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

Kent A. Riemondy[1], Sujatha Venkataraman[2,5], Nicholas Willard[3], Anandani Nellan[2,5], Bridget Sanford[2], Andrea M. Griesinger[2,5], Vladimir Amani[2,5], Siddhartha Mitra[2,5], Todd C. Hankinson[4,5], Michael H. Handler[4,5], 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[2,4,5], Andrew M. Donson[2,5,9,*], Rajeev Vibhakar[2,5,9].

[1]RNA Bioscience Initiative, [2]Department of Pediatrics, [3]Department of Pathology, and [4]Department of Neurosurgery, University of Colorado Anschutz Medical Campus, Aurora, Colorado, 80045, USA.

[5]Morgan Adams Foundation Pediatric Brain Tumor Research Program, Children's Hospital Colorado, Aurora, Colorado, 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, University of Colorado Anschutz Medical Campus, Department of Pediatrics, Mail Stop # 8302, Research Complex 1, North Tower, 12800 E. 19th Ave., Rm P18-4402A, Aurora, CO 80045. E-mail: Andrew.donson@cuanschutz.edu

Secondary contact: Rajeev Vibhakar, University of Colorado Anschutz Medical Campus, Department of Pediatrics, Mail Stop # 8302, Research Complex 1, North Tower, 12800 E. 19th Ave., Rm P18-4402A, Aurora, CO 80045. E-mail: Rajeev.vibhakar@cuanschutz.edu

Conflict of Interests

The authors have declared that no conflict of interest exists.

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 use single-cell RNA sequencing, immunohistochemistry and deconvolution of transcriptomic data to profile neoplastic and immune populations in childhood MB samples. 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, including discrete photoreceptor-like cells in MB subgroups GP3 and GP4 and nodule-associated neuronally-differentiated cells in subgroup SHH. We definitively chart the spectrum of MB immune cell infiltrates, which include subgroup/subtype-

associated developmentally-related neuron-pruning and antigen presenting myeloid cells. MB cellular diversity is recapitulated in genetically engineered mouse subgroup-specific models of MB. These findings provide a clearer understanding of both the neoplastic and immune cell heterogeneity in MB.

Introduction

Medulloblastoma (MB) is an aggressive brain tumor arising predominantly in childhood, comprised of 4 consensus molecular subgroups – GP3, GP4, SHH and WNT – that have been defined through bulk-tumor transcriptomic, methylomic and genomic profiling studies (1). 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. Subsequent studies have subdivided MB subgroups into finer subtypes with disparate outcomes and molecular features (2-4). Unfortunately, intra-tumoral cellular heterogeneity inherent in bulk-tumor samples impedes a clearer understanding of MB biology. Neoplastic and immune heterogeneity underlies a tumor's ability to proliferate, survive and evade therapeutic interventions. The emergence of single cell analysis now allows us to examine the cellular diversity inherent in biological systems, resulting in transformative findings when applied to pediatric brain tumors including MB. Hovestadt *et al.* used SMART-seq single cell RNA-seq (scRNA-seq) to interrogate MB, identifying cellular diversity within individual samples recapitulating cell cycle, progenitor and differentiated neuronal programs (5). 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 MB samples, identifying putative subgroup-specific cells of origin (6). In the present study, we add significant new knowledge by increasing the number of samples and cells per tumor and by adding MB genetrically-engineered mouse models (GEMMs) for cross-species analysis. .

Our studies identify and characterize more granular differentiation trajectories in detail, especially previously unnoted discrete photoreceptor and glutamatergic subpopulations across GP3 and GP4 tumors. Importantly we describe for the first time considerable immune cell diversity in MB in detail, including subgroup-specific immune populations. In preclinical models, we identify neoplastic subgroup-specific genetically-engineered mouse model (GEMM) 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 GEMMs 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 (Supplementary Table 1). The SHH, GP3 and GP4 samples were further assigned to more recently described subtypes (2, 4). Using the 10x Genomics droplet-sequencing-based platform we generated transcriptomes at the single-cell level as previously described (7). 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 (Supplementary Fig. 1).

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 (astrocytes, oligodendrocytes and vascular endothelium) (Fig. 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)(Fig 1b). We identified numerous MB subgroup specific copy number gains and losses that were highly concordant with CNVs identified by methylation analysis of matched samples, the most common of these being isochromosome 17q in the majority of GP4 (Supplementary Table 1). Neoplastic cell clusters were distributed broadly according to MB subgroup, as determined by methylation analysis (Fig. 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 (Supplementary Table 1). GP3 demonstrated the least cohesive clustering, with all patient samples clustering separately. In GP4 5 of 11 samples colocalized. We mapped the sum of CNVs in each sample onto the UMAP which showed a that lower CNV counts were seen in those samples that colocalized within their respective subgroup (Fig. 1c). This finding suggests that sample specific clustering within MB subgroups is a function of the extent of CNV aberration (Fig. 1c). These findings underscore the significant contribution of CNVs to the transcriptional landscape in MB.

GP3 MB contains a differentiated subpopulation corresponding to photoreceptor differentiation.

Given the observed subgroup-specific clustering, SHH, GP3 and GP4 subgroup neoplastic cells were individually re-clustered. We used Harmony alignment algorithm to identify transcriptionally similar cell types that had clustered discretely in the prior non-aligned analysis (8). Harmony revealed neoplastic clusters shared between multiple samples in each of the

subgroups (Fig. 2-4). SHH, GP3 and GP4 samples each formed clusters that exhibited transcript signatures matching those clusters identified in a recent MB single cell RNA-seq study by Hovestadt *et al.* namely cell cycle (program A), undifferentiated progenitors (program B) and neuronally-differentiated (program C)(5)(Supplementary Fig. 2). We annotated clusters accordingly, but with new subdivisions within these programs corresponding to distinct subpopulations that were identified in the present study. Subpopulations were then characterized by systematic analysis of gene expression, gene ontology (GOTerm) enrichment and inference of active transcription factors (TFs) using pySCENIC (single-cell regulatory network inference and clustering)(9)(Supplementary Tables 3-5, summarized in Table 1). To address similarities of neoplastic subpopulations between MB subgroups, we also performed a combined comparative analysis (Supplementary Fig. 3a).

GP3 tumors (n=7) formed 5 major clusters that included two differentiated cell subpopulations (GP3-C2, -C1) in addition to two progenitor (GP3-B1, -B2) and a mitotic (GP3-A) subpopulation (Fig. 2a,b). Identification of neuronally-differentiated subpopulations in GP3 differs from the findings of Hovestadt *et al.* who reported that GP3 tumors with *MYC* amplification lacked differentiated cells altogether (5). The predominant GP3 differentiated subpopulation GP3-C1 is characterized by genes related to RNA processing and axo-dendritic transport (Table 1, Supplementary Tables 3,4). The less abundant GP3-C2 was distinguished by genes and TFs related to photoreceptor development (*NRL*, *IMPG2*) (Table 1, Supplementary Tables 3-5), a phenotype that has previously been identified and studied in GP3 (10-13). We validated the presence of a photoreceptor-like subpopulation in GP3 using immunohistochemistry (IHC) for interphotoreceptor matrix proteoglycan 2 (IMPG2), a marker of GP3-C2. In a panel of GP3, GP4 and SHH MB formalin-fixed paraffin-embedded (FFPE) samples, IMPG2 protein expression was shown to be significantly higher in GP3 than GP4 and SHH ($p=1.9x10^{-7}$), with an expression score of 2 or greater in 6/7 GP3 samples (Fig. 2c). IMPG2 was expressed in discrete clusters of cells in

GP3 tumors, suggesting focal areas of GP3-C2 photoreceptor differentiation (Fig. 2d). Discrete clusters of IMPG2 expressing cells were also observed in a smaller number of GP4 tumors (Fig. 2e). The patchy distribution of *IMPG2*, and presumably other photoreceptor subpopulation-restricted markers, across areas of tumor in both GP3 and to a lesser extent in GP4 could therefore introduce a confounding factor for classification techniques that employ these markers for GP3 (14)(Supplementary Fig. 4).

Subpopulation proportions were estimated by deconvolution (Cibersort) of the Medulloblastoma Advanced Genomics International Consortium (MAGIC) dataset, a large patient cohort that includes transcriptomic data, subgroup/subtype classification and associated CNV and outcome data for the majority of samples(n=747) (2). For individual MB subgroup, estimated subpopulation proportions were compared 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 significant association between photoreceptor-differentiated subpopulation GP3-C2 and subtype GP3 alpha, that has the best clinical outcome of all GP3 subtypes per Cavalli *et al.* (2). The GP3-C2 proportion was 2.5-fold higher in GP3 alpha than other GP3 subtypes ($p=2.3x10^{-14}$), being progressively less abundant in GP3 beta and GP3 gamma (Fig. 2f). GP3 differentiated subpopulation GP3-C1 was more abundant in GP3 beta than others (p=0.0027)(Fig. 2g).

GP3 progenitor proportions are influenced by MYC and chromosome 8 gain

Examination of GP3 progenitor subpopulations GP3-B1 and GP3-B2 further substantiated the previously articulated theory that aberrant *MYC* activity is responsible for preventing differentiation of GP3 progenitor subpopulations (5). 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, Supplementary Tables 3-5). 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 (15, 16). 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)(Fig. 2b). Deconvolution of the MAGIC GP3 cohort showed a significantly higher proportion of progenitor GP3-B1 and GP3-B2 cells in GP3 gamma than in either alpha or beta subtypes (Fig. 2h,i). 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 (2). 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, with GP3-B2 proportions being higher in samples with chromosome 8 gain than in samples with chromosome 8 WT or loss (Fig. 2j). 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 (Fig. 2b). With respect to clinical outcome, a higher than median proportion of GP3-B2 conferred the shortest survival of all GP3 subpopulations (Fig. 2k, Supplementary Table 6), 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 in particular GP3-B1 (Fig. 2l, Supplementary Table 7). This analysis demonstrates a continuum of differentiated and progenitor cell ratios across GP3 alpha, beta and gamma subtypes, an important factor for correct interpretation of GP3 biological studies.

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

Six major neoplastic subpopulations were identified in 11 GP4 scRNA-seq samples, with some degree of similarity with subpopulations identified in GP3 (Fig. 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, Supplementary Tables 3,4). Deconvolution analysis of GP4 MAGIC samples (n=325) revealed that the strongest association between CNVs and estimated subpopulation proportions was between GP4-B1 and chromosome 8 aberrations (Fig. 3c). This association differs from that seen with in GP3 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 also showed variable subpopulation proportions, the most distinct being GP4 gamma having significantly lower GP4-B1 progenitors ($p=3.2x10^{-18}$)(Fig. 3c).

The predominant GP4 MB differentiated subpopulation is distinguished by expression of glutamatergic neuron lineage markers

The predominant differentiated GP4 subpopulation, GP4-C1, is characterized by gene expression similar to GP3-C1, including expression of glutamate receptor *GRIA2*, a key component of glutamatergic lineage neurons (Supplementary Table 3). In recent single cell analyses of the developing human cerebellum *GRIA2* was a marker of glutamatergic lineage neurons, most notably human excitatory cerebellar nuclei (eCN)/uniopolar brush cells (UBC) (17), although also expressed at lower levels in Purkinje cells and cerebellar granular neurons

(CGNs)(18). We compared human cerebellar lineage scRNA-seq data (17) with our MB neoplastic subpopulations showing that overlap between these cerebellar cell types was largely restricted to MB differentiated subpopulations and absent from progenitor subpopulations (Supplementary Fig. 5a-c). The strongest correlation was observed between human eCN/UBC signatures and GP4-C1, and to a progressively lesser extent with SHH-C1 and GP3-C1 (Supplementary Fig. 5d). Accordingly, IHC analysis revealed that eCN/UBC marker GRIA2 positive cells were abundantly distributed throughout the tumor parenchyma in GP4 (Fig. 3e), to a moderate extent in SHH and least abundant in GP3 (Fig. 3f,g). These findings are consistent with early MB bulk-tumor transcriptomic data that revealed a glutamatergic signature in GP4 and also in GP3, but to a lesser extent (10). The same study identified a GABAergic signature that was exclusive to GP3. Although no MB neoplastic subpopulations overlapped with GABAergic lineages, GABRA5, a key marker of the GP3 GABAergic signature (10) was expressed predominantly in GP3 progenitors and GP3-C1 (Supplementary Fig. 4a).

A photoreceptor differentiated subpopulation, GP4-C2, was also variably observed in GP4, although to a lesser extent than in GP3 (Fig. 3b, Table 1). Comparison of GP4-C2 with other neoplastic subpopulations confirmed the strong correlation with GP3-C3 (Supplementary Fig. 3a). Photoreceptor-differentiated cells in GP4 was examined by IHC (Fig. 2e), with 2/11 GP4 samples expressing IMPG2 with an IHC score of 2 or more, and isolated IMPG2 positive cells in the remainder (Fig. 2c). The presence of photoreceptor-differentiated cells in GP4 may 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 (14) in the GP4-C2 subpopulation (Supplementary Fig. 4a). By deconvolution analysis, MAGIC cohort GP4 subtypes alpha, beta and gamma also showed variable differentiated subpopulation proportions, with a significantly higher photoreceptor differentiated subpopulation GP4-C2 in GP4 gamma (11-fold higher than others, $p=5.7x10^{-12}$)(Fig. 3h).

No single GP4 subpopulation was significantly correlated with survival (Supplementary Table 6), although the ratio of progenitor (GP4-B1 and -B2 combined) to differentiated (GP4-C1 and -C2 combined) conveyed a shorter survival (Fig. 3i). These findings suggest that the better clinical survival reported for GP4-gamma compared to other GP4 subtypes (2) is likely attributable to a higher ratio of differentiated to progenitor cells.

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) (Fig. 4a,b). SHH-C2 was distinct from SHH-C1 and differentiated subpopulations in other subgroups (Supplementary Fig. 3a) with a particularly high enrichment of neuron-related gene ontologies (Fig. 4c, Supplementary Table 4). SHH-C2 was distinguished from SHH-C1 by high expression of stathmin genes (*STMN2* and *STMN4*) which are involved with neuron development and growth (19). IHC showed that SHH-C2 marker STMN2 positive cells were predominant in SHH nodules (Fig. 4d,e). Nodules of neuronal maturation are a classic histologic feature of SHH that recapitulates aspects of normal cerebellar development (20). In the MAGIC cohort, Cavalli *et al.* delineated four 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 (2). We deconvoluted the MAGIC SHH transcriptomic cohort (n=214) and showed that SHH-C2 was the most differential subtype distribution being significantly more abundant in the infant SHH beta and gamma subtypes (Fig. 4f). Consistent with SHH-C2 representing the primary population of cells that constitute SHH nodules, this subpopulation was most abundant in highly nodular SHH gamma that is associated with SHH variant MBEN.

The most abundant neuronally-differentiated SHH subpopulation SHH-C1 was most closely correlated with GP4-C1 and GP3-C1 (Supplementary Fig. 3a), with enrichment of RNA processing and axo-dendritic transport ontologies (Table 1, Supplementary Tables 3,4) and overlap with human glutamatergic eCN/UBC lineage gene expression (Supplementary Fig. 5). SHH-C1 was the only SHH subpopulation significantly associated with outcome, its relative abundance being favorable for survival (Fig. 4i, Supplementary Table 6).

SHH progenitors are distinct from GP3 and GP4 by expression of previously identified components of SHH biology. These include inferred TF activity of CGN progenitor specification gene ATOH1 (Supplementary Table 5) and high expression of genes associated with negative regulation of epithelial cell proliferation and WNT signaling (*SFRP1*, *SFRP5*) (Supplementary Tables 3,4). This is consistent with studies showing that WNT pathway activation is antagonistic to SHH-driven proliferation (21, 22). 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 (Supplementary Tables 3,4). 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 (23). Related to this, germline elongation factor *ELP1* mutations have recently been implicated as a major driver in SHH etiology (24), further supporting a potential role of transcriptional/translational dysregulation in SHH progenitor biology. SHH progenitor populations, in particular SHH-B1, were significantly inversely correlated with SHH-C2 (Supplementary Table 7), being highest in non-infant subtypes SHH alpha and delta (Fig. 4g).

CNVs were not as strongly correlated with SHH subpopulations as in GP3 or GP4, the most significant being an increase (2.1-fold higher) in cell cycle subpopulation SHH-A1 in those samples with 10q loss (Fig. 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 (25). SHH-A

subpopulations, and GP3-A and GP4-A, were enriched for cells in both G2M and S-phase, although we note that other subpopulations contain proliferating cells to a lesser extent (Supplementary Fig. 3b).

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

ScRNA-seq provided us with an opportunity to comprehensively chart tumor infiltrating immune cell subpopulations in MB. 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 (Fig. 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 (Fig. 5a). This finding is consistent with other scRNA-seq analyses of tumor infiltrating and CNS immune subpopulations that showed a greater degree of diversity than was previously assumed (26, 27). Despite apparent transcriptional overlap between subpopulations identified in these recent studies no consensus nomenclature yet exists. Accordingly, immune subpopulations described in the present populations are annotated with names descriptive of transcriptional signatures. 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 (28). 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 (Supplementary Table 8) and cannot for the most part reliably assign subpopulations to resident or peripheral myeloid origins.

The most abundant myeloid subpopulation exhibited signatures associated with neurodevelopmental roles. Specifically this subpopulation demonstrated particularly high expression of all complement component 1q subunits (*C1QA*, *C1QB* and *C1QC*) (Supplementary Table 8), that are responsible for the initiation of the classic complement cascade that is involved in neurodevelopmental synapse elimination (29). 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 (30). *MRC1* and *CD163* are widely recognized markers of anti-inflammatory or immunoregulatory M2 myeloid polarization and this population was therefore named M2-activated myeloid (M2-M). 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 (Fig. 5b). Myeloid cell scRNA-seq subpopulation signatures were used to deconvolute the MAGIC dataset (GP3, GP4, SHH and WNT; n=748) which corroborated the observation that M2-M were more abundant in SHH that other subgroups (2.2-fold; $p=2.0x10^{-15}$) (Fig. 5c). Further, M2-M abundance was significantly higher in infant subtypes SHH beta and gamma than older subtypes SHH alpha and delta (1.8-fold; $p=3.7x10^{-6}$)(Fig. 5c). IHC for M2-M marker *MRC1* (CD206), stained cells with an amoeboid morphology that frequently lined SHH nodules (Fig. 5d). The co-localization of M2-M with nodules, which are comprised of the strongly 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*)(Supplementary Table 8). 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 and were significantly less abundant in SHH in the larger MAGIC cohort (4-fold, $p=9.4x10^{-16}$)(Fig. 5e). These findings are congruent with prior studies that identified a less activated immune phenotype in GP3 and GP4 than SHH (31, 32).

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*)(Supplementary Table 8), suggesting immune cell recruitment and antigen presentation functions, respectively, and is referred to as chemokine myeloid (Chemokine-M). 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 dendritic cell-like myeloid (DC-M). 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 (Supplementary Table 6). 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 putative 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 (33). 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 (Fig. 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 differentially distributed across MB subgroups in the scRNA-seq cohort. 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 (GEMMs) 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 GEMM experimental results. Two GP3 models, the first (termed MP) driven by overexpression of cMyc and dominant negative *Trp53* (34, 35), and the second driven by co-expression of Myc and Gfi1 (and termed MG)(36, 37), and one SHH model driven by mutant Smo activated in the Atoh1 lineage (Math1-Cre/SmoM2 - termed MS)(38) were examined using scRNA-seq. Harmony was used to align GEMM with human MB subgroup scRNA-seq data, revealing that GEMM cells clustered most closely their intended corresponding human MB subgroups (Fig. 6a,b).

Both GP3 GEMMs were generated by retroviral overexpression of oncogenes in neural progenitors, and transplantation of these progenitors into the cerebellum of adult NSG mice. In model MP, scRNA-seq of 5,422 cells identified 7 neoplastic and 2 non-neoplastic subpopulations

(immune and fibroblast) (Fig. 6c). The 7 neoplastic GEMM subpopulations were compared to human GP3 neoplastic subpopulations using three methods: (i) comparison of the top 200 marker genes for GEMM neoplastic subpopulations (Supplementary Table 9) compared to corresponding subgroup subpopulations (Fig. 6d) as used previously to compare neoplastic subpopulations between subgroups; (ii) UMAPs of GEMM cells labeled with human neoplastic subpopulation metagene signatures (Supplementary Fig. 6a); and (iii) correlation of the genome-wide mean expression of each GEMM neoplastic subpopulation with human neoplastic subpopulations which are then visualized using a directed bipartite graph (Supplementary Fig. 6b)(39). Results of these analyses were used to assign GEMM 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) (Fig. 6c). Similarly, the second GP3 GEMM MG harbored neoplastic and non-neoplastic subpopulations (8,144 cells) (Fig. 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) (Fig. 6f, Supplementary Fig. 7, Supplementary Table 9), but with two subpopulations with no apparent human equivalent (MB-N1, -N2) and no apparent cell cycle subpopulations. Both GP3 GEMMs 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 (Supplementary Fig. 6a, 7a). Collectively, these data suggest that the MP GEMM caries more fidelity to human GP3 tumors

SHH GEMM 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 (12,421 cells; Fig. 6g). Neoplastic cells 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). (Fig. 6h,

Supplementary Fig. 8, Supplementary Table 9), including a neuronally-differentiated subpopulation (MS-C2) corresponding to the human SHH nodule-associated SHH-C2 subpopulation. Cumulatively we show that the GEMM 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 GEMM 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 novel neoplastic and immune subpopulations to provide deeper insights into a broad range of aspects of MB biology.

Our study identifies associations between chromosomal aberrations both with single cell clustering patterns and progenitor proportions. We show that separation of sample-specific cell clusters is driven in part by the extent 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 is 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 international consortium studies have interogated MB bulk-tumor samples, identifying consensus molecular subgroups with further subdivision of these into multiple subtypes (2-4). Our study reveals differential subpopulation proportions between MB subgroups and subtypes, and restriction of a number of consensus subgroup transcriptional markers to discrete subpopulations. These findings underscore the significant contribution of MB cellular heterogeneity to subgroup/subtype assignment.

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 (12) and more recently as a defining feature of GP3 (10, 11). 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 neurodevelopment. 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* GEMM models, that we and others have shown to recapitulate the cellular heterogeneity of SHH MB (40).

The spectrum 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 these 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 (41). 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. 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 (33). 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 GEMM 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.

Materials and methods

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 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) (Supplementary Table 1). ScRNA-seq samples were rapidly dissociated into single cells using a mechanical process as described previously (33), 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 (42).

Methods for MP and MG GP3 MB GEMM models 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 (34, 36). 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 NOD-SCID-IL2R-gamma (NSG) mice. Animals were monitored weekly and euthanized when they showed signs of tumor. Tumors were then enzymatically dissociated and viably cryopreserved prior to scRNA-seq analysis.

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

To generate mice with SHH-driven MB, 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 (18). 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 cryopreserved prior to scRNA-seq analysis.

Single-cell capture, RNA library preparation and sequencing

ScRNA-seq was performed as previously described (7). 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 libraries were 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 (Supplementary Fig. 1). 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 (8). 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 (43).

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 *et al.*, as implemented in the Seurat function AddModuleScore (44). Single-cell regulatory network inference and clustering (pySCENIC) was used to identify transcription factor regulatory networks at the single cell level (9) . Bipartite graphs were generated, using iGraph, between subpopulations using the correlation of normalized gene expression of shared variable genes as edge weights (39). Edges were removed if adjusted p-values were < 0.05 or if the correlations were less than 0.2 (GEMM 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/subtypes, and chromosomal copy number variants (CNVs) (Supplementary Table 1).

Deconvolution

Cibersort was used to perform deconvolution of bulk-tumor MB transcriptomes using scRNA-seq subpopulation signature genes (45). The combined MAGIC transcriptomic dataset (n=747) was 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 (Supplementary Table 2). Cibersort was run using these datasets with 100 permutations.

Immunohistochemistry

Immunohistochemistry was performed on 5-µm formalin-fixed, paraffin-embedded (FFPE) 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 (Abcam (clone EP929Y), 1:200 dilution), IMPG2 (Novus (Cat# NBP2-58919), 1:400), MRC1 (Sigma (clone 5C11), 1:500) and STMN2 (Novus (Cat#NBP1-49461), 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 (N.W., A.M.D.).

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.

Data and Code Availability.

Custom code used in this study is available at https://github.com/rnabioco/medulloblast. ScRNAseq and methylation data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) database and are publicly accessible through GEO SuperSeries accession number GSE156053 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE156053). All MB scRNAseq data is also available as a browsable webresource at the Pediatric Neuro-oncology Cell Atlas (pneuroonccellatlas.org).

Statistics

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

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 GEMM 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.

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).

Figure titles and legends

Fig. 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. Abbreviations: astro, astrocyte; oligo, oligodendrocyte; vasc, vascular endothelium.

Fig. 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). Representative IHC staining pattern of IMPG2 (brown) in **d** GP3 (IHC score = 3) and **e** GP4 (IHC score = 2) patient samples (scale bars = 20μm). Deconvoluted **f** GP3-C2 photoreceptor, **g** GP3-C1 neuronally-differentiated, **h** GP3-B2 and **i** GP3-B1 progenitor subpopulation fractions in the MAGIC GP3 cohort (GP3-alpha, n=67; GP3-beta, n=37; GP3-gamma, n=37). Association of **j** chromosome 8 gain, and **k** patient survival with proportion of GP3-B2 (MAGIC GP3 cohort). **l** Subpopulation fraction correlation (Pearson) between GP3-C2 and GP3-B1 (MAGIC GP3 cohort) colored by subtype.

Fig. 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: GP4-alpha, n=97; GP4-beta, n=109; GP4-gamma, n=11). Representative IHC staining pattern of neuron differentiated subpopulation marker GRIA2 (brown) in **e** a GP4 patient sample (IHC score = 2.5) and **f** a GP3 patient sample (IHC score = 0.75)(scale bars = 20μm). **g** GP4-C2 subpopulation marker GRIA2 IHC score in a cohort of MB (GP3, n=7; GP4, n=11; SHH, n=10). **h** Deconvoluted GP4-C2 photoreceptor subpopulation fractions (MAGIC GP4 cohort). **i** Association of patient survival with ratio of progenitor to differentiated deconvoluted fractions (MAGIC GP4 cohort).

Fig. 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** Heatmap showing enrichment of top neuron-related GOTerms across differentiated subpopulations. **d,e** Representative IHC staining patterns of SHH-C2 subpopulation marker STMN2 (brown) in a SHH patient samples (scale bars = **d** 40μm and **e** 20μm). 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).

Fig. 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 pattern of M2-myeloid subpopulation marker MRC1 (brown) in an SHH patient sample showing accumulation at nodule linings (scale bar = 100μm). **e** Deconvoluted non-activated microglia subpopulation in all MB subtypes (MAGIC cohort).

Fig. 6. GEMM cellular heterogeneity and fidelity with human MB subpopulations Alignment of GEMM with human MB subgroup scRNA-seq data, revealing that GEMM cells clustered most closely their intended human MB subgroups. **a** UMAP projection of Harmony aligned (theta=2) GEMM and human MB scRNA-seq data. **b** Hierarchical clustering of GEMM and human MB scRNA-seq data. **c** Non-harmonized alignment of GEMM 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 GEMM model MP and human GP3 neoplastic subpopulations. This was repeated for **e,f** GP3 GEMM model MG, and **g,h** SHH model MS. Abbr: MP, GP3 GEMM Myc/Trp53; MG, GP3 GEMM Myc/GFI1; MS, SHH GEMM Math-1/SmoM2; vasc., vascular; oligo, oligodendrocyte.

Table 1. Summary of MB subgroup neoplastic subpopulation characteristics.

program		GP3	GP4	SHH
A-cell cycle		**GP3-A**	**GP4-A1**	**SHH-A1**
	GOTerm	cell cycle	cell cycle, G2M,	cell cycle, G2M
	inferred TF activity	MYBL2	FOXM1, MYBL2	MYBL2
	marker genes	CENPF, TOP2A, TUBA1B, HMGB2, MKI67	TOP2A, HMGB2, CENPF, TUBA1B, MKI67	CENPF, TOP2A, HMGB2, MKI67
	proportion	12.5%	7.5%	12.1%
			GP4-A2	**SHH-A2**
	GOTerm		cell cycle, S phase	cell cycle: S phase
	inferred TF activity		ZNF492, E2Fs	E2Fs
	marker genes		HELLS, XRCC2, ATAD5, BRCA1	RRM2, POLQ, FANCA, XRCC2, BRCA1, HELLS
	proportion		4.6%	10%
B-progenitor		**GP3-B1**	**GP4-B1**	**SHH-B1**
	GOTerm	translation/oxidative respiration	oxidative phosphorylation/ translation	translation/oxidative respiration/neg. reg. of epithelial cell differentiation
	inferred TF activity	HMGA1, MYC, MAX	BARHL1, NEUROD2, SOX4, JUNB, JUND	ATOH1, ZEB1, HES6
	marker genes	ribosomal	STMN2, ribosomal	SFRP1, CXCR2, ribosomal
	proportion	33.3%	25.9%	28.7%
		GP3-B2	**GP4-B2**	**SHH-B2**
	GOTerm	translation/oxidative respiration	translation	translation/oxidative respiration
	inferred TF activity	HLX, ATF4, MYC	HEY1, TEAD2, ELK1	ELK1, THAP11, MAX
	marker genes	ribosomal, MYC	STMN2, ribosomal	ribosomal
	proportion	1.9%	17.3%	17.1%

		GP3-C1	GP4-C1	SHH-C1
C-neuronally differentiated	GOTerm	RNA processing/ axo-dendritic transport	RNA processing/ axo-dendritic transport	RNA processing/ axo-dendritic transport
	inferred TF activity	ARNT	ONECUT2, LHX4	POU2F2, KMT2A
	marker genes	LUC7L3, DST	GRIA2, LUC7L3, DST	GRIA2, LUC7L3, DST
	proportion	37.6%	32.7%	28.2%
		GP3-C2	**GP4-C2**	**SHH-C2**
	GOTerm	photoreceptor	photoreceptor	neuron development
	inferred TF activity	SHOX2, CRX	MEIS1, CRX	SOX4, NEUROD1
	marker genes	NRL, IMPG2	NRL, IMPG2	STMN4, STMN2
	proportion	14.6%	6.9%	3.6%

Abbreviations: TF, transcription factor.

Supplementary Tables

Supplementary Table 1. Patient and sample details.

Supplementary Table 2. Cibersort neoplastic and immune subpopulation signatures.

Supplementary Table 3. Neoplastic subpopulation marker genes.

Supplementary Table 4. Neoplastic subpopulation GOTerm enrichments.

Supplementary Table 5. Neoplastic subpopulation inferred transcription factor activity.

Supplementary Table 6. Survival analyses of deconvoluted MAGIC cohort subpopulation proportions.

Supplementary Table 7. Correlation matrix for deconvoluted MAGIC cohort subpopulation proportions.

Supplementary Table 8. Immune subpopulation marker genes.

Supplementary Table 9. GEMM subpopulation marker genes.

References

1. Northcott PA, et al. Medulloblastoma comprises four distinct molecular variants. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. Apr 10 2011;29(11):1408-14. doi:10.1200/JCO.2009.27.4324

2. Cavalli FMG, et al. Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell*. Jun 12 2017;31(6):737-754 e6. doi:10.1016/j.ccell.2017.05.005

3. Northcott PA, et al. The whole-genome landscape of medulloblastoma subtypes. *Nature*. Jul 19 2017;547(7663):311-317. doi:10.1038/nature22973

4. Sharma T, et al. Second-generation molecular subgrouping of medulloblastoma: an international meta-analysis of Group 3 and Group 4 subtypes. *Acta Neuropathol*. Aug 2019;138(2):309-326. doi:10.1007/s00401-019-02020-0

5. Hovestadt V, et al. Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature*. Aug 2019;572(7767):74-79. doi:10.1038/s41586-019-1434-6

6. Vladoiu MC, et al. Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature*. Aug 2019;572(7767):67-73. doi:10.1038/s41586-019-1158-7

7. Gillen AE, et al. Single-Cell RNA Sequencing of Childhood Ependymoma Reveals Neoplastic Cell Subpopulations That Impact Molecular Classification and Etiology. *Cell Rep*. Aug 11 2020;32(6):108023. doi:10.1016/j.celrep.2020.108023

8. Korsunsky I, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. Dec 2019;16(12):1289-1296. doi:10.1038/s41592-019-0619-0

9. Van de Sande B, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc*. Jul 2020;15(7):2247-2276. doi:10.1038/s41596-020-0336-2

10. Cho YJ, et al. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J Clin Oncol*. Apr 10 2011;29(11):1424-30. doi:10.1200/JCO.2010.28.5148

11. Kool M, et al. Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS One*. Aug 28 2008;3(8):e3088. doi:10.1371/journal.pone.0003088

12. Korf HW, et al. Immunocytochemical evidence of molecular photoreceptor markers in cerebellar medulloblastomas. *Cancer*. Oct 15 1987;60(8):1763-6. doi:10.1002/1097-0142(19871015)60:8<1763::aid-cncr2820600815>3.0.co;2-3

13. Garancher A, et al. NRL and CRX Define Photoreceptor Identity and Reveal Subgroup-Specific Dependencies in Medulloblastoma. *Cancer Cell*. Mar 12 2018;33(3):435-449 e6. doi:10.1016/j.ccell.2018.02.006

14. Northcott PA, et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. *Acta Neuropathol*. Apr 2012;123(4):615-26. doi:10.1007/s00401-011-0899-7

15. Grandori C, et al. c-Myc binds to human ribosomal DNA and stimulates transcription of rRNA genes by RNA polymerase I. *Nat Cell Biol*. Mar 2005;7(3):311-8. doi:10.1038/ncb1224

16. van Riggelen J, et al. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat Rev Cancer*. Apr 2010;10(4):301-9. doi:10.1038/nrc2819

17. Aldinger KA, et al. Spatial and single-cell transcriptional landscape of human cerebellar development. *BioRxiv*. 2020;174391174391. doi:10.1101/2020.06.30.174391

18. Ocasio J, et al. scRNA-seq in medulloblastoma shows cellular heterogeneity and lineage expansion support resistance to SHH inhibitor therapy. *Nat Commun*. Dec 20 2019;10(1):5829. doi:10.1038/s41467-019-13657-6

19. Hannan AJ, et al. Differential induction and intracellular localization of SCG10 messenger RNA is associated with neuronal differentiation. *Neuroscience*. Jun 1996;72(4):889-900. doi:10.1016/0306-4522(95)00593-5

20. Eberhart CG, et al. Apoptosis, neuronal maturation, and neurotrophin expression within medulloblastoma nodules. *J Neuropathol Exp Neurol*. May 2001;60(5):462-9. doi:10.1093/jnen/60.5.462

21. Ocasio JK, et al. GSK-3 modulates SHH-driven proliferation in postnatal cerebellar neurogenesis and medulloblastoma. *Development*. Oct 10 2019;146(20)doi:10.1242/dev.177550

22. Poschl J, et al. Wnt/beta-catenin signaling inhibits the Shh pathway and impairs tumor growth in Shh-dependent medulloblastoma. *Acta Neuropathol*. Apr 2014;127(4):605-7. doi:10.1007/s00401-014-1258-2

23. Boon K, et al. N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. *EMBO J*. Mar 15 2001;20(6):1383-93. doi:10.1093/emboj/20.6.1383

24. Waszak SM, et al. Germline Elongator mutations in Sonic Hedgehog medulloblastoma. *Nature*. Apr 2020;580(7803):396-401. doi:10.1038/s41586-020-2164-5

25. Castellino RC, et al. Heterozygosity for Pten promotes tumorigenesis in a mouse model of medulloblastoma. *PLoS One*. May 26 2010;5(5):e10849. doi:10.1371/journal.pone.0010849

26. Zilionis R, et al. Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity*. May 21 2019;50(5):1317-1334 e10. doi:10.1016/j.immuni.2019.03.009

27. Hammond TR, et al. Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. *Immunity*. Jan 15 2019;50(1):253-271 e6. doi:10.1016/j.immuni.2018.11.004

28. Li Q, Barres BA. Microglia and macrophages in brain homeostasis and disease. *Nat Rev Immunol*. Apr 2018;18(4):225-242. doi:10.1038/nri.2017.125

29. Stevens B, et al. The classical complement cascade mediates CNS synapse elimination. *Cell*. Dec 14 2007;131(6):1164-78. doi:10.1016/j.cell.2007.10.036

30. Fourgeaud L, et al. TAM receptors regulate multiple features of microglial physiology. *Nature*. Apr 14 2016;532(7598):240-244. doi:10.1038/nature17630

31. Margol AS, et al. Tumor-associated macrophages in SHH subgroup of medulloblastomas. *Clin Cancer Res*. Mar 15 2015;21(6):1457-65. doi:10.1158/1078-0432.CCR-14-1144

32. Maximov V, et al. Tumour-associated macrophages exhibit anti-tumoural properties in Sonic Hedgehog medulloblastoma. *Nat Commun*. Jun 3 2019;10(1):2410. doi:10.1038/s41467-019-10458-9

33. Griesinger AM, et al. Characterization of distinct immunophenotypes across pediatric brain tumor types. *J Immunol*. Nov 1 2013;191(9):4880-8. doi:10.4049/jimmunol.1301966

34. Pei Y, et al. HDAC and PI3K Antagonists Cooperate to Inhibit Growth of MYC-Driven Medulloblastoma. Research Support, N.I.H., Extramural

Research Support, Non-U.S. Gov't. *Cancer cell*. Mar 14 2016;29(3):311-23. doi:10.1016/j.ccell.2016.02.011

35. Pei Y, et al. An animal model of MYC-driven medulloblastoma. *Cancer Cell*. Feb 14 2012;21(2):155-67. doi:10.1016/j.ccr.2011.12.021

36. Northcott PA, et al. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Research Support, N.I.H., Extramural

Research Support, Non-U.S. Gov't. *Nature*. Jul 24 2014;511(7510):428-34. doi:10.1038/nature13379

37. Lee C, et al. Lsd1 as a therapeutic target in Gfi1-activated medulloblastoma. *Nat Commun*. Jan 18 2019;10(1):332. doi:10.1038/s41467-018-08269-5

38. Mao J, et al. A novel somatic mouse model to survey tumorigenic potential applied to the Hedgehog pathway. *Cancer Res*. Oct 15 2006;66(20):10171-8. doi:10.1158/0008-5472.CAN-06-0657

39. Laughney AM, et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat Med*. Feb 2020;26(2):259-269. doi:10.1038/s41591-019-0750-6

40. Cheng Y, et al. NeuroD1 Dictates Tumor Cell Differentiation in Medulloblastoma. *Cell Rep*. Jun 23 2020;31(12):107782. doi:10.1016/j.celrep.2020.107782

41. Schafer DP, Stevens B. Microglia Function in Central Nervous System Development and Plasticity. *Cold Spring Harb Perspect Biol*. Jul 17 2015;7(10):a020545. doi:10.1101/cshperspect.a020545

42. Wohnhaas CT, et al. DMSO cryopreservation is the method of choice to preserve cells for droplet-based single-cell RNA sequencing. *Sci Rep*. Jul 23 2019;9(1):10699. doi:10.1038/s41598-019-46932-z

43. Korsunsky I, et al. Presto scales Wilcoxon and auROC analyses to millions of observations. *BioRxiv*. 2019;653253. doi:10.1101/653253

44. Tirosh I, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. Apr 8 2016;352(6282):189-96. doi:10.1126/science.aad0501

45. Newman AM, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. May 2015;12(5):453-7. doi:10.1038/nmeth.3337

Figure 1

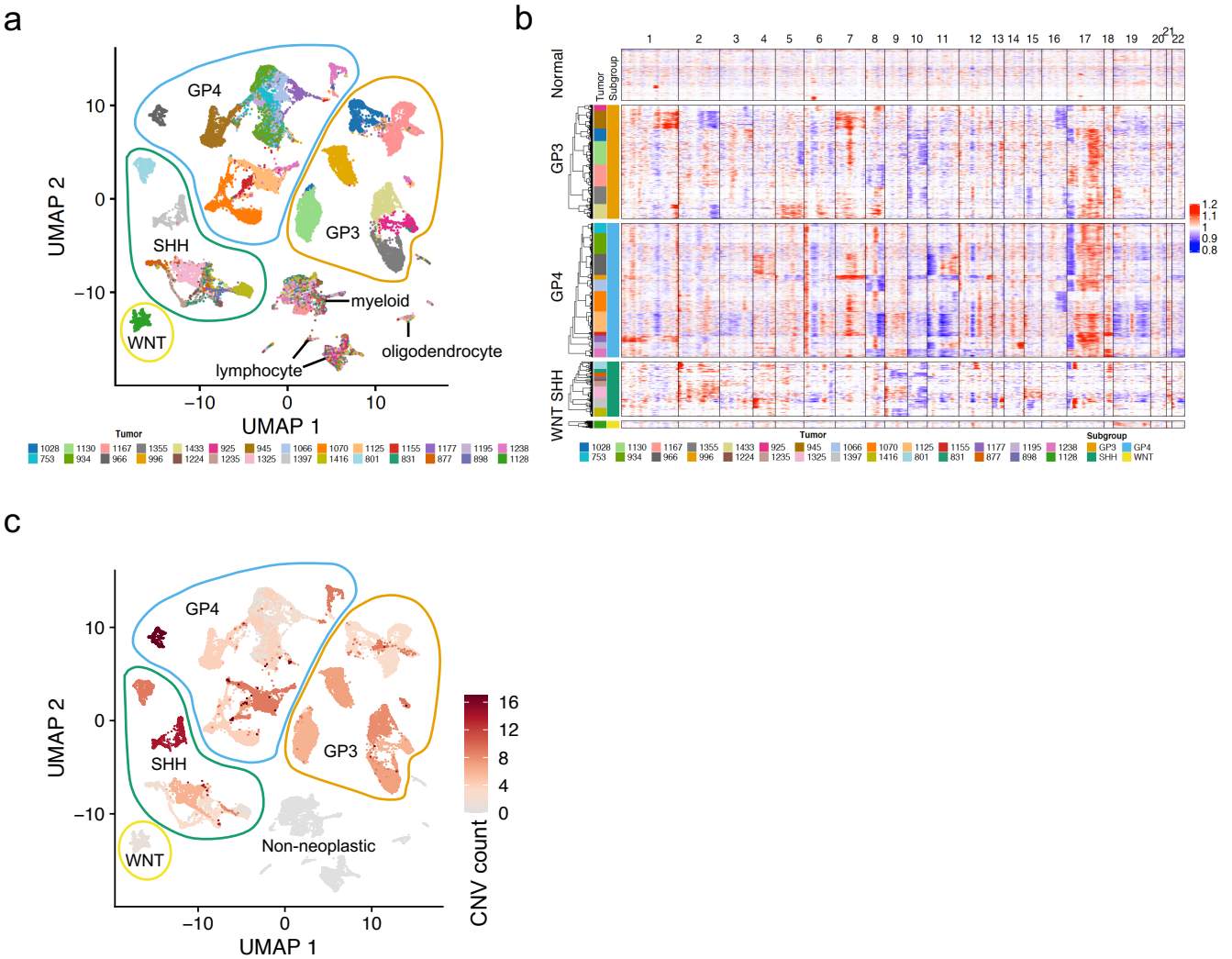

a
GP4
GP3
SHH
WNT
myeloid
oligodendrocyte
lymphocyte
UMAP 1
UMAP 2
Tumor
b
Normal
GP3
GP4
WNT SHH
Tumor
Subgroup
Subgroup
GP3
GP4
SHH
WNT
c
GP4
GP3
SHH
WNT
Non-neoplastic
UMAP 1
UMAP 2
CNV count

Figure 2

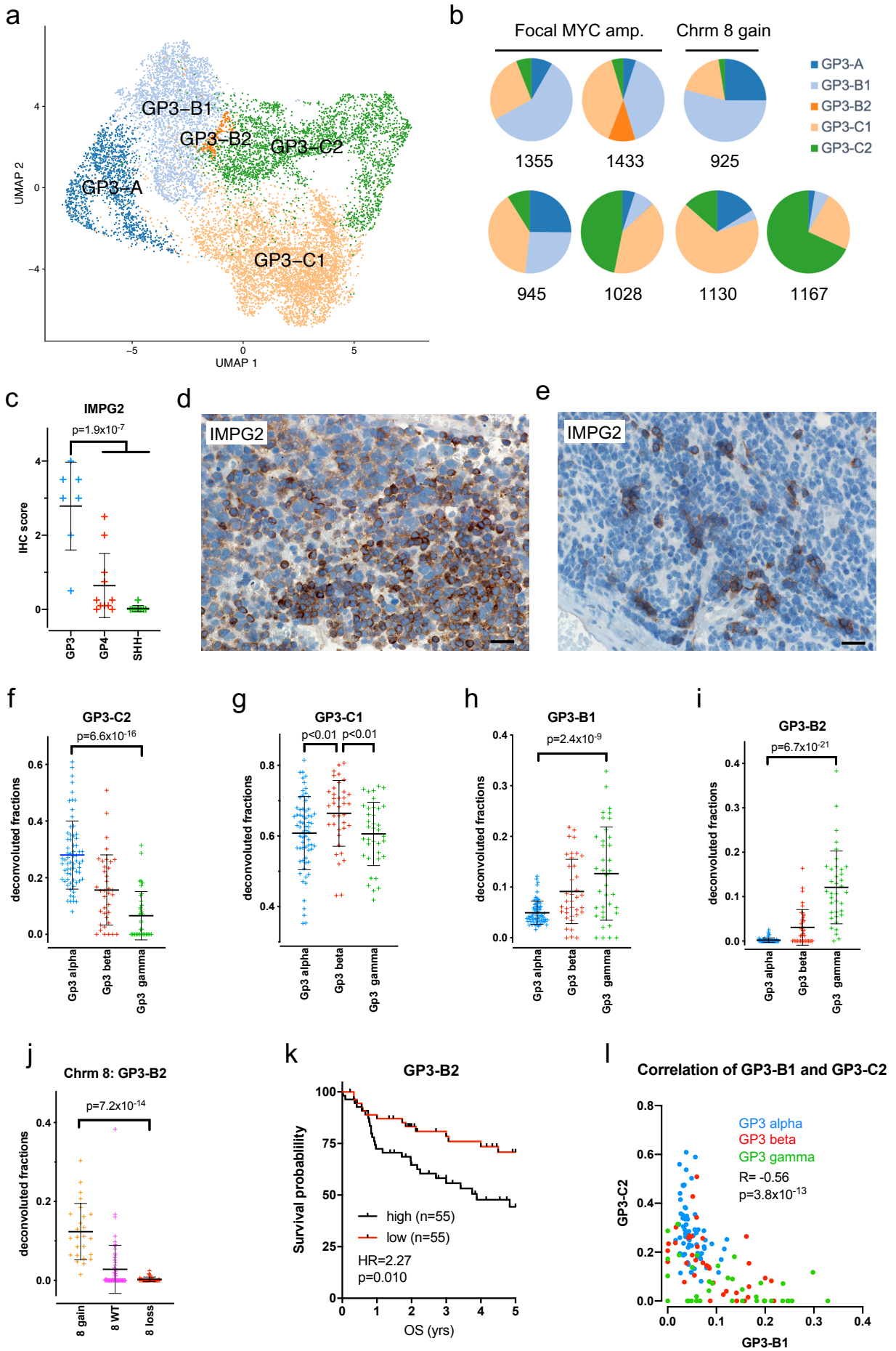

a
GP3–B1
GP3–B2
GP3–C2
GP3–A
GP3–C1
UMAP 2
UMAP 1
b
Focal MYC amp.
Chrm 8 gain
GP3-A
GP3-B1
GP3-B2
GP3-C1
GP3-C2
1355
1433
925
945
1028
1130
1167
c
IMPG2
p=1.9x10-7
IHC score
GP3
GP4
SHH
d
IMPG2
e
IMPG2
f
GP3-C2
p=6.6x10-16
deconvoluted fractions
Gp3 alpha
Gp3 beta
Gp3 gamma
g
GP3-C1
p<0.01
p<0.01
deconvoluted fractions
Gp3 alpha
Gp3 beta
Gp3 gamma
h
GP3-B1
p=2.4x10-9
deconvoluted fractions
Gp3 alpha
Gp3 beta
Gp3 gamma
i
GP3-B2
p=6.7x10-21
deconvoluted fractions
Gp3 alpha
Gp3 beta
Gp3 gamma
j
Chrm 8: GP3-B2
p=7.2x10-14
deconvoluted fractions
8 gain
8 WT
8 loss
k
GP3-B2
Survival probability
high (n=55)
low (n=55)
HR=2.27
p=0.010
OS (yrs)
l
Correlation of GP3-B1 and GP3-C2
GP3 alpha
GP3 beta
GP3 gamma
R= -0.56
p=3.8x10-13
GP3-C2
GP3-B1

Figure 3

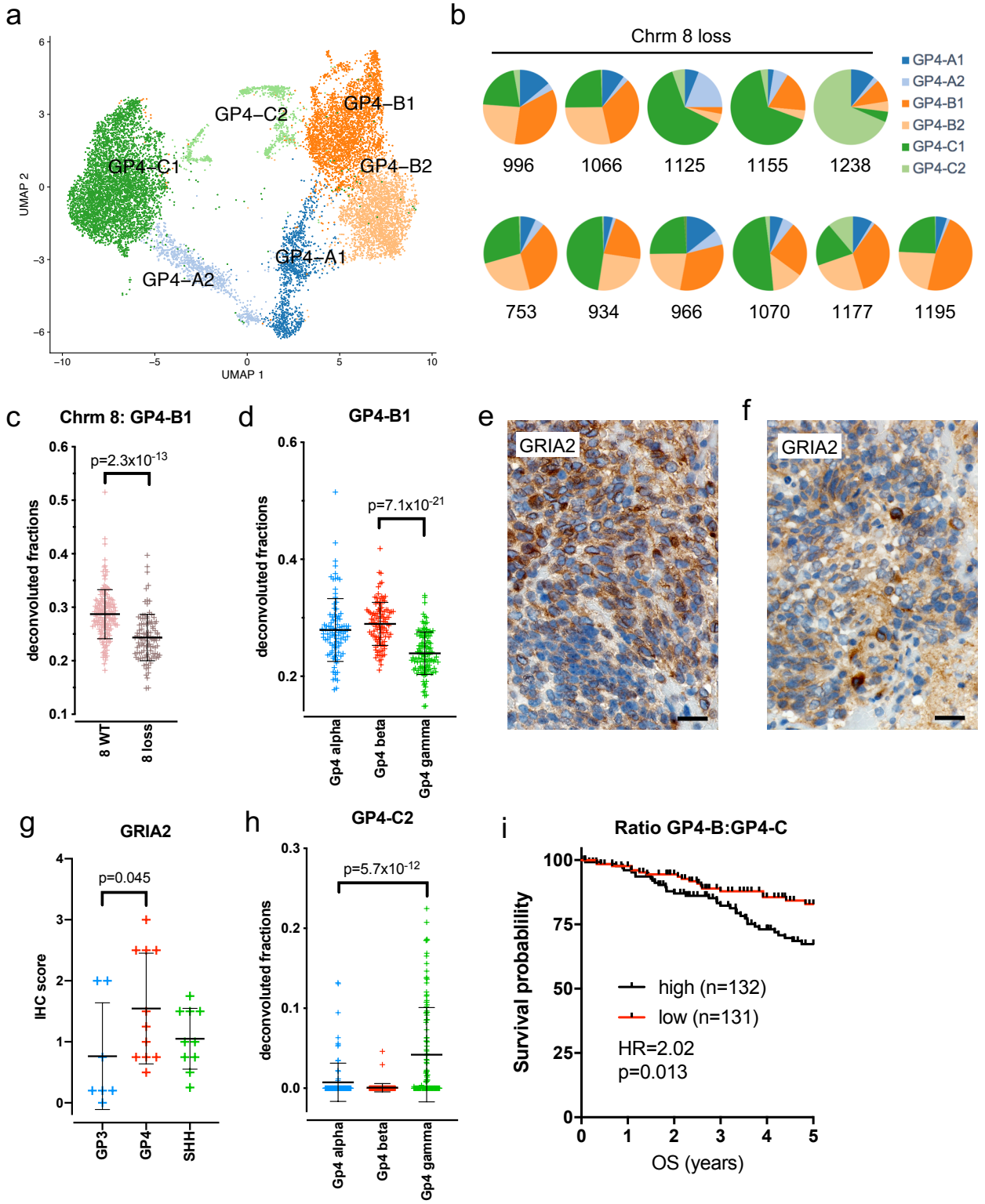

a
GP4–C1
GP4–C2
GP4–B1
GP4–B2
GP4–A1
GP4–A2
UMAP 1
UMAP 2
b
Chrm 8 loss
GP4-A1
GP4-A2
GP4-B1
GP4-B2
GP4-C1
GP4-C2
996
1066
1125
1155
1238
753
934
966
1070
1177
1195
c
Chrm 8: GP4-B1
p=2.3x10-13
deconvoluted fractions
8 WT
8 loss
d
GP4-B1
p=7.1x10-21
deconvoluted fractions
Gp4 alpha
Gp4 beta
Gp4 gamma
e
GRIA2
f
GRIA2
g
GRIA2
p=0.045
IHC score
GP3
GP4
SHH
h
GP4-C2
p=5.7x10-12
deconvoluted fractions
Gp4 alpha
Gp4 beta
Gp4 gamma
i
Ratio GP4-B:GP4-C
Survival probability
high (n=132)
low (n=131)
HR=2.02
p=0.013
OS (years)

Figure 4

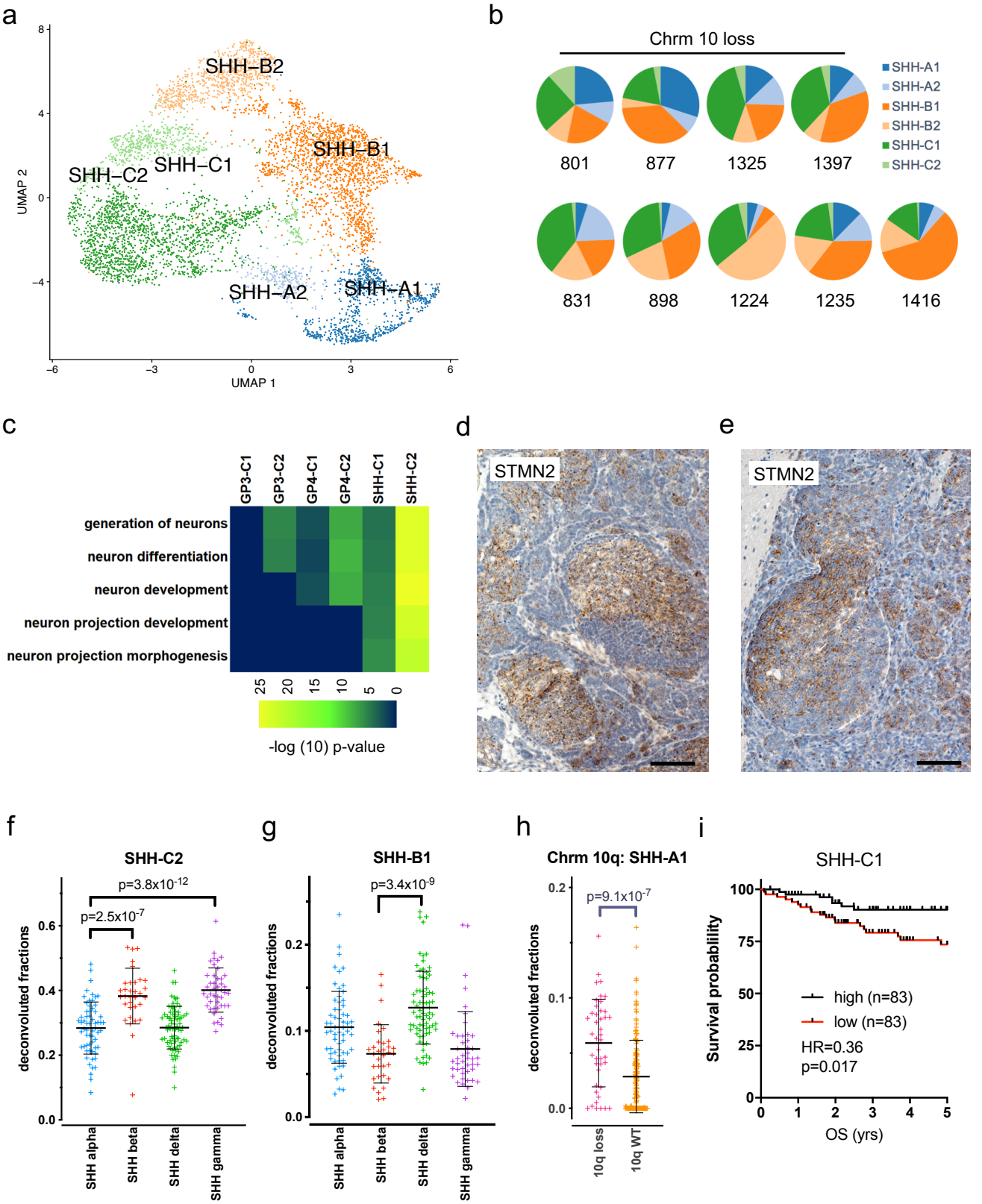

a
SHH–B2
SHH–B1
SHH–C1
SHH–C2
SHH–A2
SHH–A1
UMAP 2
UMAP 1
b
Chrm 10 loss
SHH-A1
SHH-A2
SHH-B1
SHH-B2
SHH-C1
SHH-C2
801
877
1325
1397
831
898
1224
1235
1416
c
GP3-C1
GP3-C2
GP4-C1
GP4-C2
SHH-C1
SHH-C2
generation of neurons
neuron differentiation
neuron development
neuron projection development
neuron projection morphogenesis
-log (10) p-value
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

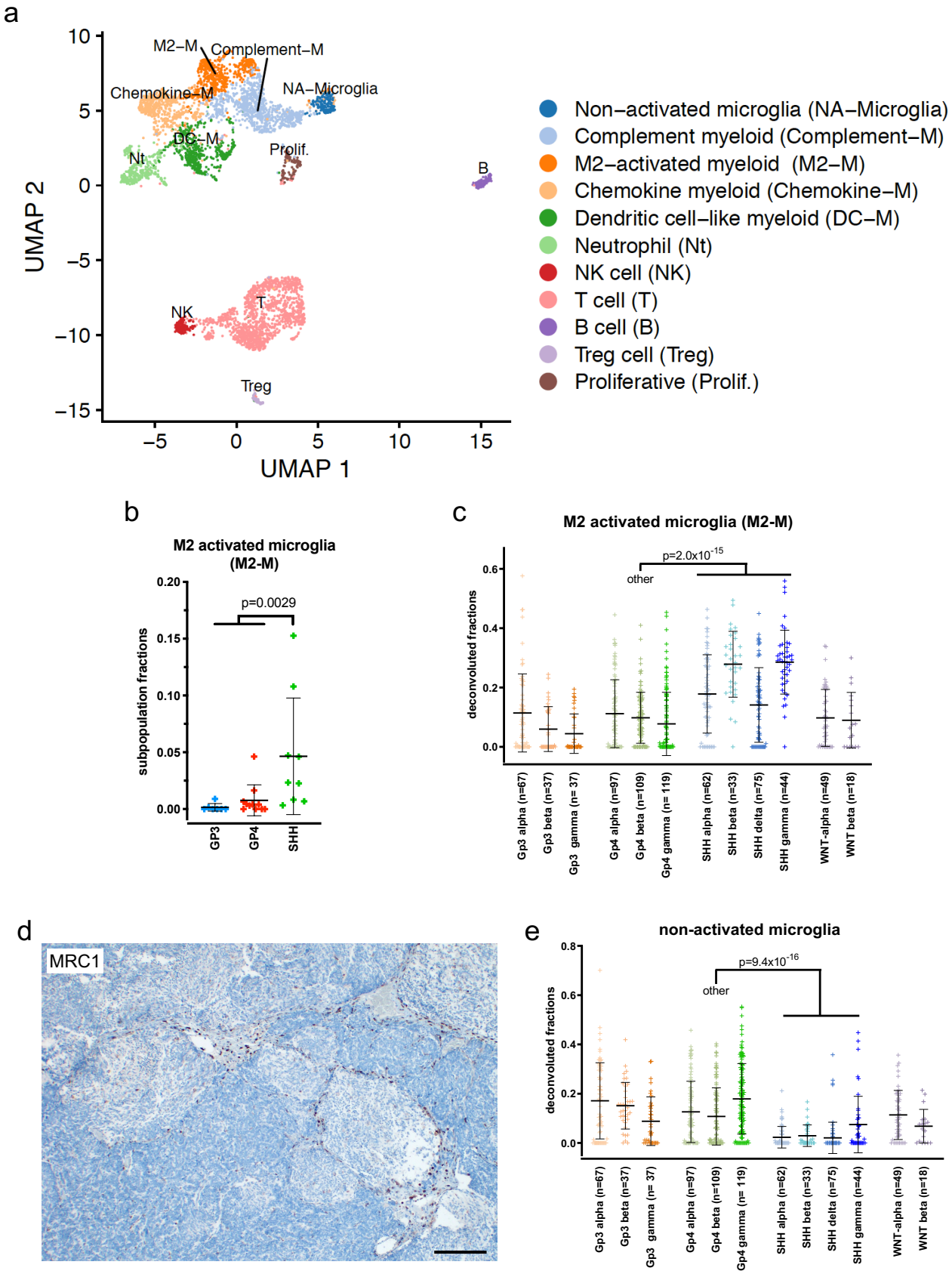
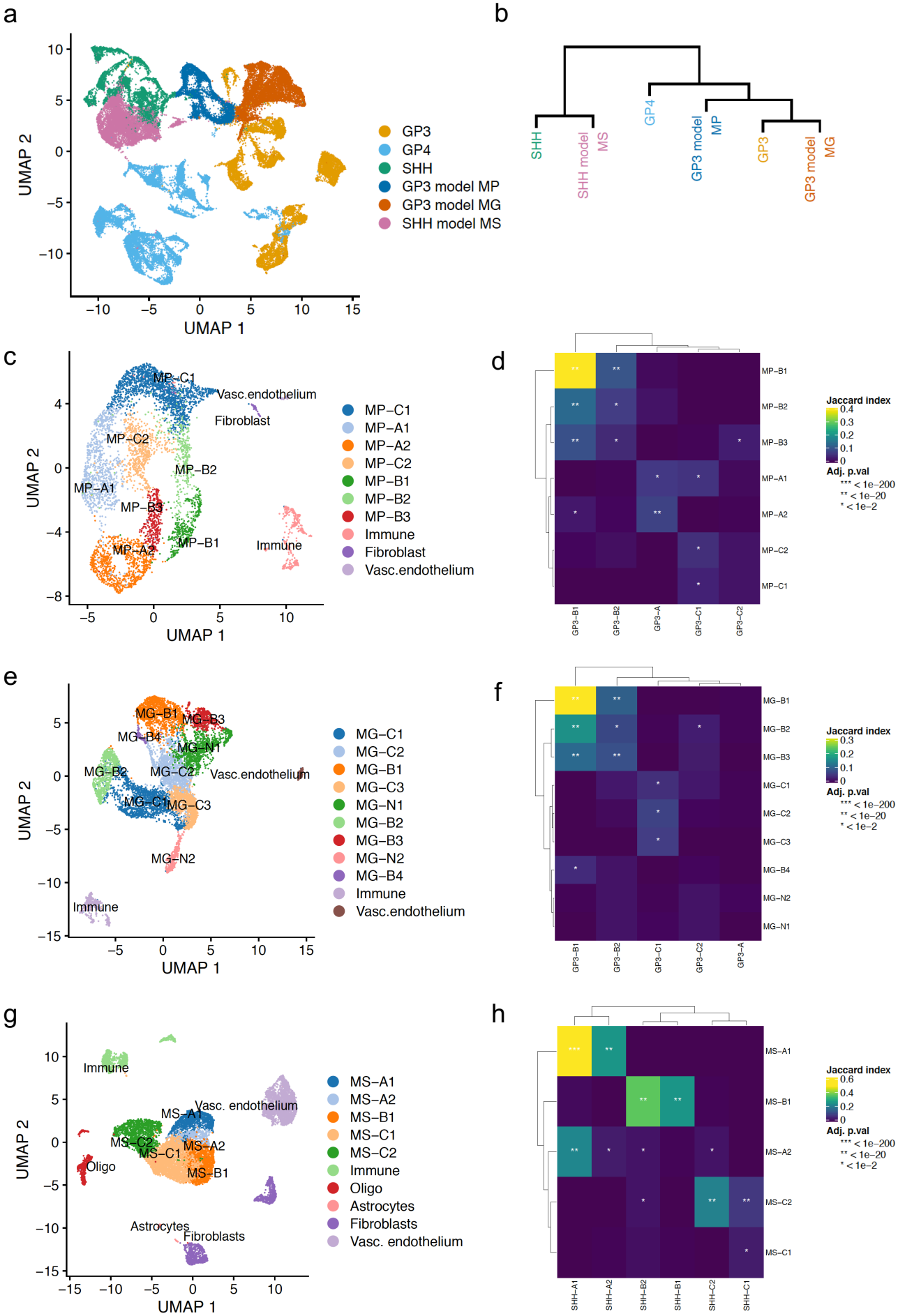

a
M2–M
Complement–M
Chemokine–M
NA–Microglia
DC–M
Nt
Prolif.
B
NK
T
Treg
UMAP 2
UMAP 1
Non–activated microglia (NA–Microglia)
Complement myeloid (Complement–M)
M2–activated myeloid (M2–M)
Chemokine myeloid (Chemokine–M)
Dendritic cell–like myeloid (DC–M)
Neutrophil (Nt)
NK cell (NK)
T cell (T)
B cell (B)
Treg cell (Treg)
Proliferative (Prolif.)
b
M2 activated microglia (M2-M)
p=0.0029
subpopulation fractions
GP3
GP4
SHH
c
M2 activated microglia (M2-M)
p=2.0x10-15
other
deconvoluted fractions
Gp3 alpha (n=67)
Gp3 beta (n=37)
Gp3 gamma (n= 37)
Gp4 alpha (n=97)
Gp4 beta (n=109)
Gp4 gamma (n= 119)
SHH alpha (n=62)
SHH beta (n=33)
SHH delta (n=75)
SHH gamma (n=44)
WNT-alpha (n=49)
WNT beta (n=18)
d
MRC1
e
non-activated microglia
p=9.4x10-16
other
deconvoluted fractions

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
Jaccard index
Adj. p.val
*** < 1e–200
** < 1e–20
* < 1e–2
MP–B1
MP–B2
MP–B3
MP–A1
MP–A2
MP–C2
MP–C1
GP3–B1
GP3–B2
GP3–A
GP3–C1
GP3–C2
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
Jaccard index
Adj. p.val
*** < 1e–200
** < 1e–20
* < 1e–2
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
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
Jaccard index
Adj. p.val
*** < 1e–200
** < 1e–20
* < 1e–2
MS–A1
MS–B1
MS–A2
MS–C2
MS–C1
SHH–A1
SHH–A2
SHH–B2
SHH–B1
SHH–C2
SHH–C1