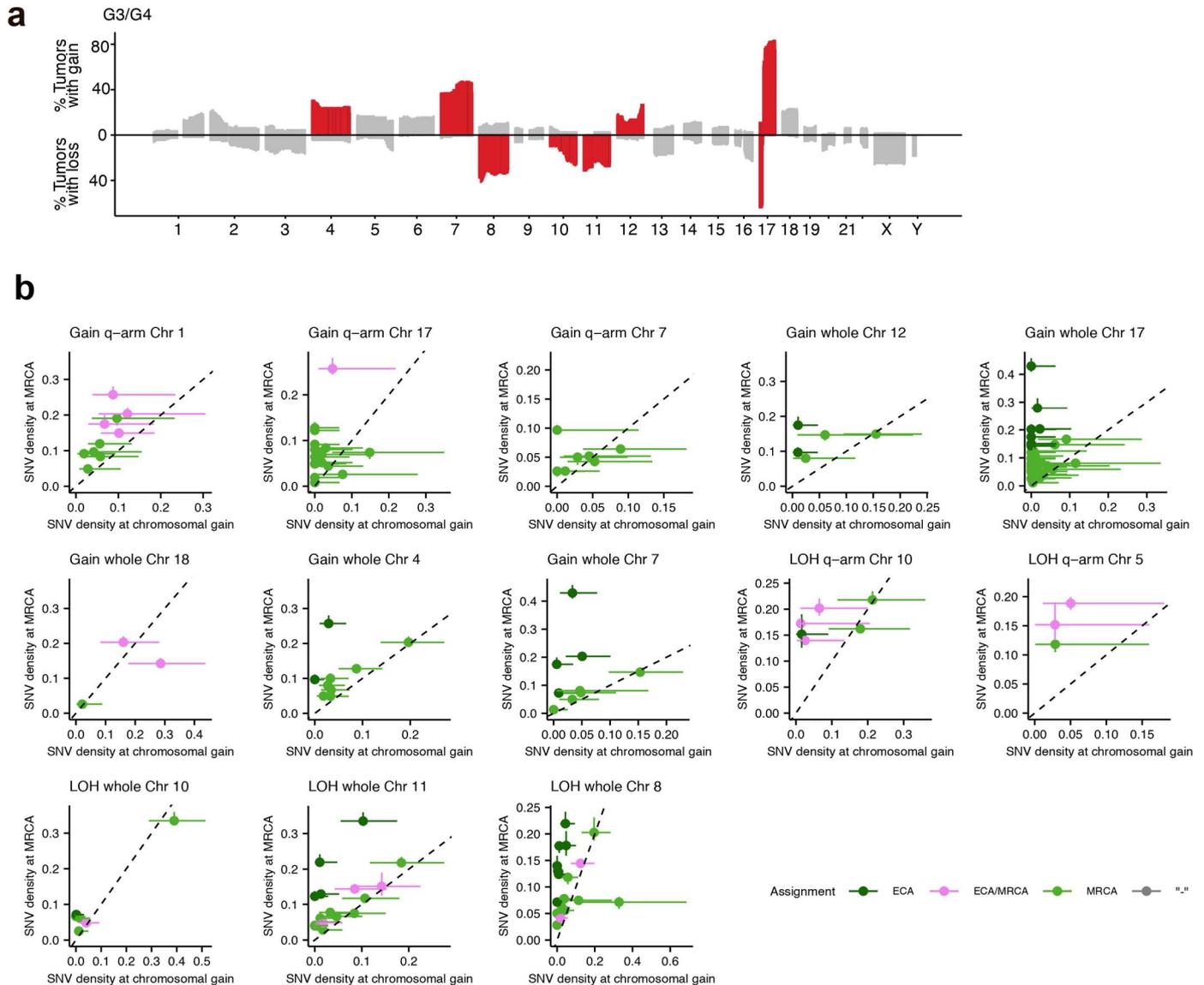
Extended Data Fig. 6 | See next page for caption.

Article

Extended Data Fig. 6 | Early evolution of Group 3/4 medulloblastoma.

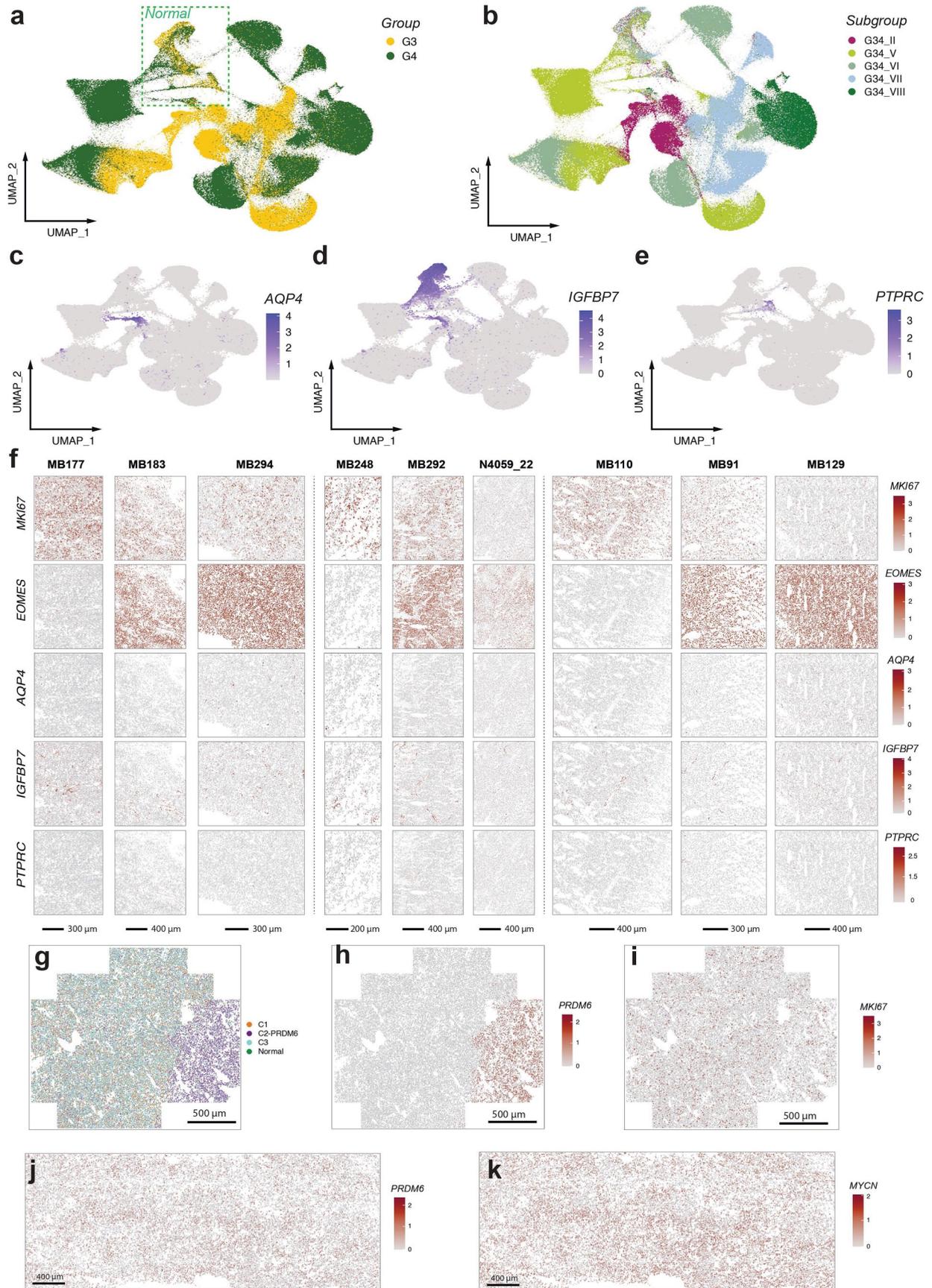
a) Medulloblastoma samples analyzed by bulk WGS. b) SNV variant allele frequencies on disomic chromosomes for sample MB104 (G34_VIII). Green line, fitted clonal SNV density; dashed line, true clonal VAF estimated with ACEseq. c), SNV densities at MRCA (39 MB G3, 69 MB G4, 21 MB SHH INF, 35 MB SHH CHL/AD, and 17 MB WNT; 4 MB SHH CHL/AD and 2 MB WNT had clonal densities between 0.7 and 2.9 SNVs/Mb and are not shown). Mean and 95% CI (estimated by bootstrapping genomic segments 1,000 times). d) Mean SNV densities at MRCA versus age at diagnosis (n = 173 cases with age information). e) As in d but for G3/4 subgroups (n = 105 cases with age information). f) Left panel, number of early CNVs per tumor. Right panel, percentage of early CNVs with SNV densities agreeing with a single ECA. Data are from G3/4 medulloblastomas with evidence

for an ECA. g), Comparison between mutation density estimates obtained in this paper and with MutationTimeR²². Estimates at gains/LOH were computed relative to the MRCA using both methods. Shown is the percentage of CNVs per tumor with overlapping 95% CIs. Data are from 38 MB G3, 66 MB G4, 15 MB SHH INF, 28 SHH CHL/AD, and 9 MB WNT with clonal gains/LOH at copy number ≤ 4 and at least 10^7 bp length. h) Population genetics model of tumors initiation in two steps. i) GSVA scores for *MYC* target genes and S-phase genes (86 G3/4 medulloblastomas with RNAseq data). j) Overall survival of 23 Group 3 and 36 Group 4 medulloblastomas with available data. k) Doubling times estimated from 35 G3/4 medulloblastomas using the population-genetics model outlined in h). l) Posterior probabilities for the model fit to all G3/4 MBs (n = 108). $\langle \mu_1, \mu_1 \rangle$, geometric mean of the driver mutation rate.



Extended Data Fig. 7 | Clonal copy number changes in Group 3/4 medulloblastoma. **a)** Percentage of tumors with copy number gains and losses ≥ 1 Mb along the genome. Red, regions where CNVs were significantly more frequent than expected, according to a Binomial test with $p_{adj} < 0.01$; Holm correction for multiple sampling. Shown are data from 109 Group 3/4

medulloblastomas. **b)** SNV densities at clonal chromosomal gains and at MRCA. Shown are mean and 95% confidence intervals (confidence intervals for SNV densities at chromosomal gains/LOH were estimated according to a Poisson distribution; confidence intervals for SNV densities at MRCA were estimated by bootstrapping genomic segments 1,000 times).

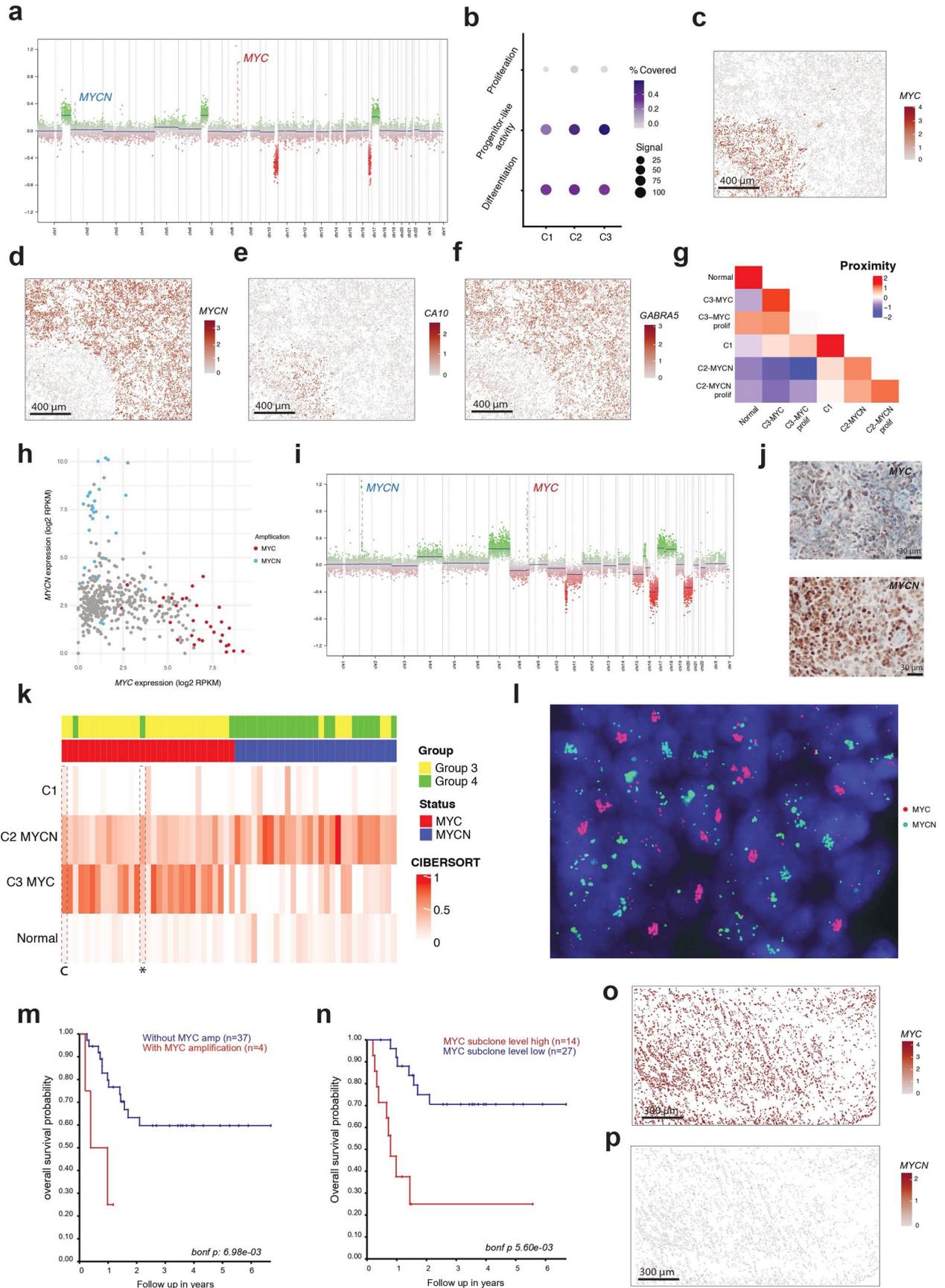


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Spatial resolution of sub-clonal tumor populations.

a) UMAP of spatial merged dataset with medulloblastoma groups annotation. Normal cells marked. b) UMAP of spatial merged dataset with medulloblastoma subgroups annotation. c-e) Feature plots showing c) *AQP4*, d) *IGFBP7* and e) *PTPRC* expression within UMAP of merged spatial dataset. f) Spatial gene

expression of *MKI67*, *EOMES*, *AQP4*, *IGFBP7* and *PTPRC* across samples. g) Spatial visualization of clones of PRDM6 sample in 2nd image fragment. h) *PRDM6* and i) *MKI67* spatial expression of sample in g. j) *PRDM6* and k) *MYCN* spatial gene expression in image fragment of sample MB292.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Independent oncogene clones may co-occur in one tumor. a) CNV profile of MB272 cases bulk methylation data. b) Per cell GSVA enrichments of proliferating, progenitor-like and differentiation in single sample MB272. Spatial expression of c) *MYC*, d) *MYCN*, e) *CA10* and f) *GABRA5* in sample MB272. g) Proximity of each compartment to each other in sample MB272 spatial data. h) Negative correlation ($R = -0.287, P = 1.08e-09$) between *MYC* and *MYCN* expression within medulloblastoma FFPE bulk RNA-seq cohort ($n = 435$). i) CNV profile of bulk methylation data from a Group 3/4 tumor with amplifications of *MYC* and *MYCN*. j) Identification of *MYC* (left) and *MYCN* (right) signals in the same sample using immunohistochemistry (IHC). k) CIBERSORT deconvolution results across subset of *MYC/MYCN* cases

from medulloblastoma bulk FFPE RNA-seq cohort. MB272 single cell data with subclones annotation used as a reference, the data from control case is marked with c, target sample marked with asterisk. l) Identification of *MYC* (red) and *MYCN* (green) signals in the highlighted target Group 3/4 sample described in panel (k) using FISH. m) Kaplan-Meier overall survival probability curves for medulloblastoma Subgroup V tumors with (red) and without (blue) *MYC* amplification as identified from bulk data CNV profiling. n) Kaplan-Meier overall survival probability curves for medulloblastoma Subgroup V tumors with high (red) and low (blue) *MYC* subclone level enrichment. o) *MYC* and p) *MYCN* expression in spatial transcriptomic relapse sample.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

Single nucleus RNA-sequencing and spatial data post analysis was performed in R 4.2 environment with the following packages: Seurat v4.1.3, dplyr v1.1.0, GSVA 1.46.0, Signac 1.10.0, InferCNV 1.10.1, scales_1.2.1 RimageROI_0.1.2, data.table 1.14.6, stringr 1.5.0, Giotto v1.1.1., QuPath v0.3.2, ggplot2 3.4.0.

Whole genome sequencing data analysis was performed in R v4.2 with the following packages: ggplot2 v3.5.1, bedr v1.0.7, openxlsx v4.2.5, NBevolution v0.0.0.9000, ggpubr v0.4.0, ggsci v2.9, pamrtools v0.5.8, tidyverse v1.3.2, mixtools v1.2.0, RColorBrewer v1.1.3, ggbeeswarm v0.6.0, survminer v0.4.9, survival v3.4.0, GSA v1.3.3, GSVA v1.44.5, msigdb v1.4.0, ExperimentHub v2.4.0, GSEABase v1.58.0, ComplexHeatmap v2.13.1, wesanderson v0.3.6, GenomicRanges v1.48.0, ggbio v1.44.1, MutationTimeR v1.0.2, VariantAnnotation v1.42.1, cdata v1.2.0, cowplot v1.1.1, mobster v1.0.0, CNAqc v1.0.0, HDInterval v0.2.2, moments v0.14.1,

Approximate Bayesian Computation was run in python v3.10.1 using pyABC v0.12.6

Cell border detection for Resolve Bioscience spatial images was performed with CellPose 2.0.5. Gene expression counts were extracted with custom Python scripts in python3.2 environment.

The source code materials for data analysis are shared via the repository:
github.com/kokonech/mbOncoAberrations

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The DNA whole genome sequencing mutation results were integrated from the corresponding medulloblastoma molecular landscape study deposited at European Genome-Phenome Archive under accession number EGAS0001001953. Single nuclei RNA and ATAC data available at GEO database under the accession numbers GSE253557 and GSE253573 accordingly. All raw images and processed data after cell segmentation from spatial transcriptomics experiments available at GEO database and can be accessed under the accession number GSE252090.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Term sex is used for patients description, provided in annotation heatmap describing the target cohort. It is not used in the research.

Reporting on race, ethnicity, or other socially relevant groupings

The samples were collected from archives such as based on material availability, we do not have information on race or ethnicity for this cohort.

Population characteristics

The study focuses on medulloblastoma brain tumors, mean age : 5 years

Recruitment

Patients were included retrospectively based on the availability of fresh frozen tumour material.

Ethics oversight

This study was performed following the regulations of the ethics committee of the Medical Faculty of Heidelberg University.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

n = 20

Sample size	Sample size is based on the availability of fresh frozen material of previously diagnosed MB G34 tumours with specific focus on cases with MYC,MYCN and SNCAIP/PRDM6 somatic alternations.
Data exclusions	No data was excluded in this study except of specific subsets not fitting quality control limits (e.g. low quality cells from single cell RNA/ATAC/spatial data)
Replication	For each target somatic change (MYC/MYCN/PRDM6) at least n=3 samples were included.
Randomization	Not applicable since the analyses were performed on collected available tumour material.
Blinding	All analyses were performed on retrospectively collected tumour material. No analyses involving living organisms were performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

- n/a | Involved in the study
- Antibodies
 - Eukaryotic cell lines
 - Palaeontology and archaeology
 - Animals and other organisms
 - Clinical data
 - Dual use research of concern
 - Plants

- n/a | Involved in the study
- ChIP-seq
 - Flow cytometry
 - MRI-based neuroimaging

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a