Neoplastic and Immune single cell transcriptomics define subgroup-specific intra-tumoral heterogeneity of childhood medulloblastoma.

Kent A. Riemondy[1], Sujatha Venkataraman[2,3], Nicholas Willard[4], Anandani Nellan[2,3], Bridget Sanford[3], Andrea M. Griesinger[2,3], Vladimir Amani[2,3], Siddhartha Mitra[2,3], Todd C. Hankinson[2,5], Michael H. Handler[2,5], Martin Sill[6], Jennifer Ocasio[7], Seth J. Weir[7], Daniel S Malawsky[7], Timothy R. Gershon[7], Alexandra Garancher[8], Robert J. Wechsler-Reya[8], Jay R. Hesselberth[1], Nicholas K. Foreman[2,3,5], Andrew M. Donson[2,3,9,*,†], Rajeev Vibhakar[2,3,9].

[1]RNA Bioscience Initiative, University of Colorado Denver, Aurora, Colorado, 80045, USA.

[2]Morgan Adams Foundation Pediatric Brain Tumor Research Program, Children's Hospital Colorado, Aurora, Colorado, 80045, USA.

[3]Department of Pediatrics, University of Colorado Denver, Aurora, Colorado, 80045, USA.

[4]Department of Pathology, University of Colorado Denver, Aurora, Colorado, 80045, USA.

[5]Department of Neurosurgery, University of Colorado Denver, Aurora, 80045, USA.

[6]Hopp Children's Cancer Center Heidelberg (KiTZ), Heidelberg, Germany. Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[7]Department of Neurology, UNC School of Medicine, Chapel Hill, NC 27599, USA

[8]Tumor Initiation and Maintenance Program, NCI-Designated Cancer Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037, USA

[9]These authors contributed equally.

*Correspondence: Andrew.donson@cuanschutz.edu

†Lead Contact: Department of Pediatrics, Mail Stop # 8302, Research Complex 1, North Tower, 12800 E. 19th Ave., Rm P18-4402A, Aurora, CO 80045.

Phone: (303) 724-4012 | Fax: (303) 724-4015 | E-mail: Andrew.donson@cuanschutz.edu

Abstract

Medulloblastoma (MB) is a heterogeneous disease in which neoplastic cells and associated immune cells contribute to disease progression. To better understand cellular heterogeneity in MB we profile neoplastic and immune populations in childhood MB samples using single-cell RNA sequencing, immunohistochemistry and deconvolution of transcriptomic data. Neoplastic cells cluster primarily according to individual sample of origin which is in part due to the effect of chromosomal copy number gains and losses. Harmony alignment reveals novel MB subgroup/subtype-associated subpopulations that recapitulate neurodevelopmental processes and are associated with clinical outcomes. We identify discrete photoreceptor-like cells in MB subgroups GP3 and GP4 and nodule-associated neuronally-differentiated cells in subgroup SHH. MB immune infiltrates consist of both developmentally-related neuron-pruning and antigen presenting myeloid cells. We show that this MB cellular diversity is recapitulated in genetically engineered mouse subgroup-specific models of MB. These findings advance our understanding of both the neoplastic and immune landscape of MB.

Introduction

Medulloblastoma (MB) is an aggressive brain tumor arising predominantly in childhood. MB is comprised of 4 consensus molecular subgroups – GP3, GP4, SHH and WNT – that were established based on a series of bulk-tumor transcriptomic, methylomic and genomic profiling studies combined with clinical data (Northcott et al., 2011). GP3 subgroup tumors often exhibit amplification of *MYC* and are associated with the poorest clinical outcome; SHH and WNT often harbor mutations in sonic hedgehog and wingless pathway genes, respectively; and the genetic driver for GP4 is less clear. Studies have subdivided MB subgroups into finer subtypes with disparate outcomes and molecular features (Cavalli et al., 2017; Northcott et al., 2017; Sharma et al., 2019). The intra-tumoral cellular heterogeneity inherent in these bulk-tumor studies impedes a clearer understanding of MB biology. Neoplastic and immune heterogeneity underlies cancer cell's ability to proliferate, survive and evade therapeutic interventions. The emergence of single cell analysis allows us to examine the cellular diversity inherent in biological systems, resulting in transformative findings when applied to pediatric brain tumors. Hovestadt *et al.* used SMART-seq single cell RNA-seq (scRNA-seq) to interrogate 25 MB (7,745 cells), identifying cellular diversity within individual samples recapitulating cell cycle, progenitor and differentiated neuronal programs (Hovestadt et al., 2019). Vladoiu *et al.* used droplet-based scRNA-seq to map mouse cerebellar developmental cellular lineages, and then compared these to childhood cerebellar brain tumors including 8 MB samples (28,311 cells), identifying putative subgroup-specific cells of origin (Vladoiu et al., 2019). In the present study, we used droplet-based scRNA-seq to analyze primary tumor samples from 28 childhood MB (approximately 40,000 cells), predominantly from GP3, GP4 and SHH subgroups. Our studies identify and characterize additional conserved MB subpopulations and validate these using histology and deconvolution of bulk-tumor transcriptomic datasets. Importantly we identify for the first time considerable immune cell diversity in MB including subgroup-specific immune populations. To examine the relationship of human tumors to experimentally induced mouse tumors that serve as preclinical models, we

identify neoplastic GEM subpopulations that correspond to those seen in human samples. Finally, we provide these data in the form of an interactive browser for users to interrogate MB subgroup neoplastic and immune cells and GEMs at single cell resolution.

Results

Sample-specific clustering in MB scRNA-seq is an effect of chromosomal copy number gains and losses.

To better understand cellular heterogeneity in MB we performed scRNA-seq on 28 primary childhood MB patient samples and a single matched sample from recurrence. MB samples were classified into subgroups (1 WNT, 9 SHH, 7 GP3 and 11 GP4) based on bulk-tumor methylome and pooled single cell transcriptome profiles (Table S1). The SHH, GP3 and GP4 samples were further assigned to more recently described subtypes (Cavalli et al., 2017; Sharma et al., 2019). Using the 10x Genomics droplet-sequencing-based platform we generated transcriptomes at the single-cell level as previously described (Gillen et al., 2020). The goal of the study was to capture sufficient cells to provide a broad view of the cellular heterogeneity of MB. After excluding poor quality cells based on low unique molecular identifier (UMI) counts and high mitochondrial proportions, 39,946 cells (~1,300 per sample) remained that passed quality controls (Figure S1).

Cell by gene expression matrices projected as 2D UMAP plots revealed multiple clusters of cells that were either unique to or shared between samples. Cell type analysis based on transcriptomic signatures identified 3 main categories of cell types – neoplastic, immune (myeloid and lymphocyte lineages) and non-neoplastic stroma (oligodendrocytes and astrocytes) (Figure 1A). The distinction between neoplastic and non-neoplastic designation was confirmed by examining copy number variants (CNVs) inferred from the single cell RNA-seq data (inferCNV). We identified numerous MB subgroup specific copy number gains and losses (Figure 1B) that

were consistent with CNVs observed by methylation analysis (Table S1). Neoplastic cell clusters were distributed broadly according to subgroup, determined by prior analysis (Figure 1A). Neoplastic clusters were most cohesive within the SHH subgroup, with 5 of the 7 SHH samples colocalizing, a reflection of underlying transcriptomic similarity. The 2 discrete SHH clusters were each patient-specific, notably from the two SHH patients with germline *TP53* mutations and particularly extensive CNV aberrations (Table S1). GP3 demonstrated the least cohesive clustering, with all patient samples clustering separately. In GP4 5 of 11 samples colocalized. These samples harbored less extensive copy number variance. Collectively, these findings suggest that sample specific clustering within subgroups is a function of the extent of CNV aberration (Figure 1C). To further investigate this observation, we calculated differential gene expression between each tumor and all other tumors (only from the neoplastic cells) to generate a list of positive and negative markers for each tumor. The proportion of markers that reside in CNVs for that tumor were compared to the proportion of all genes within a CNV for that tumor as a negative control. Positive markers of each tumor were only counted as within a CNV region if that region had a CNV gain, and negative markers were counted only within a CNV loss. For most tumors there is significant enrichment of markers within the CNV regions (Figure 1D). By extension, the broad subgroup-specific clustering regions may also be dependent on subgroup-specific CNV commonalities. These findings underscore the significant contribution of CNVs to the transcriptional landscape in MB.

Given the observed distinct subgroup-specific clustering, SHH, GP3 and GP4 subgroup neoplastic cells were individually re-clustered. We used the Harmony alignment algorithm to identify similar cell types that had clustered discretely in the prior non-aligned analysis (Korsunsky et al., 2019a). Harmony generated neoplastic clusters shared between multiple samples in each of the subgroups (Figures 2-4). SHH, GP3 and GP4 samples each formed clusters that were consistent with recent data including cell cycle (program A), undifferentiated progenitors (program

B) and neuronally-differentiated (program C)(Hovestadt et al., 2019). We therefore annotated these clusters in a similar fashion, with further subdivisions within these programs corresponding to distinct subpopulations, summarized in Table 1.

GP3 MB contain subpopulations corresponding to neuronal and photoreceptor differentiation.

GP3 tumors (n=7) formed 5 major clusters that included two differentiated cell subpopulations (GP3-C2, GP3-C1) in addition to two progenitor (GP3-B1, and GP3-B2) and a mitotic (GP3-A) subpopulation (Figure 2A,B). GP3-A, and other GP4-A and SHH-A subpopulations, were primarily enriched for cells in G2M and S-phase and therefore were labeled cell cycle subpopulations, although we note that other subpopulations contain proliferating cells to a lesser extent (Figure S2A). This is in contrast to Hovestadt *et al.* who reported that GP3 tumors with *MYC* amplification lacked differentiated cells altogether (Hovestadt et al., 2019).

Identification of differentiated subpopulations in GP3 provides insights into neuronal lineage specificity in this high-risk MB subgroup. We characterized the transcriptional profiles of these populations further using Gene Ontology Term (GOTerm) enrichment and inference of active transcription factors (TFs) using pySCENIC (single-cell regulatory network inference and clustering)(Van de Sande et al., 2020). The predominant GP3 differentiated subpopulation GP3-C1 (38% of all GP3 neoplastic cells) is characterized by genes related to RNA processing and axo-dendritic transport (Table 1, S2, S3). The less abundant GP3-C2 (15% of neoplastic cells) was distinguished by genes and TFs related to photoreceptor development (*NRL*, *IMPG2*) (Table 1, S2-4), a phenotype that has been identified and studied in GP3 but not previously recognized as a discrete subpopulation of cells (Cho et al., 2011; Garancher et al., 2018; Kool et al., 2008; Korf et al., 1987). We examined this photoreceptor like subpopulation using immunohistochemistry (IHC) to chart both morphological and spatial characteristics.

Interphotoreceptor matrix proteoglycan 2 (IMPG2), a marker of GP3-C2 that was highly enriched for photoreceptor related genes, was examined by IHC in a panel of GP3, GP4 and SHH MB FFPE samples. IMPG2 protein expression was shown to be significantly higher in GP3 than GP4 and SHH (one-tailed Student's t-test $p=1.9x10^{-7}$), with an expression score 2 or greater in 6 of 7 samples (Figure 2C). Moreover, IMPG2 was cytoplasmically expressed in discrete clusters of cells in GP3 tumors, suggesting focal areas of GP3-C2 photoreceptor differentiation (Figure 2D). Of note, *IMPG2* RNA expression has previously been identified as a classification marker of GP3 (Northcott et al., 2012).

Next using subpopulation signatures derived from scRNA-seq data we deconvoluted (Cibersort) the Medulloblastoma Advanced Genomics International Consortium (MAGIC) transcriptomic dataset to estimate subpopulation proportions in individual samples (n=747) (Cavalli et al., 2017). For each MB subgroup, we compared subpopulation proportions between samples grouped according to subtype, CNVs and clinical outcome to identify significant associations. Deconvolution analysis of the GP3 MAGIC cohort (n=141) revealed a prominent association between photoreceptor-differentiated subpopulation GP3-C2 and GP3-alpha, that has the best clinical outcome of GP3 subtypes. The GP3-C2 proportion was 2.5-fold higher in alpha than other GP3 subtypes ($p=2.3x10^{-14}$), being progressively less abundant in GP3-beta and GP3-gamma (Figure 2E). GP3 differentiated subpopulation GP3-C1 was more abundant in GP3-beta than others (p=0.0027)(Figure 2F).

GP3 progenitor proportions are influenced by MYC and chromosome 8 gain

Examination of GP3 progenitor subpopulations GP3-B1 and B2 further substantiated the previously articulated theory that aberrant *MYC* activity is responsible for preventing differentiation of GP3 progenitor subpopulations (Hovestadt et al., 2019). Both GP3-B1 and B2 shared GOTerm

enrichment of translation-related genes, including ribosomal proteins and eukaryotic translation elongation factors, and MYC TF activity (Table 1, S3,4). One normal physiological function of MYC is to directly enhance Pol I transcription of ribosomal RNA (rRNA) genes in response to the cellular demand for protein synthesis (Grandori et al., 2005; van Riggelen et al., 2010). This normal physiological role of MYC is recapitulated in the GP3-B progenitor cells, and focal amplification of *MYC* or gain of chromosome 8 is associated with increased proportions of progenitor cells versus non-*MYC* amplified GP3 samples (p=0.013)(Figure 2B). Deconvolution analysis showed a significantly higher proportion of progenitor GP3-B1 and GP3-B2 cells in GP3-gamma than in either alpha or beta subtypes (Figures 2G, H). GP3-gamma samples are distinguished by broad copy number gains on chromosome 8, focal copy number gains for *MYC* and the poorest clinical outcome of all MB subtypes (Cavalli et al., 2017). We compared all subpopulation proportions to broad CNVs reported for the MAGIC sample cohort, which showed that the strongest of these associations was between GP3-B2 and chromosome 8 CNVs, being higher in samples with chromosome 8 gain than in samples with chromosome 8 WT or loss (Figure 2I). These data are consistent with the higher proportions of GP3 progenitor subpopulations in our scRNA-seq samples with gain of chromosome 8 or *MYC* amplification (Figure 2B). With respect to clinical outcome, a higher than median proportion of GP3-B2 conferred the shortest survival of all GP3 subpopulations (Figure 2J, Table S5), consistent with the poor outcome conveyed by Gp3-gamma subtype assignment and *MYC* amplification.

The relationship between deconvoluted GP3 subpopulation proportions was examined, identifying a significant negative correlation between GP3-C2 and all other GP3 subpopulations (Table S6) in particular GP3-B1 (Figure 2K). This analysis demonstrates a continuum of differentiated and progenitor cell ratios across GP3 alpha, beta and gamma subtypes, an important finding that aid with the correct interpretation of GP3 biology.

Chromosome 8 loss is associated with lower proportion of GP4 progenitors.

Six major neoplastic subpopulations were identified in the 11 GP4 samples, closely matching those programs identified in GP3 (Figures 3A,B; Table 1). Two subpopulations were characterized by cell cycle-related gene enrichment (GP4-A1 and A2). We identified 2 progenitor subpopulations(GP4-B1 and B2) that were characterized by abundant ribosomal gene expression as seen in GP3 progenitors (Table 1, S3, S4). Deconvolution analysis of GP4 MAGIC samples (n=325) revealed that the strongest CNV association was between GP4-B1 and chromosome 8 aberrations (Figure 3C). This association differs from that seen with GP3-B2 in that GP4-B1 proportions are significantly higher in samples with WT chromosome 8 than those with chromosome 8 loss (chromosome 8 gain being largely absent in GP4 MB). This finding implicates a role for MYC in the maintenance of a progenitor subpopulation in GP4 as well as GP3 and warrants further investigation. GP4 subtypes alpha, beta and gamma also showed variable subpopulation proportions, the most distinct being GP4-gamma that had significantly lower progenitor population GP4-B1 ($p=3.2x10^{-18}$)(Figure 3D).

The predominant GP4 MB differentiated subpopulation is distinguished from other subgroups by higher expression of glutamatergic marker GRIA2

The predominant differentiated GP4 subpopulation, GP4-C1 (33% of neoplastic GP4), was characterized by gene expression similar to GP3-C1, but showed higher expression of *GRIA2*, a component of glutamatergic lineage neurons. In developing murine cerebellum neurons, *GRIA2* was identified as a marker of unipolar brush cells (UBCs)(Vladoiu et al., 2019) although also expressed in Purkinje cells and cerebellar granular neurons (CGNs)(Ocasio et al., 2019a). We compared murine cerebellar lineage scRNA-seq data (Vladoiu et al. 2019) with our MB neoplastic subpopulations which showed that murine UBC signatures were not enriched in GP4-

C1 (Figure S3A), but rather in progenitor subpopulations from all subgroups (Figure S3B). In GP4 samples, IHC analysis revealed that GRIA2 positive cells, corresponding to GP4-C1, were scattered throughout the tumor parenchyma (Figure 3E). As predicted from the scRNA-seq data, GRIA2 protein expression was highest in GP4 and lowest in GP3 (one-tailed Student's t-test p=0.045)(Figure 3F).

GP4 also showed the potential for photoreceptor differentiation, although to a much lesser extent than GP3. The small subpopulation of GP4 photoreceptor differentiated cells (GP4-C2) constituted on average 7% of GP4 neoplastic cells, and this proportion was highly variable (Figure 3B). Comparison of GP4-C2 with other neoplastic subpopulations confirmed the strong correlation with GP3-C3 (Figure S2B). The presence of photoreceptor-differentiated cells in GP4 was further examined by IHC, with 2 of 11 GP4 samples expressing IMPG2 with an IHC score of 2 or more, and isolated IMPG2 positive cells in the remainder (Figure 2G). The presence of photoreceptor-differentiated cells in GP4 may therefore explain those MB samples that are assigned to the mixed GP3/GP4 category, as illustrated by the expression of GP3 nanostring markers IMPG2 and NRL (Northcott et al., 2012) in the GP4-C2 subpopulation (Figure S4). By deconvolution analysis, GP4 subtypes alpha, beta and gamma also showed variable differentiated subpopulation proportions across the MAGIC cohort, revealing a significantly higher photoreceptor differentiated subpopulation GP4-C2 in GP4-gamma (11-fold higher than others, $p=5.7x10^{-12}$)(Figure 3G).

No single GP4 subpopulation was significantly correlated with survival (Table S5), although the ratio of progenitor (GP4-B1 and B2 combined) to differentiated (GP4-C1 and C2 combined) conveyed a significantly shorter survival (Figure 3H). Collectively these findings reveal that the better clinical survival reported for GP4-gamma compared to other GP4 subtypes (Cavalli et al., 2017) is likely attributable to a relatively higher ratio of differentiated to progenitor cells in GP4-gamma tumors.

SHH tumors contain a unique differentiated subpopulation that comprises nodular regions.

We identified six major subpopulations of neoplastic cells in SHH tumors (n=9) including 2 cell cycle (SHH-A1 and A2), 2 progenitor (SHH-B1 and B2) and 2 neuronally-differentiated subpopulations (SHH-C1 and C2) (Figure 4A,B). SHH-C2 was distinguished by stathmin gene expression (*STMN2* and *STMN4*) and axon-development ontology enrichment (Tables 1,S2,S3). In neurons, *STMN2* is involved with neuron development and growth (Hannan et al., 1996), and is expressed primarily in CGN sub-lineage postnatal excitatory cerebellar nuclei neurons, unipolar brush cells (UBCs) and UBC precursors in the developing murine cerebellum (Ocasio et al., 2019a; Vladoiu et al., 2019). In GP4, *STMN2* expression was seen in undifferentiated progenitors (Table 1), suggesting that it marks earlier stages of neuron development than *GRIA2*, consistent with its reported role in developing neurons (Stein et al., 1988). As expected, IHC revealed that STMN2 protein expressing cells were predominantly observed in SHH and GP4, with significantly lower expression in GP3 (one-tailed Student's t-test p=0.046)(Figure 4C). Furthermore, STMN2 positive SHH-C2 cells were enriched in nodules (Figures 4D,E) indicating that SHH-C2 is largely restricted to this histological feature. Nodules of neuronal maturation are a classic histologic feature of SHH that recapitulates aspects of normal cerebellar development (Eberhart et al., 2001). Consistent with nodules being exclusive to SHH, SHH-C2 signature genes are distinct from SHH-C1 and other subgroup differentiated subpopulations (Figure S2B). Using the MAGIC cohort, Cavalli *et al.* delineated a number of SHH subtypes: two infant subtypes, SHH-beta and SHH-gamma (the latter largely comprised of MB with extensive nodularity (MBEN)); a subtype in older children, SHH-alpha which harbors frequent p53 mutations; and a largely adult subtype, SHH-delta (Cavalli et al., 2017). We deconvoluted the MAGIC SHH transcriptomic cohort (n=214) and showed that SHH-C2 was the most differentially distributed subpopulation between these subtypes (Figure 4F). SHH-C2 was significantly more abundant in the infant subtypes SHH-beta

and gamma than other SHH subtypes (Figure 4F). Consistent with SHH-C2 representing the primary population of cells that constitute SHH nodules, this subpopulation was most abundant in the particularly nodular SHH-gamma subtype that is associated with SHH variant MBEN. Conversely, SHH progenitor populations, in particular SHH-B1, were significantly inversely correlated with SHH-C2 (Table S6), being highest in non-infant subtypes SHH-alpha and delta (Figure 4G). CNVs were not as strongly correlated in SHH as was seen in GP3 or GP4, the most interesting being an increase (2.1-fold higher) in cell cycle subpopulation SHH-A1 in those samples with 10q loss (Figure 4H). This supports a role of PTEN loss in upregulation of cell cycle in SHH that has was also observed in a *PTEN*-deficient SHH mouse model (Castellino et al., 2010).

The predominant neuronally-differentiated SHH subpopulation SHH-C1 (28% of neoplastic SHH cells) matched GP4-C1 in particular (Figure S2B), with enrichment of RNA processing and axo-dendritic transport ontologies (Tables 1, S2,S3). SHH-C1 was the only subpopulation significantly associated with outcome, its relative abundance being favorable for survival (Figure 4I, Table S5).

SHH progenitors are distinct from GP3 and GP4 by expression of critical components of SHH biology. These components include inferred TF activity of CGN progenitor specification gene ATOH1 (Table S4) and high expression of genes associated with negative regulation of epithelial cell proliferation and WNT signaling (*SFRP1*, *SFRP5*) (Tables S2,3). This is consistent with studies showing that WNT pathway activation is antagonistic to SHH-driven proliferation (Ocasio et al., 2019b; Poschl et al., 2014). SHH progenitor subpopulations SHH-B1 and B2 were similar to GP3 and GP4 progenitor subpopulations with respect to abundant ribosome and eukaryotic translation elongation factor gene expression (Tables S2,3). The *MYCN* gene is commonly amplified in SHH, and like cMYC can upregulate ribosomal RNA expression as a mechanism of universal upregulation of gene expression (Boon et al., 2001). Related to this, germline elongation

factor *ELP1* mutations have recently been implicated as a major driver in SHH etiology (Waszak et al., 2020), underscoring the potential importance of ribosomal biology in SHH progenitors.

The immune cell compartment of MB contains diverse myeloid cell subpopulations associated with neurodevelopment.

The immune cell compartment in MB has not yet been described in its entirety. To develop a detailed characterization of immune cell phenotypes in the MB microenvironment we specifically examined the myeloid and lymphocyte immune cell subpopulations identified in the initial analysis of the MB scRNA-seq data (Figure 1A). Cells identified as lymphocyte or myeloid (n=4669), were separated out from neoplastic cells and re-clustered, revealing four lymphocyte, six myeloid cell and a single cell cycle-related cluster (Figure 5A). Myeloid and immune cells from multiple patients' samples strongly co-clustered and therefore did not require Harmony alignment.

An unexpected degree of myeloid cell diversity was observed in the majority of MB samples analyzed (Figure 5A). Myeloid populations in the CNS are separated into two main lineages - tissue-resident microglia and peripheral bone marrow-derived myeloid cells – and both of these play roles in both immune response and neurodevelopment (Li and Barres, 2018). These shared functional roles have hindered the identification of distinguishing markers for the two etiological lineages. Consequently, we describe MB myeloid subpopulations according to neurodevelopmental or immune processes inferred from transcriptomic signatures (Table S7) and cannot for the most part reliably assign subpopulations to resident or peripheral myeloid origins.

The predominant myeloid subpopulation (34% of myeloid cells) exhibited signatures associated with neurodevelopmental roles. Specifically this subpopulation expressed all complement component 1q subunits (*C1QA*, *C1QB* and *C1QC*) (Table S7), that are responsible for the initiation of the classic complement cascade that is involved in neurodevelopmental

synapse elimination (Stevens et al., 2007). This population was thus termed complement myeloid (Complement-M). A second myeloid neurodevelopment-associated subpopulation was characterized by expression of phagocytosis receptors *MRC1* (CD206), CD163, and *MERTK* that are crucial for microglia-mediated clearance of apoptotic cell debris from the developing neurogenic niche (Fourgeaud et al., 2016). *MRC1* and *CD163* are widely recognized markers of anti-inflammatory or immunoregulatory M2 myeloid polarization. This population is referred to as the M2-activated (M2-M) myeloid population. M2-M cells showed the strongest subgroup association, being more abundant (8.8-fold) in SHH than GP3 and GP4 combined in scRNA-seq samples (Figure 5B). Myeloid cell scRNA-seq subpopulation signatures were used to deconvolute the MAGIC dataset (GP3, GP4, SHH and WNT; n=748). This analysis corroborated the observation that M2-M were more abundant in SHH that other subgroups (2.2-fold; $p=2.0x10^{-15}$) (Figure 5C). Further, M2-M abundance was significantly higher in infant SHH subtypes beta and gamma than older SHH subtypes alpha and delta (1.8-fold; $p=3.7x10^{-6}$)(Figure 5C). IHC for M2-M marker *MRC1* (CD206), stained cells with an amoeboid morphology that frequently lined SHH nodules (Figure 5D). The co-localization of M2-M with nodules, which are comprised of the neuronally-differentiated SHH-C2 subpopulation cells support the hypothesis that this myeloid subpopulation is programmed for developmental activity, such as synaptic pruning and elimination of apoptotic neuron debris, rather than an immune role. This hypothesis is also supported by age restriction of M2-M, being significantly more abundant in infant SHH subtypes that are likely more developmentally influenced.

A single myeloid subpopulation could confidently be designated as microglia, expressing widely established microglial markers (*P2RY12, TMEM119, SALL1, CX3CR1*)(Table S7). Due to the relatively low expression of myeloid cell immune activation markers (including HLA class II and cytokines/chemokines) this myeloid population was classified as non-activated microglia (NA-microglia). In contrast to M2-M, NA-microglia trended toward being less abundant in SHH

compared to GP3 and GP4 in scRNAseq samples (27-fold)(Figure 5E) and were significantly less abundant in SHH in the larger MAGIC cohort (4-fold, $p=9.4x10^{-16}$)(Figure 5F). These findings are congruent with prior studies that identified a less immune active phenotype in GP3 and GP4 than SHH (Margol et al., 2015; Maximov et al., 2019).

Two further myeloid subpopulations exhibited transcriptomic profiles associated with antigen presentation, both expressing MHC class II genes, suggesting an active immune functional role as opposed to the developmental/homeostatic roles inferred for the prior myeloid subpopulations. The first of these was distinguished from previous myeloid subpopulations by expression of chemokines (*CCL3, CCL4, CXCL8* and *IL1B*) and MHC class II (*HLA-DRB1*, *HLA-DRA, HLA-DPA*)(Table 7), suggesting immune cell recruitment and antigen presentation functions, respectively. This population is referred to as the chemokine myeloid (Chemokine-M) cell population. A second antigen presenting myeloid subpopulation was distinguished by C-lectins (*CLEC10A, CLEC4A, CLEC12A*), *CD1* subunits (*CD1C, CD1D* and *CD1E*) and MHC class II (*HLA-DQA1, HLA-DQB1*) all of which are associated with antigen presentation by macrophages and/or dendritic cells, and was therefore named the dendritic cell-like myeloid (DC-M) subpopulation. Outcome analysis of deconvoluted myeloid immune subpopulation factions did not identify any significant association with survival for antigen presentation-associated myeloid subpopulations Chemokine-M or DC-M (Table S5). A survival advantage was instead observed in those GP3 and GP4 with a higher than median proportion of complement-M cells, suggesting a significant role for myeloid cells associated with neurodevelopmental processes in MB biology.

Analysis of lymphocytes in MB was limited by their rarity in this tumor type compared to other pediatric brain tumors, including glioblastoma, as has previously been observed (Griesinger et al., 2013). The most abundant MB lymphocyte population was identified as T-cells (*CD3D, TRAC*), constituting ~80% of lymphocytes, with no separation of CD4 from CD8 T-cells at this level of resolution (Figure 5A). The remaining lymphocyte clusters were NK cells (*NKG7, GNLY*),

B-cells (*MS4A1*, *CD79A*) and regulatory T-cells (*FOXP3, CTLA4*). None of these lymphocyte subpopulations was significantly more abundant in any subgroup in scRNA-seq samples. Deconvolution of the larger MAGIC cohort did not yield usable results, likely due to the scarcity of lymphocyte specific transcripts in bulk transcriptomic data.

Genetically engineered mouse models of MB contain subpopulations corresponding to human subgroup-specific subpopulations

Genetically engineered mouse models (GEMs) for GP3 and SHH MB provide valuable *in vivo* experimental models. We sought to define the cellular heterogeneity present in these with respect to identified human MB subpopulations, which is critical for correct interpretation of GEM experimental results. Two GP3 models, the first (termed MP) driven by overexpression of cMyc and dominant negative *Trp53* (Pei et al., 2016; Pei et al., 2012), and the second driven by co-expression of Myc and Gfi1 (and termed MG)(Lee et al., 2019; Northcott et al., 2014), and one SHH model driven by mutant Smo activated in the Atoh1 lineage (Math1-Cre/SmoM2 - termed MS)(Mao et al., 2006) were examined using scRNA-seq. Harmony was used to align GEM with human MB subgroup scRNA-seq data, revealing that GEM cells clustered most closely their intended corresponding human MB subgroups (Figure 6A,B).

Both GP3 GEMs were generated by retroviral overexpression of oncogenes in neural progenitors, and transplantation of these progenitors into the cerebellum of adult NOD-SCID-IL2R-gamma (NSG) mice. In model MP, scRNA-seq of 5,422 cells identified 7 neoplastic and 2 non-neoplastic subpopulations (immune and fibroblast) (Figure 6C). The 7 neoplastic GEM subpopulations were compared to human GP3 neoplastic subpopulations using three methods: (i) comparison of the top 200 marker genes for GEM neoplastic subpopulations (Table S8) compared to corresponding subgroup subpopulations (Figure 6D) as used previously to compare

neoplastic subpopulations between subgroups; (ii) UMAPs of GEM cells labeled with human neoplastic subpopulation metagene signatures (Figure S5A); and (iii) correlation of the genome-wide mean expression of each GEM neoplastic subpopulation with human neoplastic subpopulations which are then visualized using a directed bipartite graph (Figure S5B)(Laughney et al., 2020). Results of these analyses were used to assign GEM subpopulations to cell cycle (A), progenitor (B) and differentiated (C) phenotypes. This approach identified MP subpopulations corresponding to human GP3 cell cycle (MP-A1, -2), progenitor (MP-B1,-B2, -B3) and differentiated neoplastic subpopulations (MP-C1, -C2) (Figure 6C). Similarly, the second GP3 GEM MG harbored neoplastic and non-neoplastic subpopulations (8,144 cells) (Figure 6E), which were comprised of 9 neoplastic clusters that grossly corresponded to progenitor (MG-B1, -B2, -B3, -B4) and differentiated subpopulations (MB-C1, -C2, -C3) (Figures 6F, S6, Table S8), but with two subpopulations with no apparent human equivalent (MB-N1, -N2) and no apparent cell cycle subpopulations. Both GP3 GEMs harbored subpopulations corresponding to the neuronally-differentiated human GP3-C1, but with only a very minor subpopulation of cells that were enriched for photoreceptor differentiated GP3-C2 metagenes (Figures S5A, S6A). These data suggest that the MP GEM caries more fidelity to human GP3 tumors

SHH GEM model MS was comprised of subpopulations that also corresponded to human SHH neoplastic subpopulations. Five neoplastic subpopulations and 8 non-neoplastic subpopulations were revealed by scRNA-seq analysis (12421 cells; Figure 6G). Neoplastic cells (n=7,123) showed strong concordance with human SHH neoplastic subpopulations, comprising 2 cell cycle (MS-A1, -2), 1 progenitor (MS-B1) and 2 differentiated subpopulations (MS-C1, -C2). (Figure 6H, Figure S7, Table S8), including a neuronally-differentiated subpopulation (MS-C2) corresponding to the human SHH nodule-associated SHH-C2 subpopulation. Cumulatively we show that the GEM MB models are consistent with human MB and can be used to test therapeutic response and resistance to novel agents.

Human and mouse MB cell atlas browser

We have created a publicly available scRNA-seq resource, the Pediatric Neuro-oncology Cell Atlas (pNeuroOncCellAtlas.org). This interactive resource allows users to study expression of transcripts at the single cell level for neoplastic and immune cells in human samples and GEM models, with extensive annotations.

Discussion

Recognition of the inherent intra-tumoral cellular heterogeneity in MB is critical for the correct interpretation of MB cancer biology, an aggressive childhood brain tumor that has seen few treatment advances in 20 years. The present study addresses this problem using scRNA-seq to study childhood MB, revealing neoplastic and immune subpopulations that had not previously been identified. We characterize these previously unexplored subpopulations to provide deeper insights into a broad range of aspects of MB biology.

Our study identified associations between chromosomal aberrations both with single cell clustering patterns and progenitors proportions. We show that separation of sample-specific cell clusters is driven in part by the expression of genes located in areas of chromosomal copy number gain and loss. Copy number variance of chromosome 8 correlated with the proportion of progenitor populations in GP3, as previously observed, but this association was also observed in GP4. These findings underscore the important role that CNVs play in MB cellular heterogeneity and their potential impact on subgroup/subtype assignment. MB cellular heterogeneity may confound tumor sub-classifications that are based on sequencing of bulk-tumor samples. Several very large international consortium studies have accumulated hundreds of patient samples and subjected bulk-tumor samples to genomic, transcriptomic and methylomic analyses (Cavalli et al., 2017; Northcott et al., 2017; Sharma et al., 2019). This work has resulted in identification of well-

established molecular subgroups of MB, and further subdivision of these into multiple subtypes. In addition to the effect of CNVs, our study reveals differential subpopulation proportions between MB subgroups and subtypes, and some restriction of consensus subgroup transcriptional markers to discrete subpopulations. It will be critically important to determine how differences in CNVs and cellular heterogeneity correspond with difference between tumor subgroups.

Our analysis revealed the presence of a photoreceptor-differentiated subpopulation that was seen predominantly in GP3 and to a lesser extent in GP4 tumor samples. This finding in GP3 samples is concordant with numerous prior studies that had identified a photoreceptor gene signature in MB (Korf et al., 1987) and more recently as a defining feature of GP3 (Cho et al., 2011; Kool et al., 2008). The proportion of photoreceptor differentiated cells was variable in GP3, being inversely proportional to MYC-associated progenitor subpopulations. In the MAGIC transcriptomic dataset, the photoreceptor subpopulation was estimated to be significantly greater in the GP3-alpha than GP3-gamma MYC-associated subtype, and consequently is associated with a comparatively favorable clinical outcome. The presence of photoreceptor subpopulation in a subset of GP4 samples runs counter to the prior understanding of photoreceptor differentiation as a hallmark of GP3 and serves as an example of the underlying biological commonalities between GP3 and GP4 that confound subgroup assignment based on bulk-tumor sample analysis.

We show that neuronally-differentiated subpopulation SHH-C2 is the major cell type that constitutes SHH nodules. The SHH-C2 subpopulation was estimated to be the most abundant in infant subtypes SHH-beta and -gamma, further supporting the association of nodule formation with early development. SHH-C2 is particular abundant in the SHH-gamma subtype that is largely composed of the MBEN histological type that carries a particularly favorable prognosis. Functional analyses of the molecular pathways driving differentiation toward the SHH-C2 is feasible now that this subpopulation has been molecularly defined and may reveal therapeutically relevant insights.

This approach can explored by application of scRNA-seq to *in vivo* GEM models, that we and others have shown to recapitulate the cellular heterogeneity of SHH MB (Cheng et al., 2020).

The diversity of MB infiltrating myeloid subpopulations revealed by our study is significantly greater than what was previously understood using flow and histological studies. A number of myeloid subpopulations are distinguished either by neurodevelopment or antigen-presenting gene signatures. During early development, myeloid cells guide neural development, in part by interacting with developing neurons, phagocytosing apoptotic cells, pruning synapses, modulating neurogenesis, and regulating synapse plasticity and myelin formation (Schafer and Stevens, 2015). Two MB myeloid subpopulations exhibited neurodevelopmental-related characteristics, which differs from the immune role that has been assumed for MB tumor infiltrating immune cells. In SHH we identified an M2-myeloid subpopulation that was particularly abundant in nodular linings. In GP3 and GP4, we identified a complement-expressing myeloid subpopulation with potential neurodevelopmental roles that was associated with survival. These neurodevelopmentally-related myeloid subpopulations may be either subgroup or age restricted, M2-myeloid cells being specific to SHH that is predominant in infants. In addition to those M2-myeloid subpopulation with a potential role in neurodevelopment, we identified a number of subpopulations harboring gene expression profiles indicative of active immune roles, in particular chemokine- and DC-like myeloid subpopulations that are both characterized by MHC class-II expression. Conversely, the presence of naïve microglia is consistent with previous observations of a particularly immunosuppressed microenvironment in MB (Griesinger et al., 2013). As this “cold” immunophenotype is likely to impede immunotherapy in MB, strategies targeting the specific immunophenotype of MB are necessary and can be advanced based on the findings of the present study.

Collectively, our study provides further insight into the neoplastic and immune cellular heterogeneity of the most common medulloblastoma subgroups and how this intra-tumoral

heterogeneity is reflected in GEM models of MB. These data show that tumor cells present a range of differentiation states that varies with each subgroup, and also show subgroup-specific interaction of the tumors with their immune microenvironments. We also provide interactive browsers for each of the described datasets that will facilitate the on-going interpretation of complex childhood MB biological data.

Acknowledgements

K.A.R. is supported as an Informatics Fellow of the RNA Bioscience Initiative, University of Colorado School of Medicine. This study was supported by the Morgan Adams Foundation. The University of Colorado Denver Genomics and Microarray, Flow Cytometry, and Histology Shared Resources are supported by the University of Colorado's NIH/NCI Cancer Center (P30CA046934).

Author Contributions

A.M.D. and S.V., performed experiments with assistance from A.N., A.M.G., V.A., A.G., T.C.H and M.H.H. K.A.R. performed bioinformatic analyses and B.S., S.J.W., D.S.M. and M.S. made additional bioinformatic contributions. N.W. assisted with neuropathology concerns. T.R.G., A.G., R.J.W-R. and J.O. provided GEM samples. K.A.R., R.V., S.M., J.R.H., N.K.F. and A.M.D. were responsible for the design of the study. All authors assisted with manuscript preparation.

Competing Interests statement

None of the authors have competing interests

Figure titles and legends

Figure 1. Sample specific clustering in MB scRNA-seq is a function of the extent of chromosomal copy number gain and loss. (A) Unaligned UMAP projection of single-cell expression data of 28 MB patient samples reveals neoplastic clusters and non-neoplastic lymphocyte and myeloid clusters. (B) Inference of CNVs in scRNA-seq data. (C) Copy number gain or loss event count overlaid onto unaligned UMAP projection of MB cells. (D) Proportions of sample-specific over- and under-expressed marker genes in regions of copy number gain and loss respectively, compared to portion of all genes in areas of CNV (hypergeometric enrichment analysis, * p= < 0.01, ** p< 1e-10, *** p< 1e-100.)

Figure 2. GP3 MB are comprised of progenitor and differentiated neoplastic subpopulations including a photoreceptor-differentiated cluster. (A) Harmony alignment of GP3 neoplastic cells colored by identified clusters. (B) Subpopulation proportions in each GP3 sample. (C) GP3-C2 photoreceptor differentiated subpopulation marker IMPG2 IHC score in MB (GP3, n=7; GP4, n=11; SHH, n=10). (D) Representative IHC staining pattern of IMPG2 (brown) in a GP3 patient sample (scale bar = 50μm). Deconvoluted (E) GP3-C2 photoreceptor, and (F) GP3-C1 neuronally-differentiated subpopulation fractions i(MAGIC GP3 cohort). Deconvoluted (G) GP3-B2, and (H) GP3-B1 progenitor subpopulation fractions in the (MAGIC GP3 cohort: GP3-alpha, n=67; GP3-beta, n=37; GP3-gamma, n=37). Association of (I) chromosome 8 gain, and (J) patient survival with proportion of GP3-B2 (MAGIC GP3 cohort) (see also Table S5). (K) Subpopulation fraction correlation (Pearson) between GP3-C2 and GP3-B1 (MAGIC GP3 cohort) colored by subtype (see also Table S6).

Figure 3. GP4 MB neoplastic cell heterogeneity. (A) Harmony alignment of GP4 neoplastic cells colored by identified clusters.(B) Subpopulation proportions in each GP4 sample. (C) Association of chromosome 8 loss, and (D) subtype with deconvoluted proportion of GP4-B1 (MAGIC GP4 cohort). (E) Representative IHC staining pattern of GP4-C1 neuron differentiated subpopulation marker GRIA2 (brown) in a GP4 patient sample (scale bar = 50μm). (F) GP4-C2 subpopulation marker GRIA2 IHC score in a cohort of MB (GP3, n=7; GP4, n=11; SHH, n=10). (G) Deconvoluted GP4-C2 photoreceptor subpopulation fractions (MAGIC GP4 cohort: GP4-alpha, n=97; GP4-beta, n=109; GP4-gamma, n=119). (H) Association of patient survival with ratio of progenitor to differentiated deconvoluted fractions (MAGIC GP4 cohort) (see also Table S5).

Figure 4. SHH MB neoplastic subpopulations include a nodule-associated neuronally-differentiated subpopulation. (A) Harmony alignment of SHH neoplastic cells colored by identified clusters. (B) Subpopulation proportions in each SHH sample. (C) SHH-C2 nodule-associated subpopulation marker STMN2 IHC score in a cohort of MB (GP3, n=7; GP4, n=11; SHH, n=10). (D,E) Representative IHC staining patterns of SHH-C2 subpopulation marker STMN2 (brown) in a SHH patient samples (scale bars = 100μm (D) and 50μm (E)). Deconvoluted (F) SHH-C2 nodule-associated, and (G) SHH-B1 progenitor subpopulation fractions (MAGIC SHH cohort: SHH-alpha, n=62; SHH-beta, n=33; SHH-delta, n=75; SHH-gamma, n=44). (H) Association of chromosome 10q loss with deconvoluted proportion of GP4-A1 (MAGIC SHH cohort). (I) Association of patient survival with deconvoluted SHH-C1 neuron-differentiated fraction (MAGIC SHH cohort) (see also Table S5).

Figure 5. The immune landscape of MB. (A) non-harmonized alignment of MB tumor infiltrating immune cells colored by identified clusters.(B) M2-myeloid subpopulation fractions (scRNA-seq

cohort; GP3, n=7; GP4, n=11; SHH, n=9). (C) Deconvoluted M2-myeloid (M2-M) subpopulation in MB subtypes (MAGIC cohort). (D) Representative IHC staining patterns of M2-myeloid subpopulation marker MRC1 (brown) in an SHH patient sample showing accumulation at nodule linings (scale bar = 100μm). (E) Non-activated microglia (NAM) subpopulation fractions (scRNA-seq cohort). (f) Deconvoluted non-activated microglia (NAM) subpopulation in all MB subtypes (MAGIC cohort).

Figure 6. GEM cellular heterogeneity and fidelity with human MB subpopulations Alignment of GEM with human MB subgroup scRNA-seq data, revealing that GEM cells clustered most closely their intended human MB subgroups. (A) UMAP projection of Harmony aligned (theta=2) GEM and human MB scRNA-seq data. (B) Hierarchical clustering of GEM and human MB scRNA-seq data. (C) Non-harmonized alignment of GEM model MP single cells colored by identified clusters, and labelled according to corresponding subgroup human neoplastic subpopulations, and (D) Jaccard index for the top 200 marker genes between GEM model MP and human GP3 neoplastic subpopulations. This was repeated for (E,F) GP3 GEM model MG, and (G,H) SHH model MS. Abbr: MP, GP3 GEM Myc/Trp53; MG, GP3 GEM Myc/GFI1; MS, SHH GEM Math-1/SmoM2; vasc., vascular; oligo, oligodendrocyte.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
GRIA2 mouse MAb (1:200)	Abcam	Clone EP929Y
IMPG2 rabbit PAb (1:400)	Novus	NBP2-58919
MRC1 mouse Mab (1:500)	Sigma	Clone 5C11
STMN2 rabbit PAb (1:1000)	Novus	NBP1-49461
Biological Samples		
Human surgical samples (refer to Table S1)	University of Colorado / Morgan Adams Foundation tumor bank	Refer to Table S1
Chemicals, Peptides, and Recombinant Proteins		
Critical Commercial Assays		
Allprep RNA/DNA Mini Kit	Qiagen	Cat #80204
Chromium Single Cell V2 and V3 Chemistry Library Kit, Gel Bead & Multiplex Kit and Chip Kit	10x Genomics	Cats #120237 #120262 #120236
Deposited Data		
scRNAseq and Methylation data superseries		To be uploaded to GEO
MB scRNAseq browsable webresource		pneuroonccellatlas.org
Experimental Models: Cell Lines		
Mouse MB model MP: MycT58A/ DNp53	Tumor Initiation and Maintenance Program, NCI-Designated Cancer Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037, USA	MycT58A/ DNp53

Mouse MB model MG: MycT58A/ GFI1	Tumor Initiation and Maintenance Program, NCI-Designated Cancer Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037, USA	MycT58A/ GFI1
Mouse MB model MS: Math1-Cre/SmoM2	Department of Neurology, UNC School of Medicine, Chapel Hill, NC 27599, USA	Math1-Cre/SmoM2
Experimental Models: Organisms/Strains		
Oligonucleotides		
Software and Algorithms		
R (version 4.0.2)	CRAN	https://cran.r-project.org/
Bioconductor R (version 3.11)	Roswell Park Comprehensive Cancer Center	https://www.bioconductor.org/
CellRanger (version 3.0.2)	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome
Seurat (version 3.2)	Satija laboratory	https://satijalab.org/seurat
pySCENIC (version 0.9.18+1.ge5f9b7c)		https://github.com/aertslab/pySCENIC
MolecularNeuropathology.org version 12	German Cancer Research Center (DKFZ)	https://www.molecularneuropathology.org/mnp
InferCNV (version 1.3.6)	Broad Institute	https://github.com/broadinstitute/inferCNV
Custom code	Github	https://github.com/rnabioco/medulloblast
Prism 8.0.1	GraphPad	NA
Other		

STAR Methods

RESOURCE AVAILABILITY

Lead Contact.

Further information and requests for reagents may be directed to and will be fulfilled by the Lead Contact, Andrew Donson (andrew.donson@cuanschutz.edu).

Materials Availability.

This study did not generate any new unique reagents

Data and Code Availability.

Custom code used in this study is available at https://github.com/rnabioco/medulloblast. ScRNAseq and methylation data to be deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) database. All MB scRNAseq data will also be made available as a browsable web resource at the Pediatric Neuro-oncology Cell Atlas (pneuroonccellatlas.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient sample tumor cell preparation

MB samples for the scRNA-seq study were dissociated and viably banked at the time of surgery at our institution over a 10-year period. Patient material was collected at our institution at the time of surgery with consent (COMIRB 95-500). Samples were collected into PBS for subsequent scRNA-seq analysis and snap frozen for bulk-tumor methylome analysis (28 primary

and 1 matched recurrence sample) (Table S1). ScRNA-seq samples were rapidly dissociated into single cells using a mechanical process as described previously (Griesinger et al., 2013), viably cryopreserved and banked at <80°C for later use. In this way we were able to batch samples and thus limit experimental variance without compromising sample quality, as DMSO cryopreservation has been shown to match and in some cases exceed the quality of fresh samples with respect to scRNA-seq analysis (Wohnhaas et al., 2019).

Methods for MP and MG Tumor Generation and Tumor Cell Preparation

Cerebellar stem/progenitor cells (Prominin1+ cells) were purified by fluorescence activated cell sorting (FACS) from the cerebella of postnatal day 5–7 (P5–P7) C57BL/6J pups as previously described (Northcott et al., 2014; Pei et al., 2016). To generate MP or MG tumors, cells were infected with viruses produced from pMSCV-MycT58A-IRES-Luc and pMSCV-DNp53-IRES-GFP (for MP tumors) or pMSCV-GFI1-IRES-GFP (for MG tumors). After overnight infection, cells were washed and 10,000 cells were stereotactically injected into the cerebellum of 6- to 8-week-old NSG mice. Animals were monitored weekly and euthanized when they showed signs of tumor. Tumors were then enzymatically dissociated and viably frozen prior to scRNA-seq analysis.

Method for Math1-Cre/SmoM2 SHH GEM tumor generation and tumor cell preparation

To generate mice with SHH-driven medulloblastoma, we crossed SmoM2 mice (Jackson Labs, stock # 005130) with Math1-Cre (Jackson Labs, stock #011104) to generate M-Smo mice. All mice were of species Mus musculus and crossed into the C57BL/6 background through at least five generations. All animal studies were carried out with the approval of the University of North Carolina Institutional Animal Care and Use Committee under protocols (19-098). Mice were raised until P15 and then tumors were harvested under general anesthesia. Tumor samples were

dissociated using the Papain Dissociation System (Worthington Biochemical) as previous described (Ocasio et al., 2019a). Briefly, tumor samples were incubated in papain at 37°C for 15 min, then triturated and the suspended cells were spun through a density gradient of ovomucoid inhibitor. Pelleted cells were then resuspended in 2 ml of Bambanker media (fisher Scientific # NC9582225) and viably frozen prior to scRNA-seq analysis.

METHOD DETAILS

Tissue harvest for single cell suspension, single-cell RNA library preparation and sequencing

ScRNA-seq was performed as previously described (Gillen Cell Reports 2020). Briefly, samples were thawed in batches and flow sorted (Astrios EQ) to obtain viable single cells based on propidium iodide (PI) exclusion. With the study goal of performing scRNA-seq on 2,000 cells per sample, we utilized a Chromium Controller in combination with Chromium Single Cell V2 and V3 Chemistry Library Kits, Gel Bead & Multiplex Kit and Chip Kit (10X Genomics). Transcripts were converted to cDNA, barcoded and sequenced on Illumina NovaSeq6000 to obtain approximately 50 thousand reads per cell.

ScRNA-seq data analysis

Raw sequencing reads were demultiplexed, mapped to the human reference genome (build GRCh38) and gene-expression matrices were generated using CellRanger (version 3.0.1). The resulting count matrices were further filtered in Seurat 3.1.0 (https://satijalab.org/seurat/) to remove cell barcodes with less than 200 genes or 500 UMIs, more than 30% of UMIs derived from mitochondrial genes, or more than 40,000 UMIs. This filtering resulted in 39,946 single cells

across all samples (Figure S1). After normalization, these cells were clustered using the Seurat workflow based on dimensionality reduction by PCA using the 4,000 most variable genes. Coarse cell types were defined (Immune, myeloid, and stroma cells) were assigned based on marker gene expression. Tumor cells were readily identified based on their discrete clustering patterns, and the presence of CNVs generated using inferCNV(v.1.3.6). Tumor cells from each subgroup were then reanalyzed separately. We applied Harmony alignment (theta = 1.5) to correct for inter-sample variation due to experimental or sequencing batch effects (Korsunsky et al., 2019a). After assessment of clustering using a variety of dimensions, we used 50 harmony dimensions to cluster the data and perform dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP). Differential expression and marker gene identification was performed using Presto (Korsunsky et al., 2019b).

Chromosomal CNVs of single cells from scRNA-seq were inferred on the basis of average relative expression in variable genomic windows using InferCNV (https://github.com/broadinstitute/inferCNV). Cells classified as non-neoplastic were used to define a baseline of normal karyotype such that their average copy number value was subtracted from all cells.

Neoplastic subpopulations were characterized by direct examination of neoplastic-subpopulation specific gene lists, GO-term enrichment analysis, and inference of transcription factor regulatory networks. GO-term enrichments were calculated by gProfiler2 (v.0.1.9). Module scores for each subpopulation gene signature were generated using the top 200 markers from each subpopulation ranked by adjusted p-value, then secondarily ranked by positive fold change. Modules were calculated based on the methods in Tirosh, I et al, as implemented in the Seurat function AddModuleScore (Tirosh et al., 2016). Single-cell regulatory network inference and clustering (pySCENIC) was used to identify transcription factor regulatory networks at the single cell level (Van de Sande et al., 2020) . Bipartite graphs were generated, using iGraph, between

subpopulations using the correlation of normalized gene expression of shared variable genes as edge weights (Laughney et al., 2020). Edges were removed if adjusted p-values were < 0.05 or if the correlations were less than 0.2 (GEM to human comparisons) or less than 80% of the maximum correlations for human to human comparisons. One-to-one orthologs were used when comparing human and mouse single cell datasets and gene signatures. Harmony alignment (theta = 2) was performed to generate a UMAP projection containing mouse and human neoplastic subpopulations.

Bulk-tumor methylome analyses

DNA was extracted from snap frozen EPN-PF surgical tumor samples (Qiagen, Allprep DNA/RNA mini kit). Twenty eight surgical tumor samples were from initial presentation and 1 matched sample from a metastatic first recurrence. Methylome analysis of DNA from presentation samples was performed using the Illumina 850K methylation array. Resulting IDAT files were uploaded to MolecularNeuropathology.org (https://www.molecularneuropathology.org/mnp) which provided subgrouping into MB molecular subgroups and chromosomal copy number variants (CNVs) (Table S1). Methylomic profiles of MB samples were further classified to MB subtypes by Dr. Martin Sill.

Deconvolution

Cibersort was used to perform deconvolution of bulk-tumor MB transcriptomes using scRNA-seq subpopulation signature genes (Newman et al., 2015). The combined MAGIC transcriptomic dataset (n=747) was used unlogged and used as the mixture file. A signature gene input matrix was generated from log2 values of normalized raw scRNA-seq expression data

averaged for neoplastic and immune clusters (Table S9). Cibersort was run using these datasets with 100 permutations.

Immunohistochemistry

Immunohistochemistry was performed on 5-μm formalin-fixed, paraffin-embedded tumor tissue sections using a Ventana autostainer. Antigen retrieval was performed by incubation in citrate solution pH 6.0 for 10 min at 110°C. Slides were treated with primary antibodies for GRIA2 (1:200 dilution), IMPG2 (1:400), MRC1 (1:500) and STMN2 (1:1000) for 32 minutes at 37°C. All immunostained sections were counterstained with hematoxylin. Neuropathological review of staining and blinded IHC scoring was then performed (A.M.D., N.W.).

Survival studies

Survival data was obtained for a subset of the MAGIC transcriptomic cohort (n=539). Survival analyses were performed using Prism (GraphPad) software, with outcome censored at 5 years. Hazard ratios (HRs) for progression free survival (PFS) and overall survival (OS) were estimated using log-rank (Mantel–Cox) analysis of high versus low (i) expression of each subpopulation as defined by median deconvoluted subpopulation fractions, and (ii) IHC as defined by median score or present versus absent.

STATISTICAL ANALYSIS

Statistical analyses were performed using R, Prism (GraphPad), and Excel (Microsoft) software. Details of statistical tests performed are included in figure legends. For all tests, statistical significance was defined as $P < 0.05$.

Supplementary Tables

Table S1. Patient and sample details, related to Figure 1.

Table S2. Neoplastic subpopulation marker genes, related to Figures 2-4 and Table 1.

Table S3. Neoplastic subpopulation GOTERM enrichments, related to Figure 2 and Table 1.

Table S4. Neoplastic subpopulation inferred transcription factor activity, related to Figure 2 and Table 1.

Table S5. Survival analyses of deconvoluted MAGIC cohort subpopulation proportions, related to Figure 2-5.

Table S6. Correlation matrix for deconvoluted MAGIC cohort subpopulation proportions, related to Figure 2.

Table S7. Immune subpopulation marker genes, related to Figure 5.

Table S8. GEM subpopulation marker genes, related to Figure 6.

References

Boon, K., Caron, H.N., van Asperen, R., Valentijn, L., Hermus, M.C., van Sluis, P., Roobeek, I., Weis, I., Voute, P.A., Schwab, M.*, et al.* (2001). N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. EMBO J *20*, 1383-1393.

Castellino, R.C., Barwick, B.G., Schniederjan, M., Buss, M.C., Becher, O., Hambardzumyan, D., Macdonald, T.J., Brat, D.J., and Durden, D.L. (2010). Heterozygosity for Pten promotes tumorigenesis in a mouse model of medulloblastoma. PLoS One *5*, e10849.

Cavalli, F.M.G., Remke, M., Rampasek, L., Peacock, J., Shih, D.J.H., Luu, B., Garzia, L., Torchia, J., Nor, C., Morrissy, A.S.*, et al.* (2017). Intertumoral Heterogeneity within Medulloblastoma Subgroups. Cancer Cell *31*, 737-754 e736.

Cheng, Y., Liao, S., Xu, G., Hu, J., Guo, D., Du, F., Contreras, A., Cai, K.Q., Peri, S., Wang, Y.*, et al.* (2020). NeuroD1 Dictates Tumor Cell Differentiation in Medulloblastoma. Cell Rep *31*, 107782.

Cho, Y.J., Tsherniak, A., Tamayo, P., Santagata, S., Ligon, A., Greulich, H., Berhoukim, R., Amani, V., Goumnerova, L., Eberhart, C.G.*, et al.* (2011). Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *29*, 1424-1430.

Eberhart, C.G., Kaufman, W.E., Tihan, T., and Burger, P.C. (2001). Apoptosis, neuronal maturation, and neurotrophin expression within medulloblastoma nodules. J Neuropathol Exp Neurol *60*, 462-469.

Fourgeaud, L., Traves, P.G., Tufail, Y., Leal-Bailey, H., Lew, E.D., Burrola, P.G., Callaway, P., Zagorska, A., Rothlin, C.V., Nimmerjahn, A.*, et al.* (2016). TAM receptors regulate multiple features of microglial physiology. Nature *532*, 240-244.

Garancher, A., Lin, C.Y., Morabito, M., Richer, W., Rocques, N., Larcher, M., Bihannic, L., Smith, K., Miquel, C., Leboucher, S.*, et al.* (2018). NRL and CRX Define Photoreceptor Identity and Reveal Subgroup-Specific Dependencies in Medulloblastoma. Cancer Cell *33*, 435-449 e436.

Gillen, A.E., Riemondy, K.A., Amani, V., Griesinger, A.M., Gilani, A., Venkataraman, S., Madhavan, K., Prince, E., Sanford, B., Hankinson, T.C.*, et al.* (2020). Single-Cell RNA Sequencing of Childhood Ependymoma Reveals Neoplastic Cell Subpopulations That Impact Molecular Classification and Etiology. Cell Rep *32*, 108023.

Grandori, C., Gomez-Roman, N., Felton-Edkins, Z.A., Ngouenet, C., Galloway, D.A., Eisenman, R.N., and White, R.J. (2005). c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. Nat Cell Biol *7*, 311-318.

Griesinger, A.M., Birks, D.K., Donson, A.M., Amani, V., Hoffman, L.M., Waziri, A., Wang, M., Handler, M.H., and Foreman, N.K. (2013). Characterization of distinct immunophenotypes across pediatric brain tumor types. J Immunol *191*, 4880-4888.

Hannan, A.J., Henke, R.C., Weinberger, R.P., Sentry, J.W., and Jeffrey, P.L. (1996). Differential induction and intracellular localization of SCG10 messenger RNA is associated with neuronal differentiation. Neuroscience *72*, 889-900.

Hovestadt, V., Smith, K.S., Bihannic, L., Filbin, M.G., Shaw, M.L., Baumgartner, A., DeWitt, J.C., Groves, A., Mayr, L., Weisman, H.R.*, et al.* (2019). Resolving medulloblastoma cellular architecture by single-cell genomics. Nature *572*, 74-79.

Kool, M., Koster, J., Bunt, J., Hasselt, N.E., Lakeman, A., van Sluis, P., Troost, D., Meeteren, N.S., Caron, H.N., Cloos, J.*, et al.* (2008). Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. PLoS One *3*, e3088.

Korf, H.W., Czerwionka, M., Reiner, J., Schachenmayr, W., Schalken, J.J., de Grip, W., and Gery, I. (1987). Immunocytochemical evidence of molecular photoreceptor markers in cerebellar medulloblastomas. Cancer *60*, 1763-1766.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019a). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods *16*, 1289-1296.
Korsunsky, I., Nathan, A., Millard, N., and Raychaudhuri, S. (2019b). Presto scales Wilcoxon and auROC analyses to millions of observations. BioRxiv.
Laughney, A.M., Hu, J., Campbell, N.R., Bakhoum, S.F., Setty, M., Lavallee, V.P., Xie, Y., Masilionis, I., Carr, A.J., Kottapalli, S.*, et al.* (2020). Regenerative lineages and immune-mediated pruning in lung cancer metastasis. Nat Med *26*, 259-269.
Lee, C., Rudneva, V.A., Erkek, S., Zapatka, M., Chau, L.Q., Tacheva-Grigorova, S.K., Garancher, A., Rusert, J.M., Aksoy, O., Lea, R.*, et al.* (2019). Lsd1 as a therapeutic target in Gfi1-activated medulloblastoma. Nat Commun *10*, 332.
Li, Q., and Barres, B.A. (2018). Microglia and macrophages in brain homeostasis and disease. Nat Rev Immunol *18*, 225-242.
Mao, J., Ligon, K.L., Rakhlin, E.Y., Thayer, S.P., Bronson, R.T., Rowitch, D., and McMahon, A.P. (2006). A novel somatic mouse model to survey tumorigenic potential applied to the Hedgehog pathway. Cancer Res *66*, 10171-10178.
Margol, A.S., Robison, N.J., Gnanachandran, J., Hung, L.T., Kennedy, R.J., Vali, M., Dhall, G., Finlay, J.L., Erdreich-Epstein, A., Krieger, M.D.*, et al.* (2015). Tumor-associated macrophages in SHH subgroup of medulloblastomas. Clinical cancer research : an official journal of the American Association for Cancer Research *21*, 1457-1465.
Maximov, V., Chen, Z., Wei, Y., Robinson, M.H., Herting, C.J., Shanmugam, N.S., Rudneva, V.A., Goldsmith, K.C., MacDonald, T.J., Northcott, P.A.*, et al.* (2019). Tumour-associated macrophages exhibit anti-tumoural properties in Sonic Hedgehog medulloblastoma. Nat Commun *10*, 2410.
Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat Methods *12*, 453-457.
Northcott, P.A., Buchhalter, I., Morrissy, A.S., Hovestadt, V., Weischenfeldt, J., Ehrenberger, T., Grobner, S., Segura-Wang, M., Zichner, T., Rudneva, V.A.*, et al.* (2017). The whole-genome landscape of medulloblastoma subtypes. Nature *547*, 311-317.
Northcott, P.A., Korshunov, A., Witt, H., Hielscher, T., Eberhart, C.G., Mack, S., Bouffet, E., Clifford, S.C., Hawkins, C.E., French, P.*, et al.* (2011). Medulloblastoma comprises four distinct molecular variants. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *29*, 1408-1414.
Northcott, P.A., Lee, C., Zichner, T., Stutz, A.M., Erkek, S., Kawauchi, D., Shih, D.J., Hovestadt, V., Zapatka, M., Sturm, D.*, et al.* (2014). Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature *511*, 428-434.
Northcott, P.A., Shih, D.J., Remke, M., Cho, Y.J., Kool, M., Hawkins, C., Eberhart, C.G., Dubuc, A., Guettouche, T., Cardentey, Y.*, et al.* (2012). Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta neuropathologica *123*, 615-626.
Ocasio, J., Babcock, B., Malawsky, D., Weir, S.J., Loo, L., Simon, J.M., Zylka, M.J., Hwang, D., Dismuke, T., Sokolsky, M.*, et al.* (2019a). scRNA-seq in medulloblastoma shows cellular heterogeneity and lineage expansion support resistance to SHH inhibitor therapy. Nat Commun *10*, 5829.
Ocasio, J.K., Bates, R.D.P., Rapp, C.D., and Gershon, T.R. (2019b). GSK-3 modulates SHH-driven proliferation in postnatal cerebellar neurogenesis and medulloblastoma. Development *146*.
Pei, Y., Liu, K.W., Wang, J., Garancher, A., Tao, R., Esparza, L.A., Maier, D.L., Udaka, Y.T., Murad, N., Morrissy, S.*, et al.* (2016). HDAC and PI3K Antagonists Cooperate to Inhibit Growth of MYC-Driven Medulloblastoma. Cancer cell *29*, 311-323.
Pei, Y., Moore, C.E., Wang, J., Tewari, A.K., Eroshkin, A., Cho, Y.J., Witt, H., Korshunov, A., Read, T.A., Sun, J.L.*, et al.* (2012). An animal model of MYC-driven medulloblastoma. Cancer Cell *21*, 155-167.

Poschl, J., Bartels, M., Ohli, J., Bianchi, E., Kuteykin-Teplyakov, K., Grammel, D., Ahlfeld, J., and Schuller, U. (2014). Wnt/beta-catenin signaling inhibits the Shh pathway and impairs tumor growth in Shh-dependent medulloblastoma. Acta Neuropathol *127*, 605-607.
Schafer, D.P., and Stevens, B. (2015). Microglia Function in Central Nervous System Development and Plasticity. Cold Spring Harb Perspect Biol *7*, a020545.
Sharma, T., Schwalbe, E.C., Williamson, D., Sill, M., Hovestadt, V., Mynarek, M., Rutkowski, S., Robinson, G.W., Gajjar, A., Cavalli, F.*, et al.* (2019). Second-generation molecular subgrouping of medulloblastoma: an international meta-analysis of Group 3 and Group 4 subtypes. Acta Neuropathol *138*, 309-326.
Stein, R., Mori, N., Matthews, K., Lo, L.C., and Anderson, D.J. (1988). The NGF-inducible SCG10 mRNA encodes a novel membrane-bound protein present in growth cones and abundant in developing neurons. Neuron *1*, 463-476.
Stevens, B., Allen, N.J., Vazquez, L.E., Howell, G.R., Christopherson, K.S., Nouri, N., Micheva, K.D., Mehalow, A.K., Huberman, A.D., Stafford, B.*, et al.* (2007). The classical complement cascade mediates CNS synapse elimination. Cell *131*, 1164-1178.
Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G.*, et al.* (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189-196.
Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q.*, et al.* (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc *15*, 2247-2276.
van Riggelen, J., Yetil, A., and Felsher, D.W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. Nat Rev Cancer *10*, 301-309.
Vladoiu, M.C., El-Hamamy, I., Donovan, L.K., Farooq, H., Holgado, B.L., Sundaravadanam, Y., Ramaswamy, V., Hendrikse, L.D., Kumar, S., Mack, S.C.*, et al.* (2019). Childhood cerebellar tumours mirror conserved fetal transcriptional programs. Nature *572*, 67-73.
Waszak, S.M., Robinson, G.W., Gudenas, B.L., Smith, K.S., Forget, A., Kojic, M., Garcia-Lopez, J., Hadley, J., Hamilton, K.V., Indersie, E.*, et al.* (2020). Germline Elongator mutations in Sonic Hedgehog medulloblastoma. Nature *580*, 396-401.
Wohnhaas, C.T., Leparc, G.G., Fernandez-Albert, F., Kind, D., Gantner, F., Viollet, C., Hildebrandt, T., and Baum, P. (2019). DMSO cryopreservation is the method of choice to preserve cells for droplet-based single-cell RNA sequencing. Sci Rep *9*, 10699.

Table 1

program	GP3	GP4	SHH	
A-cell cycle	**GP3-A**	**GP4-A1**	**SHH-A1**	
	cell cycle	cell cycle, G2M,	cell cycle, G2M	GOterm
	MYBL2	FOXM1, MYBL2	MYBL2	inferred TF activity
	CENPF, TOP2A, TUBA1B, HMGB2, MKI67	TOP2A, HMGB2, CENPF, TUBA1B, MKI67	CENPF, TOP2A, HMGB2, MKI67	marker genes
	12.5%	7.5%	12.1%	proportion
		GP4-A2	**SHH-A2**	
		cell cycle, S phase	cell cycle: S phase	GOterm
		ZNF492, E2Fs	E2Fs	inferred TF activity
		HELLS, XRCC2, ATAD5, BRCA1	RRM2, POLQ, FANCA, XRCC2, BRCA1, HELLS	marker genes
		4.6%	10%	proportion
B- progenitor	**GP3-B1**	**GP4-B1**	**SHH-B1**	
	translation/oxidative respiration	oxidative phosphorylation/ translation	translation/oxidative respiration/neg. reg. of epithelial cell differentiation	GOterm
	HMGA1, MYC, MAX	BARHL1, NEUROD2, SOX4, JUNB, JUND	ATOH1, ZEB1, HES6	inferred TF activity
	ribosomal	STMN2, ribosomal	SFRP1, CXCR2, ribosomal	marker genes
	33.3%	25.9%	28.7%	proportion
	GP3-B2	**GP4-B2**	**SHH-B2**	
	translation/oxidative respiration	translation	translation/oxidative respiration	GOterm
	HLX, ATF4, MYC	HEY1, TEAD2, ELK1	ELK1, THAP11, MAX	inferred TF activity
	ribosomal, MYC	STMN2, ribosomal	ribosomal	marker genes
	1.9%	17.3%	17.1%	proportion
C-neuronally differentiated	**GP3-C1**	**GP4-C1**	**SHH-C1**	
	RNA processing/ axo-dendritic transport	RNA processing/ axo-dendritic transport	RNA processing/ axo-dendritic transport	GOterm
	ARNT	ONECUT2, LHX4	POU2F2, KMT2A	inferred TF activity
	LUC7L3, DST	GRIA2, LUC7L3, DST	GRIA2, LUC7L3, DST	marker genes
	37.6%	32.7%	28.2%	proportion
	GP3-C2	**GP4-C2**	**SHH-C2**	
	photoreceptor	photoreceptor	axon development	GOterm
	SHOX2, CRX	MEIS1, CRX	SOX4, NEUROD1	inferred TF activity
	NRL, IMPG2	NRL, IMPG2	STMN4, STMN2	marker genes
	14.6%	6.9%	3.6%	proportion

Figure 1

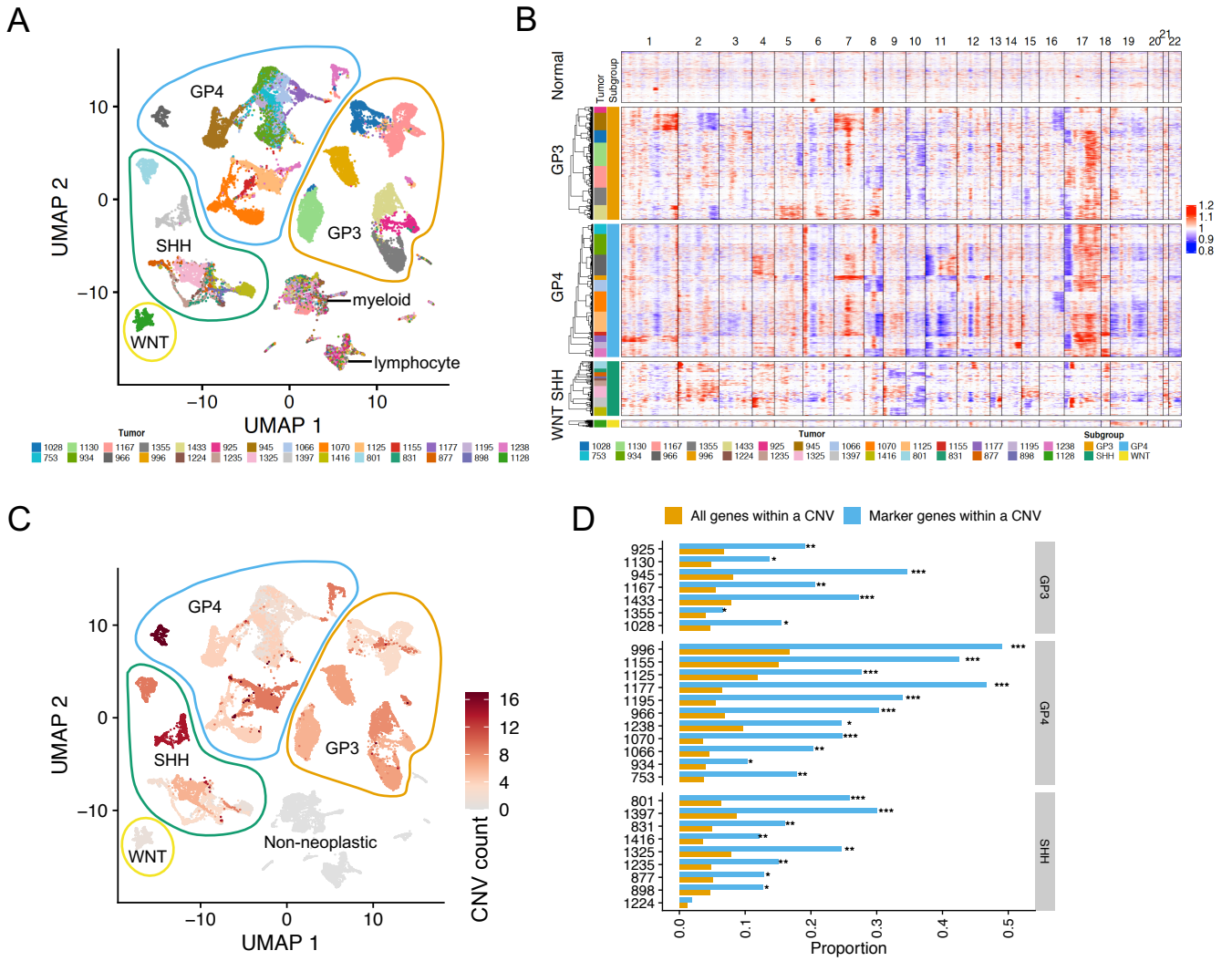

A
GP4
GP3
SHH
WNT
myeloid
lymphocyte
UMAP 2
UMAP 1
Tumor
B
Normal
GP3
GP4
WNT SHH
Tumor
Subgroup
C
GP4
GP3
SHH
WNT
Non-neoplastic
CNV count
D
All genes within a CNV
Marker genes within a CNV
GP3
GP4
SHH
Proportion

Figure 2

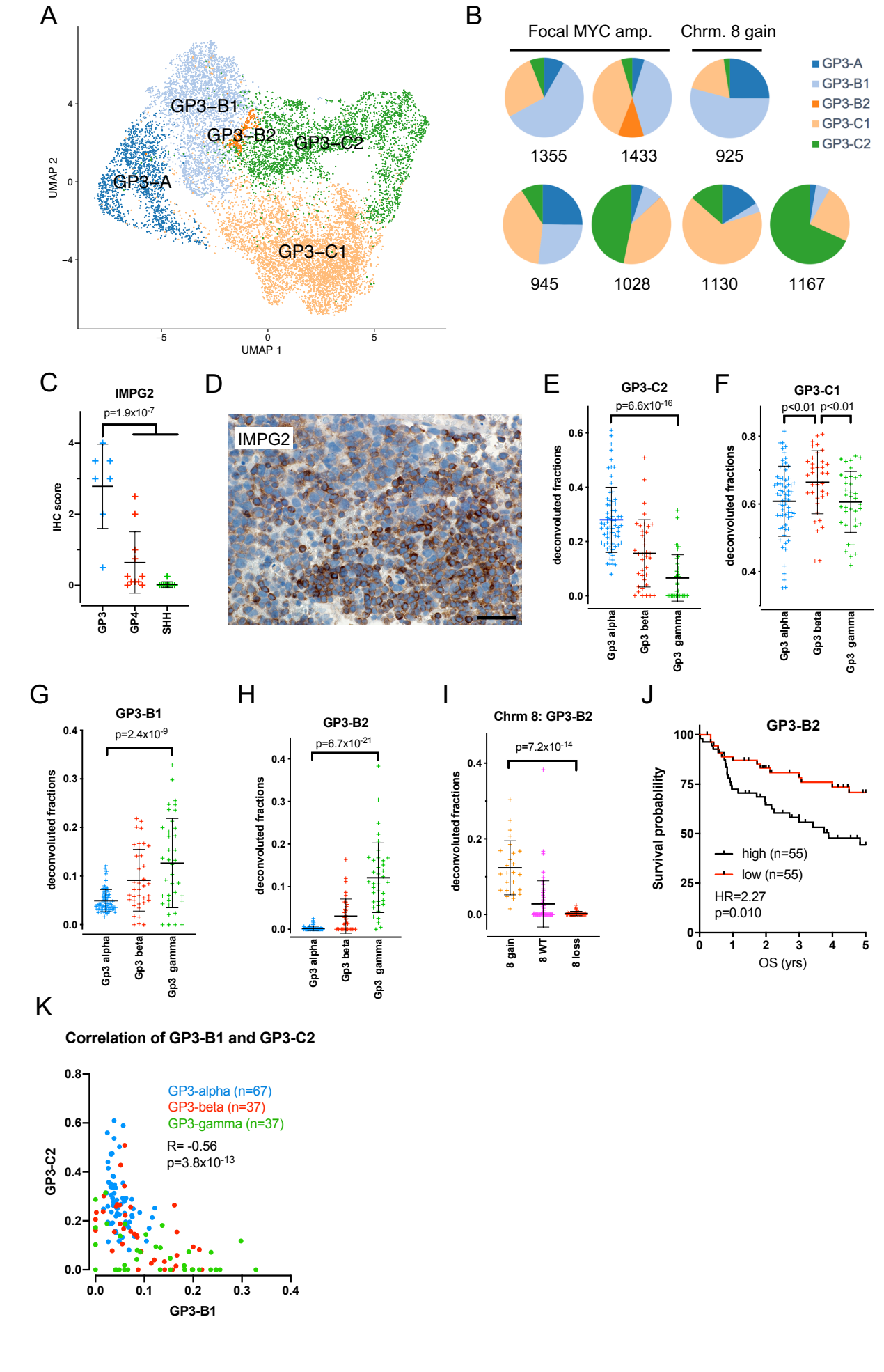

A
GP3–B1
GP3–B2
GP3–C2
GP3–A
GP3–C1
UMAP 2
UMAP 1
B
Focal MYC amp.
Chrm. 8 gain
GP3-A
GP3-B1
GP3-B2
GP3-C1
GP3-C2
1355
1433
925
945
1028
1130
1167
C
IMPG2
p=1.9x10-7
IHC score
GP3
GP4
SHH
D
IMPG2
E
GP3-C2
p=6.6x10-16
deconvoluted fractions
Gp3 alpha
Gp3 beta
Gp3 gamma
F
GP3-C1
p<0.01
p<0.01
deconvoluted fractions
Gp3 alpha
Gp3 beta
Gp3 gamma
G
GP3-B1
p=2.4x10-9
deconvoluted fractions
Gp3 alpha
Gp3 beta
Gp3 gamma
H
GP3-B2
p=6.7x10-21
deconvoluted fractions
Gp3 alpha
Gp3 beta
Gp3 gamma
I
Chrm 8: GP3-B2
p=7.2x10-14
deconvoluted fractions
8 gain
8 WT
8 loss
J
GP3-B2
Survival probability
high (n=55)
low (n=55)
HR=2.27
p=0.010
OS (yrs)
K
Correlation of GP3-B1 and GP3-C2
GP3-alpha (n=67)
GP3-beta (n=37)
GP3-gamma (n=37)
R= -0.56
p=3.8x10-13
GP3-C2
GP3-B1

Figure 3

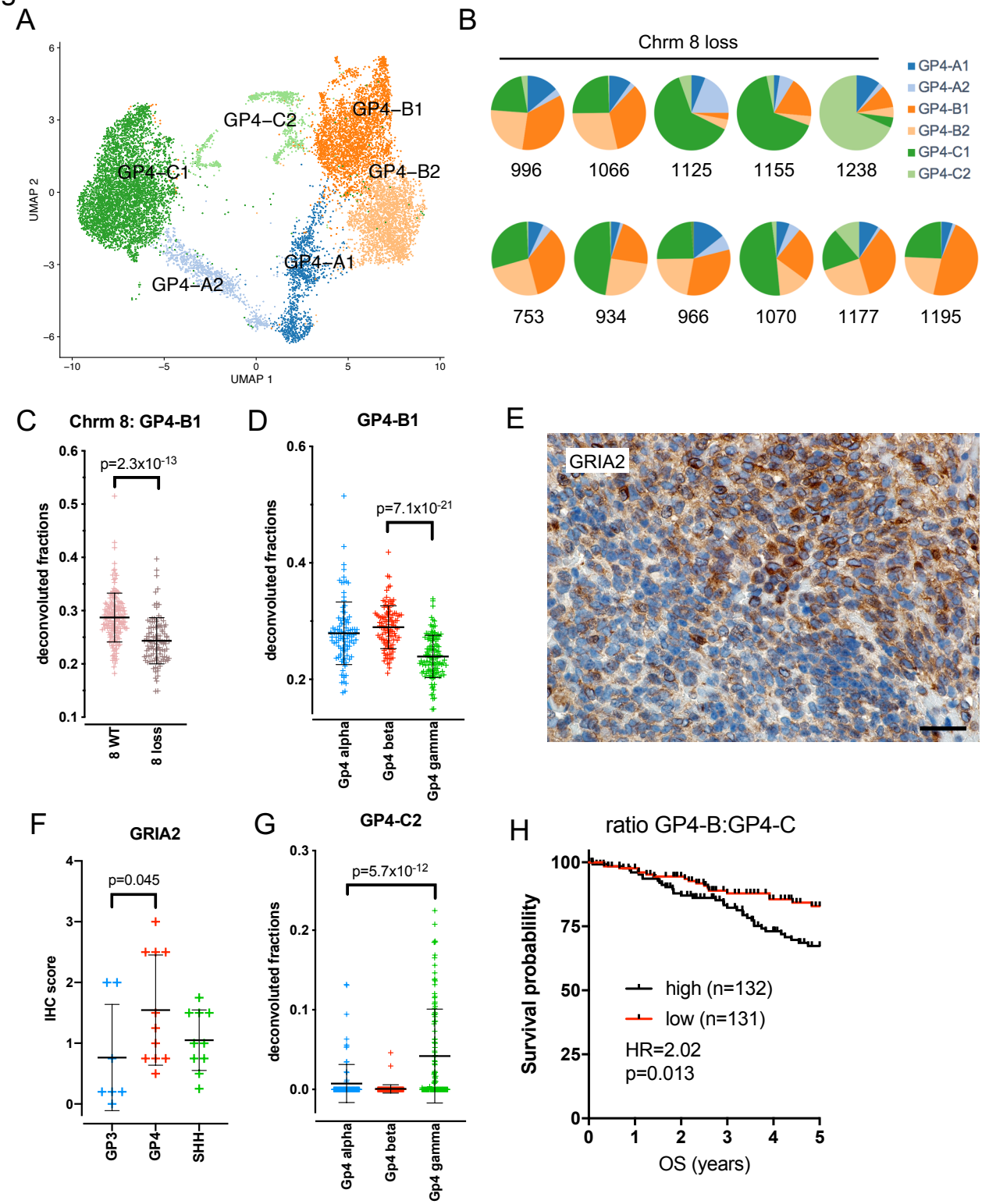

A
GP4–C1
GP4–C2
GP4–B1
GP4–B2
GP4–A1
GP4–A2
UMAP 1
UMAP 2
B
Chrm 8 loss
GP4-A1
GP4-A2
GP4-B1
GP4-B2
GP4-C1
GP4-C2
996
1066
1125
1155
1238
753
934
966
1070
1177
1195
C
Chrm 8: GP4-B1
p=2.3x10-13
deconvoluted fractions
8 WT
8 loss
D
GP4-B1
p=7.1x10-21
deconvoluted fractions
Gp4 alpha
Gp4 beta
Gp4 gamma
E
GRIA2
F
GRIA2
p=0.045
IHC score
GP3
GP4
SHH
G
GP4-C2
p=5.7x10-12
deconvoluted fractions
Gp4 alpha
Gp4 beta
Gp4 gamma
H
ratio GP4-B:GP4-C
Survival probability
high (n=132)
low (n=131)
HR=2.02
p=0.013
OS (years)

Figure 4

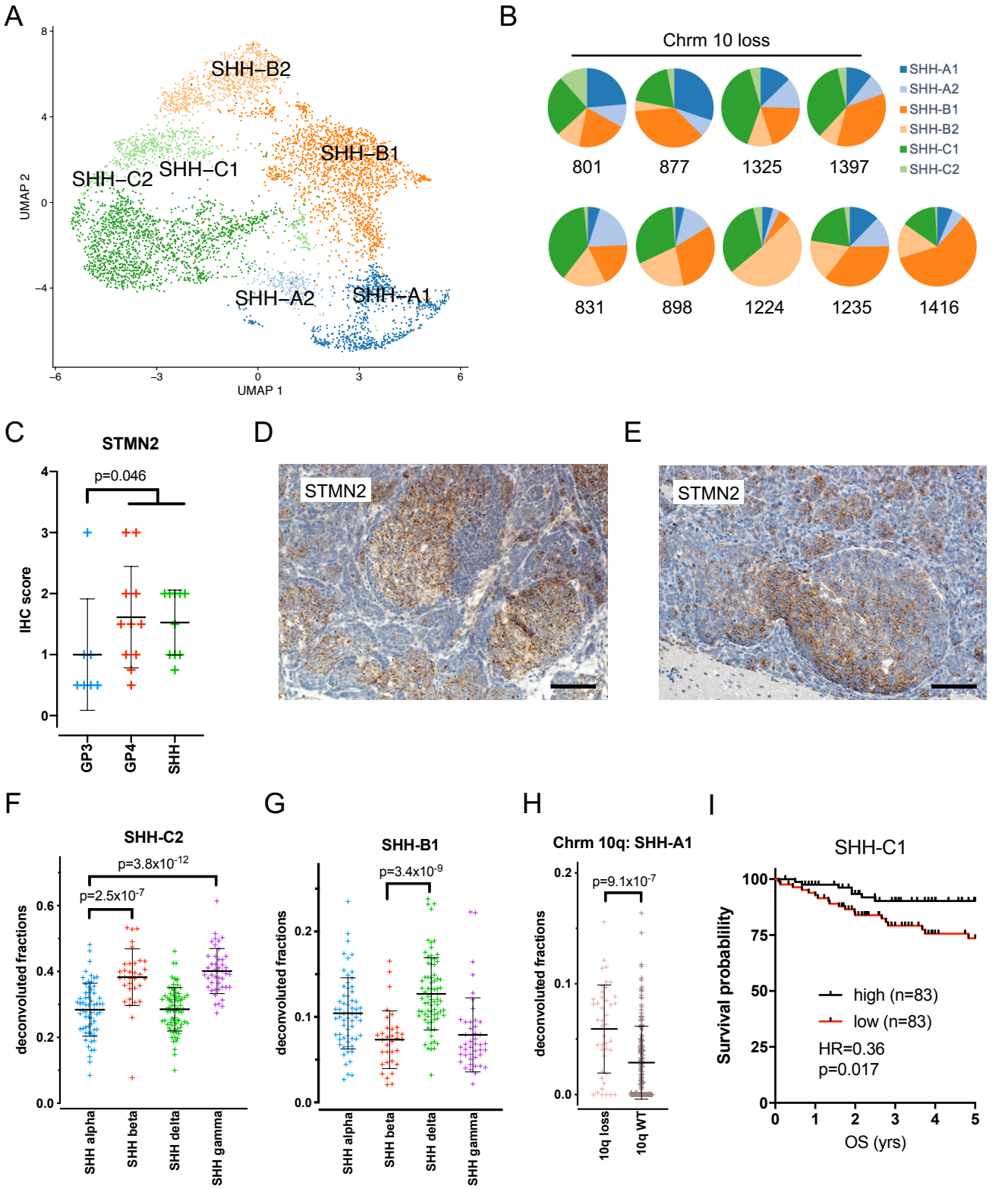
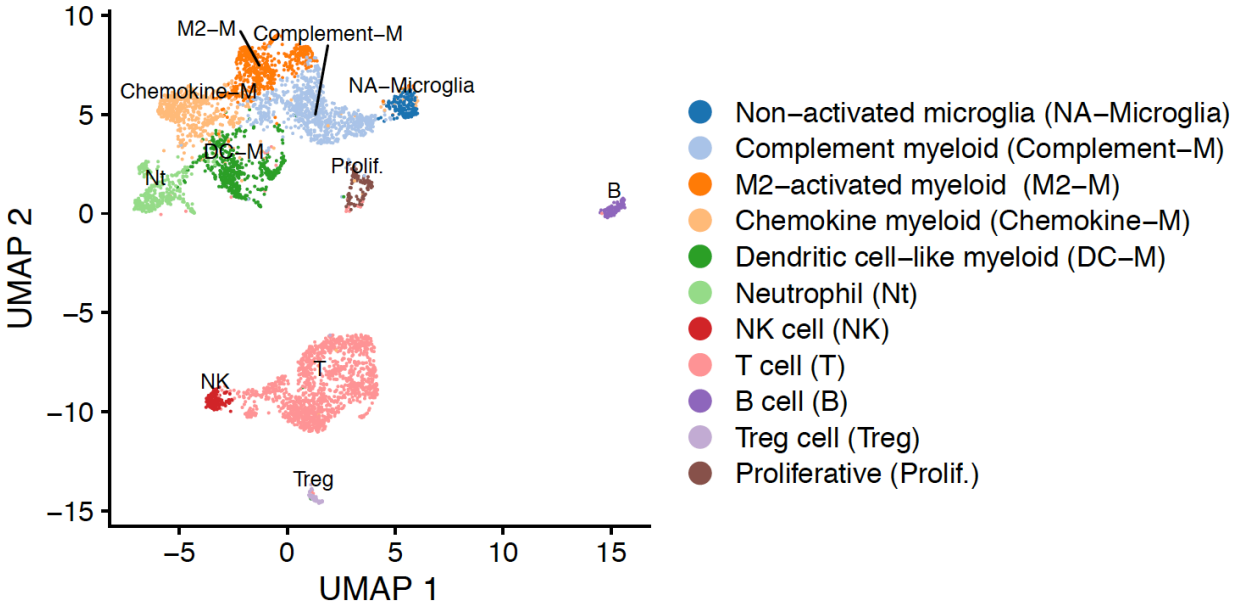
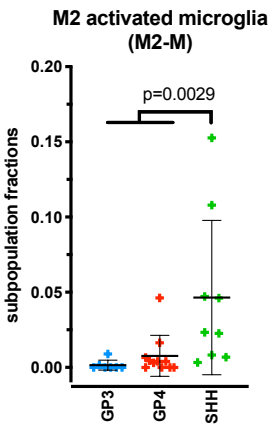

A
SHH–B2
SHH–B1
SHH–C1
SHH–C2
SHH–A2
SHH–A1
UMAP 2
UMAP 1
B
Chrm 10 loss
SHH-A1
SHH-A2
SHH-B1
SHH-B2
SHH-C1
SHH-C2
801
877
1325
1397
831
898
1224
1235
1416
C
STMN2
p=0.046
IHC score
GP3
GP4
SHH
D
STMN2
E
STMN2
F
SHH-C2
p=3.8x10-12
p=2.5x10-7
deconvoluted fractions
SHH alpha
SHH beta
SHH delta
SHH gamma
G
SHH-B1
p=3.4x10-9
deconvoluted fractions
SHH alpha
SHH beta
SHH delta
SHH gamma
H
Chrm 10q: SHH-A1
p=9.1x10-7
deconvoluted fractions
10q loss
10q WT
I
SHH-C1
Survival probability
high (n=83)
low (n=83)
HR=0.36
p=0.017
OS (yrs)

Figure 5

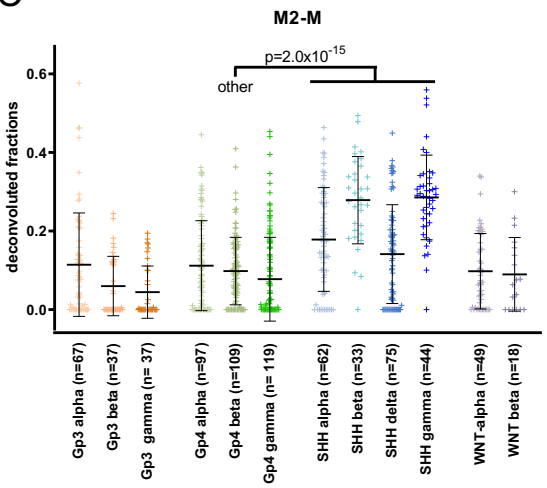
A



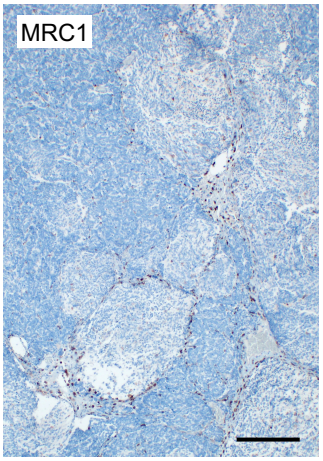
B



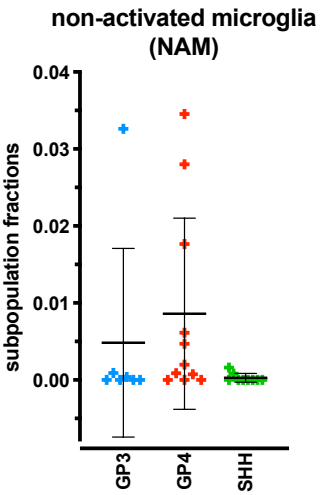
C



D



E



F

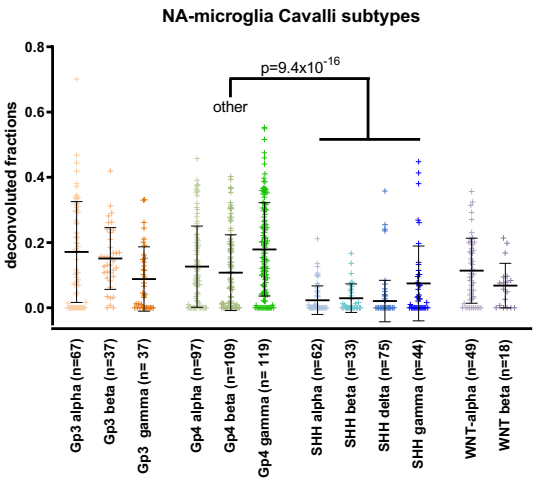
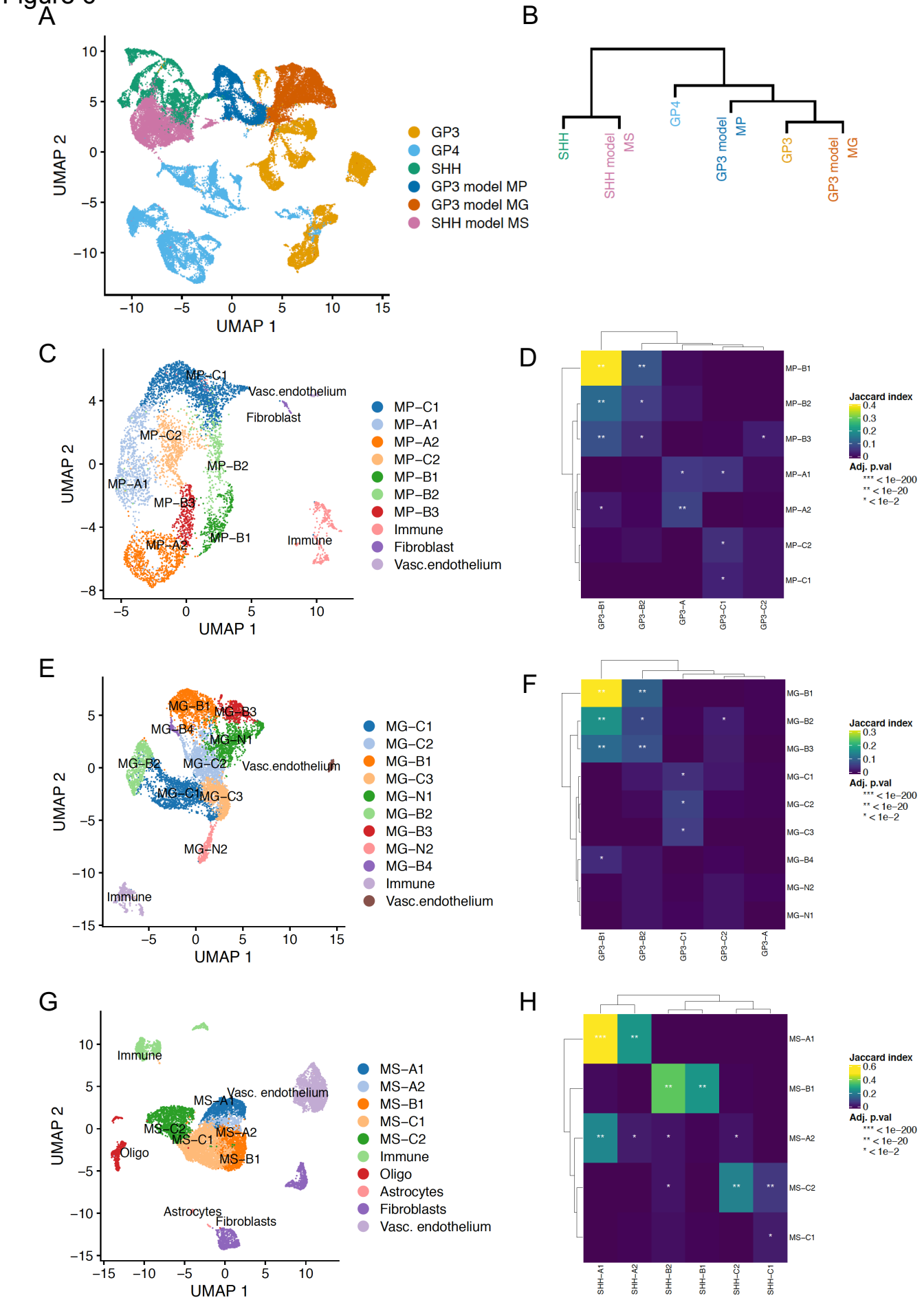


Figure 6


A
UMAP 2
UMAP 1
GP3
GP4
SHH
GP3 model MP
GP3 model MG
SHH model MS
B
SHH
SHH model
MS
GP4
GP3 model
MP
GP3
GP3 model
MG
C
MP–C1
Vasc.endothelium
Fibroblast
MP–C2
MP–B2
MP–A1
MP–B3
MP–A2
MP–B1
Immune
MP–C1
MP–A1
MP–A2
MP–C2
MP–B1
MP–B2
MP–B3
Immune
Fibroblast
Vasc.endothelium
D
MP–B1
MP–B2
MP–B3
MP–A1
MP–A2
MP–C2
MP–C1
GP3–B1
GP3–B2
GP3–A
GP3–C1
GP3–C2
Jaccard index
0.4
0.3
0.2
0.1
0
Adj. p.val
*** < 1e–200
** < 1e–20
* < 1e–2
E
MG–B1
MG–B3
MG–B4
MG–N1
MG–B2
MG–C2
Vasc.endothelium
MG–C1
MG–C3
MG–N2
Immune
MG–C1
MG–C2
MG–B1
MG–C3
MG–N1
MG–B2
MG–B3
MG–N2
MG–B4
Immune
Vasc.endothelium
F
MG–B1
MG–B2
MG–B3
MG–C1
MG–C2
MG–C3
MG–B4
MG–N2
MG–N1
GP3–B1
GP3–B2
GP3–C1
GP3–C2
GP3–A
Jaccard index
0.3
0.2
0.1
0
Adj. p.val
*** < 1e–200
** < 1e–20
* < 1e–2
G
Immune
Vasc. endothelium
MS–A1
MS–C2
MS–A2
MS–C1
Oligo
MS–B1
Astrocytes
Fibroblasts
MS–A1
MS–A2
MS–B1
MS–C1
MS–C2
Immune
Oligo
Astrocytes
Fibroblasts
Vasc. endothelium
H
MS–A1
MS–B1
MS–A2
MS–C2
MS–C1
SHH–A1
SHH–A2
SHH–B2
SHH–B1
SHH–C2
SHH–C1
Jaccard index
0.6
0.4
0.2
0
Adj. p.val
*** < 1e–200
** < 1e–20
* < 1e–2