Analysis of single-cell RNA-sequencing data to identify quiescent and proliferating neural cell populations in Glioblastoma

Rajeev Vikram *[1], Wen Cheng Chou[1], Pei-Ei Wu[1], Wei-Ting Chen[1], Chen-Yang Shen *[1,2]

[1]Institute of Biomedical Sciences, Taipei 115, Taiwan, [2]Graduate Institute of Environmental Science, China Medical University, Taichung 404, Taiwan

*corresponding authors: rajeev.vkrm@ibms.sinica.edu.tw, bmcys@ibms.sinica.edu.tw

ABSTRACT

Background:

Diffuse Glioblastoma (GBM) has high mortality and remains one of the most challenging type of cancer to treat. Identifying and characterizing the cells populations driving tumor growth and therapy resistance has been particularly difficult owing to marked inter and intra tumoral heterogeneity observed in these tumors. These tumorigenic populations contain long lived cells associated with latency, immune evasion and metastasis.

Methods:

Here, we analyzed the single-cell RNA-sequencing data of high grade glioblastomas from four different studies using integrated analysis of gene expression patterns, cell cycle stages and copy number variation to identify gene expression signatures associated with quiescent and cycling neuronal tumorigenic cells.

Results:

The results show that while cycling and quiescent cells are present in GBM of all age groups, they exist in a much larger proportion in pediatric glioblastomas. These cells show similarities in

their expression patterns of a number of pluripotency and proliferation related genes. Upon unbiased clustering, these cells explicitly clustered on their cell cycle stage. Quiescent cells in both the groups specifically overexpressed a number of genes for ribosomal protein, while the cycling cells were enriched in the expression of high-mobility group and heterogeneous nuclear ribonucleoprotein group genes. A number of well-known markers of quiescence and proliferation in neurogenesis showed preferential expression in the quiescent and cycling populations identified in our analysis. Through our analysis, we identify ribosomal proteins as key constituents of quiescence in glioblastoma stem cells.

Conclusions:

This study identifies gene signatures common to adult and pediatric glioblastoma quiescent and cycling stem cell niches. Further research elucidating their role in controlling quiescence and proliferation in tumorigenic cells in high grade glioblastoma will open avenues in more effective treatment strategies for glioblastoma patients.

Keywords: Glioblastoma, scRNA-seq, Quiescence, proliferation, neurogenesis, GSC, stem cells

INTRODUCTION

Diffuse gliomas are tumors of the central nervous system with histological similarity to glial cells. Worldwide, approximately 100,000 new cases of diffuse glioma are reported every year[1]. Despite it being a relatively rare cancer type, diffuse gliomas have a very poor prognosis with high mortality burden. The 2016 WHO classification of gliomas divides them into astrocytomas, oligodendrogliomas and oligoastrocytomas with subgroupings based on IDH mutations and 1p19q co-deletion status[2]. Irrespective of the categories, the tumors are graded from one to four according to the histological degree of malignancy.

The grade IV diffuse astrocytoma (IDH-wildtype) also called as glioblastoma (GBM) accounts for about 75% of all diffuse gliomas with a median survival of about one to two years after therapy, making it the most lethal of gliomas[2, 3]. Intratumoral heterogeneity in GBM is a key challenge to developing effective therapeutic strategies. Neurodevelopmental bi-lineage hierarchy does partially explain the heterogeneity in IDH-mutant and pediatric gliomas, however, this bi-lineage hierarchy model fails to explain the widespread phenotypic heterogeneity and evolving phenotypic states in GBM. The cancer stem cells (CSC) theory suggests that Glioblastoma Stem-like Cells (GSCs) are at the center of the tumor organization and instrumental in generating and replicating intratumoral phenotypic heterogeneity[4]. Indeed, similar to other cancers, intratumoral heterogeneity, resistance to treatment and relapse in GBM has been attribute to this small subpopulation in a number of studies. Although GSCs seem to be a key target in GBM therapy, their existence and cellular nature remains a hotly debated topic. While there is evidence pointing to the presence of such GSCs, identification of these cells has remained a challenge. Primarily because there is no marker which can be considered to universally identify GSCs[5].

In most cancers including GBM, single surface marker approach has been used to identify CGSCs. A number of cell membrane antigens like CD133, CD15/SSEA, CD44, PDGFRA, EGFR or A2B5 are shown to be associated with potential GSCs. Earlier studies showing the tumorigenic potential of cells isolated using one or a combination of these markers [6–11] did not address the tumorigenic potential of marker negative cells. Later studies however show that both the marker positive and negative glioma fractions can show multipotent behavior [8, 12]. Recent reports show that marker negative cells are able to generate marker positive cells and replicate the tumor heterogeneity[13]. Thus the evidence so far indicates a non-hierarchical model where cells niches with strong cellular plasticity are at the core of recreating intratumoral heterogeneity. Evidence supporting intratumoral cell niches with high cellular plasticity in glioblastoma comes from a recent study by Jung, Erik, et al.[14]. The authors show the existence two complementary cellular niches driving tumor progression and therapy resistance in GBM. Further evidence of the plastic nature of GSC niches come from studies which demonstrate the generation of potential GSCs from non-tumorigenic glioma cells [15]. Based on these and a number of other recent studies, it seems that within the tumor, potential GSCs remain in interactive niches and are highly plastic in addition to being able to acquire therapy resistance and tumorigenicity. Recent studies in tumor immunosurveillance evasion suggests that these niches are composed of small populations of cycling and quiescent stem cell like cells [16].

Advances in newer methods to study cellular transcriptomes at the single cell level especially massively parallel single-cell RNA-sequencing (scRNAseq) has significantly enhanced our understanding of spatial and temporal heterogeneity in glioblastoma. Recent single cell studies of gliomas have shown the presence of a progenitor type population [17–19]. Interestingly, in terms of conventional markers, tumors from different patients show variability in their expression,

suggesting heterogeneity within GSCs[20]. As more and more genomic and transcriptomic data from single cell experiments in gliomas becomes available, it is becoming more evident that although canonical GSC markers seem to be associated with proliferative cells in low grade gliomas, such correlation is not evident in GBM[21]. Thus, projects designed to identify GSCs on the hierarchical CGSC model have largely been ineffective.

To develop a better diagnostic and treatment strategy for GBM as well as low grade glioma, is it important understand the dynamics of the tumor microenvironment, especially the intrinsic plasticity of the cell niches. Understanding the mechanism of maintenance of these highly plastic cell subpopulations within the tumor, the role of the microenvironment dynamics in selection and survival of such populations and their propagation are instrumental in unearthing the reasons for the development of resistance to Temozolomide (TMZ) chemotherapy and radiotherapy.

A number of recent studies have utilized scRNAseq to study gliomas of different origin and grade generating a wealth of data on the transcriptomic nature of cells within the tumor[14, 17, 19, 22]. While these studies primarily focused on different states of gliomas and tumor-immune cell interaction, few studies have tried to delineate differences in cellular states within neural cells in GBM. Reanalysis of these scRNA-seq datasets can give us a deeper understanding of the genomic and transcriptomic commonalities within malignant neural subpopulations. Identification of GSC niches with quiescent population perhaps holds and the key to developing strategies which can target genes/transcripts involved in maintaining GSC plasticity and crosstalk in high grade gliomas.

Here, we analyze the scRNAseq data from four different studies encompassing thousands of tumor and peripheral cells from pediatric and adult IDH-wildtype glioblastoma patients to identify and study quiescent and cycling GSCs. The tumor types range from primary to relapsed

tumors. Our results show that cycling and quiescent like-cell subpopulation are present in most GBM tumors with a gene expression signature associated with ribosomal biogenesis, cell cycle activation and malignancy. Key overexpressed genes include *DCX*, *SOX4* and *DLL3*, known markers for quiescence, stemness and tumorigenicity. These cells have a Copy number variance pattern distinguishing them from other neural cells subtypes within the tumor. and key genes/transcripts expression pattern in these niches across GBM tumor types.

RESULTS

Selection of datasets and identification of neural cells

To identify common GSC like populations across GBMs, we selected scRNAseq datasets from different studies representing IDH-wildtype, grade IV glioblastomas. Included datasets represent major patient groups (pediatric, adult and recurrent). Table 1 shows the major characteristics of the included datasets. For differential gene expression analysis of GBM subpopulations, we also included brain metastasis (lung squamous cell carcinoma) data set from the study GSE117891.

Table 1 Dataset Summary

Dataset	No. of Cells	Location	No. of Patients	Tumor Type	Tumor Grade
GSE84465	1745	Tumor/Periphery	4	Primary	IDH-wildtype
GSE131928	1460	Tumor	6	Primary/ Recurrent	IDH-wildtype
GSE103224	3679	Tumor	2	Recurrent	IDH-wildtype
GSE117891	359	Tumor/Periphery	1	Metastatic (Lung Squamous Cell Carcinoma)	NA

Significant inter and intratumoral heterogeneity is a challenge in identifying GSC like niches because gene expression patterns of different cell types and sample origin (i.e., transcriptomic diversity of the samples) induce strong variance which often masks the similarities in cellular programs in small subpopulations. Malignant cells with stem cell like properties are typically a very small subset of the tumor population. We therefore first sought to enrich our datasets based on clear cell identity and malignancy. The datasets were first clustered using the standard Seurat-sctarnsform pipeline. As, GBMs primarily originate in neural cells, we first identified the clusters of clear neural/glial (neuronal) or myeloid/immune origin by measuring the expression of known markers for immune cells (*CD4, CD83* & *HLA-DRA*) and neuronal cells (*S100B*, *OLIG1* & *SCG2*) [23, 24] (Figure 1 B, C). Clusters are represented using Uniform Manifold Approximation and Projection (UMAP) (Figure 1A). The neuronal cluster significantly overexpressed *EGFR*, a gene overexpressed in 30-50% of all GBMs and associated with neoplasia. Interestingly, non-GBM tumor cells (GSE117891) did not express EGFR and showed markedly different non myeloid/immune cluster profile. (Figure 1D). These clusters showed distinct expression profiles for glial cell markers for astrocyte(S100B) and oligodendrocytes(OLIG2). Indicating the presence of transformed non neural cells in these clusters (Figure 1 B, C).

Table 1 Study Group Composition

Group	Source	Patient-ID	Molecular Subtype	No. of Transformed Cells
Adult	GSE84465	BT-S1	Classical	351
		BT-S2	Classical	713
		BT-S4	Classical	367
		BT-S6	Proneural	143
	GSE103224	PJ-32	Mesenchymal	574
		PJ-35	Classical	3105
	GSE117891	GS-15	NA	352
Pediatric	GSE131928	BT-749	Proneural	253
		BT-771	Mesenchymal	256
		BT-786	Proneural	160
		BT-830	Mesenchymal	166
		BT-920	Mesenchymal	07
		BT-1160	Proneural	332
		MGH-85	Proneural	245

Integrated cluster analysis

To identify malignant or transformed stem-like cells, we decided to focus exclusively on neuronal cells and examine them in detail. To do so, we first removed the immune/myeloid clusters from the study. Next, we used the mutual nearest neighbors (MNNs) method in conjugation with Seurat to integrate the datasets. For comparative analysis, we created two

integrated datasets comprising of adult and pediatric GBM patients. Table 2 shows the patient wise detail of the adult and pediatric groups. Both adult and pediatric groups clustered largely according to cell types and cell stages (figure 2 A, B). Comparison of adult and pediatric groups showed that a number of genes specific to cell cycle, neuronal and glial cells like *TOP2A*, *SLC1A2*, *SSR4*, *APOD*, *PLP1*, *CD74*, *RTN1*, *HMGB2*, *CHI3L1*, *SERPINE1* were commonly overexpressed in certain clusters in both the groups. (figure 2 B). Whilst SLC1A2, APOD, PLP1, CD74, and RTN1 are cell type specific proteins and expressed by astrocytes, oligodendrocytes, OPCs and endothelial cells respectively[25], TOP2A and HMGB2 expression is specific to transcriptional activation and cell cycle[26, 27]. CHI3L1 and SERPINE1 are both proteins related to cell differentiation and malignant transformation[28, 29]. Comparatively, Adult GBM dataset exhibited greater number of distinct clusters which might be reflective of higher cell type and disease stage diversity. The presence of specific clusters overexpressing TOP2A and SOX4 in both adult and pediatric groups indicated a similarity in gene expression profile of these clusters across groups. Cycling or actively dividing and quiescent stem cells are known to coexist in adult stem cell niches[30]. Hence, we assessed the expression distribution of known markers for proliferation (KI67and CD44) and maintenance of cellular plasticity (SOX11 and DCX). The expression of these genes coincided with clusters1,2 and 3 the adult group while in the pediatric group, all clusters except clusters 3 and 7 showed high expression of these genes (Figure 2 C). Interestingly, CD44 expression did not follow this pattern, its expression was confined to clusters unrelated to the expression of other markers. Indeed, recent reports have questioned the validity of CD44 as a reliable marker of proliferation in glioblastoma [20, 31]. As, *TOP2A*, *MKI67*, *DCX*,*SOX4* and *SOX11* genes are known to show high expression in cycling and Quiescent stem cells respectively[26, 32], we suspected the presence of similar niches within

the above mentioned clusters in the GBM groups. To verify this hypothesis, we decided to further analyze the cell type context of these clusters. Top 50 differentially expressed genes for both adult and pediatric groups are included in additional file 1.

Identification of types and cell cycle stages

To determine the presence of cycling and Quiescent cells in both groups, we sought to identify and distinguish these cells from mature neural and glial cells. We used single cell datasets of adult and embryonic brain, from Darmanis, Spyros, et al.[25] as reference dataset and performed unbiased cell type recognition using SingleR package (see methods). The parameters used are described in the methods section. The results confirmed the initial clustering based predictions and showed the presence of both cycling stem cell like (cGSC) and Quiescent stem cell like (qGSC) cells in patient samples from both groups (Figure 3 A). Comparatively, the proportion of cGSCs and qGSCs was much higher in the pediatric group. Specifically, qGSC population was markedly low in adult group. Patient samples BT-S2 and BT-S4 had no identified qGSCs, while only one cell could be identified as qGSC in PJ-32, while PJ-35 had highest number of both qGSCs and cGSCs in the adult group (Figure 3 D), probably because a large number of cells in the group are from PJ-35. Expectedly, both these populations were absent in the non-GBM metastatic tumor sample (GS-15), confirming their importance as key originators of GBM.

Analysis of cycling and quiescent cells

To do a comparative analysis of the gene expression patterns between the groups, differential gene expression analysis was performed using Model-based Analysis of Single Cell Transcriptomics (MAST) package from R. Significant differentially expressed genes for both qGSCs and cGSCs were compared with other cell types in both adult and pediatric groups (figure 3 B). In both groups, qGSCs had a markedly high expression of *DCX, SOX4, SOX11* and

DLL3 genes. While SOX4 and SOX 11 are both critical in the development and maintenance of neural pluripotent cells, DCX is an essential factor in neurogenesis in neuronal migration. DLL3 is a ligand for the Notch pathway and plays a pleotropic role in notch pathway regulation [32]. On the other hand, cGSCs in both groups were marked by overexpression of HMGB2, HSP90B1 and KPNA2 apart from TOP2A (figure 3 C). HMGB2 is a member of the high mobility protein family, functioning as a modulator of chromatin structure. However, recent study has shown its role in transition from quiescence to activated state in neuronal stem cells (NSCs) [27]. Similarly, HSP90B1, a member of the heat shock protein family, has a role in maintaining embryonic pluripotency [36], whereas KPNA2 is known to be associated with a number of cancers[37]. comparatively, in the adult group, qGSCs and cGSCs have a marked difference in their expression profiles, but less so in the pediatric group. A possible reason for this distinction perhaps is the fact that pediatric brain cells are primed for development.

This is also evident from the cell cycle stage prediction. Previous studies have shown that pluripotency in stem cells is intricately related with cell cycle stages. Whilst a short or truncated G1 (gap1) phase is considered a hallmark of pluripotent state, lengthening of G1 phase is observed when the cells enter cycling phase of rapid differentiation[33–35]. To further confirm the cellular states of these populations, we did a cell cycle state pseudotime prediction using Tricycle R package as described in the methods section. As, the Quiescent or G0 state is not exclusively defined in continuous cell state pseudotime embedding, we expected to find the qGSC cells to be predicted in the G0/G1 phase range, whilst cGSCs to be in G2/M state range. The results were as expected with qGSC almost exclusively in G0/early G1 state whilst cGSCs in late G1 to M states (Supplementary Figure S1). Interestingly, the distribution of cells within cellular states was continuous showing the presence of cells in intermediate states, indicating a

transition between qGSC and cGSC states. As, the stem cell like nature of these clusters was supported by both cell type and cell state analysis, we separated these clusters from other cell types and did a comparative analysis of underlying gene expression patterns between adult and pediatric groups to identify universal expression signatures of qGSCs and cGSCs.

Cluster analysis of cycling and quiescent cells

qGSCs and cGSCs from both groups were reclustered using the same unbiased approach of batch correction as described earlier. Five clusters were observed in both groups (Figure 4 A). Cell cycle pseudotime analysis of the clusters revealed a clear distinction in the cell cycle phase of the clusters in both groups (Figure 4 B). Differential gene expression analysis of the clusters not only revealed the genes involved in quiescence and activation in both groups (Supplementary Figure S2), but also showed a marked similarity between the clusters. we found that a number of highly overexpressed genes in clusters 1 and 2 from the adult group were also highly expressed in clusters 2 and 3 in the pediatric group (Figure 4 C). while the set of clusters overexpressing a large number of ribosomal genes (RP), especially *RPL23, RPL34, RPS3, RPS13, RPS29* also correlate with qGSC cells, the cGSC dominant set was marked by the overexpression of a number of high mobility group (HMG) genes including *HMGB1* and *HMGB2* and heterogeneous nuclear ribonucleoprotein (hnRNP) genes, notably *HNRNPA3* and *HNRNPD*. Recent research points to intricate relation between ribosomal activity and quiescence in stem cells[38, 39]. Indeed, high level of ribosomal presence can block stem cell differentiation. On the other hand, while HMG and TOP2A transcription regulators represent dynamic cell division, hnRNPs are key factors in pre-mRNA processing and transport. Indeed, based on gene expression pattern and cell cycle pseudotime analysis, a picture of sequential progression between the clusters is indicated with the overexpression of ribosomal proteins positively correlating with true

quiescence while the overexpression of HMG and hnRNPs indicating progression into cycling phase. In terms of disease model, it is likely that the mechanism of transition from quiescent to cycling states in GBMs remains similar to that of NSCs. In terms of patient samples, all patient samples from the adult group had the presence of ribosome overexpressing cluster (cluster 1), whereas in the pediatric group the ribosome overexpressing cluster (cluster 2) was absent in BT-830. Similarly, the hnRNP/ HMG overexpressing cluster in the adult group (cluster 2) was absent in PJ-32 (Figure 4 D). This may be because the number of cells included in the study may not represent the total tumoral heterogeneity or that the qGSC and cGSC states are interconvertible. Cluster wise differentially expressed genes for both adult and pediatric groups are included in additional file 2.

Gene ontology (GO) analysis of the clusters further confirmed our observations with the ribosome overexpressing cluster enriched in the biological process of cotranslational protein targeting to membrane or endoplasmic reticulum. The HMG and hnRNPs overexpressing clusters were enriched in cell cycle stages of DNA replication and sister chromatid separation. These clusters are likely representative of cycling cells from S to G2 phases. Interestingly, in the pediatric group, we found that the cluster 5 which comprised of a few cells from samples BT-1160, BT-749, BT-771 and MGH-85 was enriched for neuronal development and differentiation (Figure 5 A). However, we could not determine if this subpopulation is a transformed NSC precursor of a specific lineage.

Analysis of copy number variations (CNVs)

Copy number alterations (gain and/or loss) of the DNA are known to be associated with disease progression in various cancers[40]. Based on the gene expression patterns in the qGSCs and cGSCs, we reasoned that the variations in the chromosomal regions of these cells is likely

distinct from normal or differentiated neoplastic glial/neuronal cells. To confirm this hypothesis, we compared the CNV status of GSC (qGSC and cGSC) clusters with non-malignant and differentiated neoplastic cells. For comparison, we used the gene expression counts of 332 of adult normal brain cells from dataset: GSE67835 as reference[25]. The results did not show a clear similarity in CNV patterns of adult and pediatric groups, however, we did observe a pattern of copy number gain at chromosomes 19 and 11 in some patients in the pediatric group (BT-1160, BT-749, BT-771, MGH-85). In terms of pediatric group, we found a consistent CNV pattern in the GSC population (Figure 5 B) (see also supplementary Figure S3). Chromosomes 19 showed a copy number gain while chromosome10 showed a loss of copy number. Locus gain at chromosome 19 is relevant in this study's context because Chromosome 19 which has a high gene density, also harbors a large number of ribosomal genes [41]. While there seems to be a correlation between copy number alteration at chromosome 19 and ribosomal protein abundance, we could not verify this correlation in terms of causation. However, we consider this an interesting finding, which needs to be explored in detail in the future.

DISCUSSION

Glioblastomas present marked inter and intra tumoral heterogeneity which is a key hurdle in identifying tumorigenic cell populations and therefore designing robust therapeutic strategies to target them. Recent advances in single cell techniques has helped immensely in studying and intratumoral cellular niches. However, identification of GSCs which are considered to be the drivers of tumor progression and therapy resistance largely remains a challenge because they exist as a small population of cycling and quiescent cells within these niches. The dynamic nature of tumor microenvironment means that these cells show considerable phenotypic plasticity. This behavior of GSCs would suggest that a marker based strategy, although very informative in describing cellular state at a given time, is insufficient in identifying stem cell like tumor populations with marked cellular plasticity. The existence of GSC-like cells in proliferative and quiescent states within these niches is largely agreed upon, however, the identification of quiescent GSC-like population has remained a challenge. Through our study of scRNAseq data from a diverse panel GBM samples we have identified gene signature patterns uniquely associated with cycling and quiescent states in GBM cells.

A number of single cell sequencing based studies in recent years have focused on identifying GSCs[15, 18, 20, 22, 42], these studies support the theory that malignancy in glioblastoma is a function of a small number of progenitor like cells which may exist in neural, mesenchymal, oligodendrocyte or astrocyte like states. The results of this analysis show that although most of the cycling and quiescent GSCs do share similarities with endothelial and glial like cells, they have unique transcriptomic profiles which suggest that these cells are maintained in their own niches within the microenvironment surrounded or populated with mature endothelial, neural

and/or glial cells. It is important to underline the observation that while transformed cells from the non–GBM tumor (GS-15) had endothelial, microglia and oligodendrocyte like populations, they lacked both cGSC and qGSC like cells. This provides evidence that glioblastoma is inherently a disease of neural stem/progenitor cells.

A number of studies on cancer stem cells have largely focused on the expression of genes related to a few key transcription factors (Yamanaka factors), OCT4, SOX2, KLF4, and MYC[43] including NANOG which are well known markers for pluripotency. Similarly, makers for proliferation like MKI67 and CD44 are the focus of most of the research on proliferating and pluripotent malignant cells in GBM. However, such approaches overlook the molecular processes involved in maintaining quiescence or triggering proliferation. Indeed, we observed no correlation between CD44 expression and cGSCs. On the other hand, our analysis gives further evidence that SOX4, SOX11 and DCX do overexpress in cycling and quiescent GSCs, however, studying these markers alone won't explain the molecular process involved in maintaining quiescence or triggering proliferation.

The results of this analysis provides strong evidence that quiescent and cycling stem like cells in GBM share common molecular pathway to maintain quiescence. By comparing the difference in the gene expression profiles of qGSC and cGSC clusters, we have been able to captures the changes in cellular processes of the cells transitioning from quiescent to cycling state. Notably, the quiescent state is underlined by the overexpression of a number of ribosomal genes while the cycling state is marked by the overexpression of HMG and hnRNPs. This switch from ribosomal gene family to HMG and hnRNP genes suggests triggering of entry into cycling phase is accompanied by profound changes in cell physiology. The presence of migration and neuronal differentiation related genes like *S100B, VIM* and *SPARC* in a quiescent like separate cluster

(cluster 5) may show the presence of NSC like or progenitor cells. further research is required to understand their interaction with other clusters. In terms of patient groups, the results show that adult tumor samples had a much lower proportion of qGSCs, this is probably reflective of the differences in developmental state of brain. Importantly, the results suggest that qGSCs and cGSCs of NSC nature are a key feature of the glioblastoma tumor microenvironment. These populations are largely interconvertible but are maintained predominantly by the expression of genes distinct to these states.

The role of ribosomal proteins in stem cell maintenance is an area of active research[44]. Recent research on mouse NSCs has shown low protein synthesis rate as a hallmark of quiescent state[45]. This phenomenon of quiescent NSCs is probably due to reduced activity of the mTOR (mammalian target of rapamycin) kinase which acts as a key bridge linking ribosome biogenesis and protein synthesis to induction of pluripotency, self-renewal and differentiation in adult stem cells[46, 47]. Experimentally, it has been shown that knockdown of 4E-BP1 (an mTOR target) promotes differentiation in mouse NSCs[48], mTOR signaling drops when the cell exits cell cycle, leading to suppression of ribosome synthesis, controlling NSC differentiation[49]. It is thus likely that a number of ribosomal proteins are maintained in the quiescent cells to trigger differentiation, thus ensuring and effective transformation of the stem cell state upon receiving environmental signal[47, 48, 50, 51]. While this is a possible theory for the overexpression of ribosomal genes in quiescent cells, further studies are needed to understand this phenomenon in glioblastoma.

In conclusion, this study provides vital insight in the expression profile of cycling and quiescent like cells in glioblastoma. Therapy designs targeting these cells holds great promise in the treatment of GBM patients because studies have shown that these cells are key to developing

therapy resistance, migration and proliferation. Targeting quiescent GSCs is critical to overcome tumor relapse. This work is an important step in understanding the molecular processes that govern the quiescent and cycling states in GBMs.

MATERIALS AND METHODS

Data resource and selection

All Single-Cell RNA-Seq raw read count matrices and metadata files (wherever available) were were downloaded from Gene Expression Omnibus (GEO) repository. Specifically, gene/cell expression counts from datasets GSE84465[17],GSE117891[18],GSE131928[52] and GSE103224[19] downloaded. The expression matrix for patients included in the study was then curated from the raw expression files. Raw counts for patient data from GSE131928 was not made available by the authors. Log2 transformed count (available) was used instead.

Data filtration and normalization

All datasets were filtered and analyzed using Seurat V4[53]. Raw data matrix was first filtered using the slandered Seurat protocol to remove possible low quality cells, cells with <200 or >3000 transcripts were excluded from the analysis. In addition, cells of poor quality, recognized as cells with >5% of their transcripts coming from mitochondrial genes, were excluded from the downstream analysis.

Clustering techniques

Primary clustering of the datasets was done following Seurat protocol. Briefly, after filtering and removal of mitochondrial counts, the data was log normalized and highly variable features were calculated (for this study, we kept the nfeature setting to 6000). Next, the data was scaled before performing linear dimensional reduction. High variable principal components were selected based on percentage variance. Next, the K nearest neighbor graph was constructed based on

calculated principal components and clustered. Finally, the dimensional reduction and visualization was done using Uniform Manifold Approximation and Projection (UMAP).

Differential gene expression analysis

Analysis of differentially expressed genes for each clusters was done by implementing Model-based Analysis of Single Cell Transcriptomics (MAST) package with Seurat[54].

Integration of datasets

For integrated analysis, following initial Seurat protocol, the fast mutual nearest neighbors (fastMNN) R/Bioconductor package was applied to correct for differences between data sets (batch effect correction)[55]. Clustering was done using default parameters.

Reference based cell type identification

Cell type identification was performed using singleR[56]. which is an R/Bioconductor package to perform unbiased cell type recognition from single-cell RNA sequencing data, by using reference datasets of pure cell types to identify the cell type of individual single cells independently. Here we used the dataset from Darmanis, Spyros, et al.[25] as reference dataset for cell type identification. Cells not recognized as either of the cell types (NAs) were removed from further analysis.

Cell cycle trajectory inference

Cell cycle phase of the integrated datasets was inferred using the Tricycle R/Bioconductor package, which uses a fixed reference dataset to infer cell cycle phase of the test dataset[57]. Here, we used the reference dataset provided in tricycle with default parameters to infer cell cycle positions of cells in integrated data. The inferred positions were then project on to the UMAP for visualization. The estimated cell cycle position is bound between 0 and 2pi. The cycle

positions approximately relate to theta as: 0.25pi-1.75pi to G0/G1 stage, 0.5pi to start of S stage, pi to start of G2M stage and 1.5pi the middle of M stage.

CNV analysis

To compare the copy number variations between clusters and datasets, we used CONICSmat (Copy-Number Analysis in Single-Cell RNA-Sequencing from an expression matrix) R package which compares average gene expression of genes within a region to calculate the variance in copy number across samples[58]. Although reference data is not explicitly required, yet, for added certainty, we used normal brain expression matrix of 322 normal brain cells from Darmanis, Spyros, et al[25]. Analysis was done following the default protocol.

Gene Ontology (GO) analysis

Up to 50 (wherever possible) overexpressing genes for each cluster were analyzed for the enrichment of associated GO terms. Top 5 terms were selected based on fold change and represented graphically.

Statistical analysis

All data analysis was performed with R. Specific packages used are mentioned in the above sections.

FIGURE LEGENDS

Figure 1: Identification of Immune and neural cells clusters

(A) UMAP representation of clusters of all datasets. Immune cells are colored yellow; neural cells are colored blue, for GSE117891, blue represents transformed cluster, yellow-immune and green, neuronal-like. (B) Bar plot showing cluster wise expression levels of immune cell markers, bars are color coded according to the clusters in A. (C) Bar plot showing expression levels of neuronal markers, bars are color coded according to clusters in A.(D) Bar plot showing expression levels of EGFR, bars are colored red. All expression values are log transformed.

Figure 2: Cell type comparison of Adult and pediatric groups

(A) UMAP representation of Adult and Pediatric groups. Clusters are numbered and color coded. (B) Violin plot of cluster wise expression distribution of cell cycle and neural cell type markers. (C) UMAP representation of expression distribution of markers for quiescence, proliferation and migration in adult and pediatric clusters. Color transition represents expression levels with high expression represented in deep blue. All expression values are log transformed.

Figure 3: Group wise comparison of identified cell types

(A) Heat map of cluster wise expression comparison of markers for cell types in adult and pediatric groups. Cells with no strong matches are marked as NA. (B) Violin plot of top nine overexpressed genes in the cycling cells common to both adult and pediatric groups. (C) Violin plot of top eight overexpressed genes in the quiescent cells common to both adult and pediatric groups. (D) Patient sample wise composition of cell types in adult and pediatric groups. All expression values are log transformed.

Figure 4: Group wise comparison of cycling and quiescent clusters

(A) UMAP representation of integrated clustering of cycling and quiescent cells in Adult and Pediatric groups. Clusters are numbered and color coded. (B) UMAP representation of cell cycle status of the clusters from A. Theta values correspond cell cycle stages as follows: 0.25-1.75π ~ G0/G1 stage, 0.5π ~ start of S stage, π ~ start of G2M stage and 1.5π ~ middle of M stage. (C) Cluster wise composition of cycling quiescent populations in patient samples from adult and pediatric groups.

Figure 5: Pathway enrichment and copy number variations in cycling and quiescent cells

(A) Bar plots of top five biological processes enriched in cycling and quiescent clusters based on common overexpressed genes in adult and pediatric groups. log fold enrichment of biological processes is represented as bars. (B) Representative images of inferred CNVs in cycling and quiescent cells from one pediatric patient sample (BT-1160) compared to normal cells.

LIST OF ABBRIVIATIONS

GBM: Glioblastoma

GSC: Glioblastoma stem cell

NSC: Neural stem cell

UMAP: Uniform Manifold Approximation and Projection

cGSC: Cycling glioblastoma stem cell

qGSC: Quiescent glioblastoma stem cell

HMG: High mobility group protein

hnRNP: Heterogeneous nuclear ribonucleoprotein

ADDITIONAL FILES AND SUPPLEMENTARY FIGURES

Additional File 1: Group wise list of differentially expressed genes according to respective clusters for Figure 2 A.

Additional File 2: Group wise list of differentially expressed genes and common genes according to respective clusters for Figure 4.

Supplementary figure S1: Cell Cycle pseudotime representation of unclustered cycling and quiescent GSCs from adult and pediatric groups.

Supplementary figure S2: Dot plot representation of top ten genes per cluster for figure 4.

Supplementary figure S: CNV profiles of GSCs from pediatric group patient samples.

563

564 # REFERENCES

565 1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating
566 the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J
567 Cancer. 2019;144:1941–53.

568 2. Wesseling P, Capper D. WHO 2016 Classification of gliomas. Neuropathology and Applied
569 Neurobiology. 2018;44:139–50.

570 3. Berger K, Turowski B, Felsberg J, Malzkorn B, Reifenberger G, Steiger H-J, et al. Age-
571 stratified clinical performance and survival of patients with IDH-wildtype glioblastoma
572 homogeneously treated by radiotherapy with concomitant and maintenance temozolomide. J
573 Cancer Res Clin Oncol. 2021;147:253–62.

574 4. de Weille J. On the Genesis of Neuroblastoma and Glioma. International Journal of Brain
575 Science. 2014;2014:e217503.

576 5. Suvà ML, Tirosh I. The Glioma Stem Cell Model in the Era of Single-Cell Genomics. Cancer
577 Cell. 2020;37:630–6.

578 6. Brescia P, Ortensi B, Fornasari L, Levi D, Broggi G, Pelicci G. CD133 is essential for
579 glioblastoma stem cell maintenance. Stem Cells. 2013;31:857–69.

580 7. Beier D, Hau P, Proescholdt M, Lohmeier A, Wischhusen J, Oefner PJ, et al. CD133+ and
581 CD133− Glioblastoma-Derived Cancer Stem Cells Show Differential Growth Characteristics and
582 Molecular Profiles. Cancer Res. 2007;67:4010–5.

583 8. Ogden AT, Waziri AE, Lochhead RA, Fusco D, Lopez K, Ellis JA, et al. Identification of
584 A2B5+CD133- tumor-initiating cells in adult human gliomas. Neurosurgery. 2008;62:505–14;
585 discussion 514-515.

586 9. Anido J, Sáez-Borderías A, Gonzàlez-Juncà A, Rodón L, Folch G, Carmona MA, et al. TGF-β
587 Receptor Inhibitors Target the CD44high/Id1high Glioma-Initiating Cell Population in Human
588 Glioblastoma. Cancer Cell. 2010;18:655–68.

589 10. Paugh BS, Zhu X, Qu C, Endersby R, Diaz AK, Zhang J, et al. Novel Oncogenic PDGFRA
590 Mutations in Pediatric High-Grade Gliomas. Cancer Res. 2013;73:6219–29.

591 11. Schultz C, Lemke N, Ge S, Golembieski WA, Rempel SA. Secreted Protein Acidic and Rich
592 in Cysteine Promotes Glioma Invasion and Delays Tumor Growth in Vivo. Cancer Res.
593 2002;62:6270–7.

594 12. Shmelkov SV, Butler JM, Hooper AT, Hormigo A, Kushner J, Milde T, et al. CD133
595 expression is not restricted to stem cells, and both CD133+ and CD133– metastatic colon cancer
596 cells initiate tumors. J Clin Invest. 2008;118:2111–20.

597 13. Wang J, Sakariassen PØ, Tsinkalovsky O, Immervoll H, Bøe SO, Svendsen A, et al. CD133
598 negative glioma cells form tumors in nude rats and give rise to CD133 positive cells. Int J
599 Cancer. 2008;122:761–8.

600 14. Jung E, Osswald M, Ratliff M, Dogan H, Xie R, Weil S, et al. Tumor cell plasticity,
601 heterogeneity, and resistance in crucial microenvironmental niches in glioma. Nat Commun.
602 2021;12:1014.

603 15. Suvà ML, Rheinbay E, Gillespie SM, Patel AP, Wakimoto H, Rabkin SD, et al.
604 Reconstructing and reprogramming the tumor propagating potential of glioblastoma stem-like
605 cells. Cell. 2014;157:580–94.

606 16. Vinay DS, Ryan EP, Pawelec G, Talib WH, Stagg J, Elkord E, et al. Immune evasion in
607 cancer: Mechanistic basis and therapeutic strategies. Seminars in Cancer Biology.
608 2015;35:S185–98.

609 17. Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, et al. Single-
610 Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human
611 Glioblastoma. Cell Rep. 2017;21:1399–410.

612 18. Yu K, Hu Y, Wu F, Guo Q, Qian Z, Hu W, et al. Surveying brain tumor heterogeneity by
613 single-cell RNA-sequencing of multi-sector biopsies. Natl Sci Rev. 2020;7:1306–18.

614 19. Yuan J, Levitin HM, Frattini V, Bush EC, Boyett DM, Samanamud J, et al. Single-cell
615 transcriptome analysis of lineage diversity in high-grade glioma. Genome Medicine. 2018;10:57.

616 20. Dirkse A, Golebiewska A, Buder T, Nazarov PV, Muller A, Poovathingal S, et al. Stem cell-
617 associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the
618 microenvironment. Nat Commun. 2019;10:1787.

619 21. Guo M, Peng Y, Gao A, Du C, Herman JG. Epigenetic heterogeneity in cancer. Biomarker
620 Research. 2019;7:23.

621 22. Giraddi RR, Chung C-Y, Heinz RE, Balcioglu O, Novotny M, Trejo CL, et al. Single-Cell
622 Transcriptomes Distinguish Stem Cell State Changes and Lineage Specification Programs in
623 Early Mammary Gland Development. Cell Rep. 2018;24:1653-1666.e7.

624 23. Summers KL, Hock BD, McKenzie JL, Hart DNJ. Phenotypic Characterization of Five
625 Dendritic Cell Subsets in Human Tonsils. Am J Pathol. 2001;159:285–95.

626 24. Campanelli JT, Sandrock RW, Wheatley W, Xue H, Zheng J, Liang F, et al. Expression
627 profiling of human glial precursors. BMC Dev Biol. 2008;8:102.

628 25. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, et al. A survey of human
629 brain transcriptome diversity at the single cell level. PNAS. 2015;112:7285–90.

630 26. Thakurela S, Garding A, Jung J, Schübeler D, Burger L, Tiwari VK. Gene regulation and
631 priming by topoisomerase IIα in embryonic stem cells. Nat Commun. 2013;4:2478.

632 27. Kimura A, Matsuda T, Sakai A, Murao N, Nakashima K. HMGB2 expression is associated
633 with transition from a quiescent to an activated state of adult neural stem cells. Dev Dyn.
634 2018;247:229–38.

635 28. Rehli M, Niller H-H, Ammon C, Langmann S, Schwarzfischer L, Andreesen R, et al.
636 Transcriptional regulation of CHI3L1, a marker gene for late stages of macrophage
637 differentiation. J Biol Chem. 2003;278:44058–67.

638 29. Li S, Wei X, He J, Tian X, Yuan S, Sun L. Plasminogen activator inhibitor-1 in cancer
639 research. Biomedicine & Pharmacotherapy. 2018;105:83–94.

640 30. Li L, Clevers H. Coexistence of Quiescent and Active Adult Stem Cells in Mammals.
641 Science. 2010;327:542–5.

642 31. Wang H-H, Liao C-C, Chow N-H, Huang LL-H, Chuang J-I, Wei K-C, et al. Whether CD44
643 is an applicable marker for glioma stem cells. Am J Transl Res. 2017;9:4785–806.

644 32. Codega P, Silva-Vargas V, Paul A, Maldonado-Soto AR, DeLeo AM, Pastrana E, et al.
645 Prospective Identification and Purification of Quiescent Adult Neural Stem Cells from Their In
646 Vivo Niche. Neuron. 2014;82:545–59.

647 33. Pauklin S, Vallier L. The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity.
648 Cell. 2013;155:135–47.

649 34. White J, Dalton S. Cell cycle control of embryonic stem cells. Stem Cell Rev. 2005;1:131–8.

650 35. Cheshier SH, Morrison SJ, Liao X, Weissman IL. In vivo proliferation and cell cycle kinetics
651 of long-term self-renewing hematopoietic stem cells. PNAS. 1999;96:3120–5.

652 36. Bradley E, Bieberich E, Mivechi NF, Tangpisuthipongsa D, Wang G. Regulation of
653 embryonic stem cell pluripotency by heat shock protein 90. Stem Cells. 2012;30:1624–33.

654 37. Han Y, Wang X. The emerging roles of KPNA2 in cancer. Life Sciences. 2020;241:117140.

655 38. Sharifi S, da Costa HFR, Bierhoff H. The circuitry between ribosome biogenesis and
656 translation in stem cell function and ageing. Mechanisms of Ageing and Development.
657 2020;189:111282.

658 39. van Velthoven CTJ, Rando TA. Stem Cell Quiescence: Dynamism, Restraint, and Cellular
659 Idling. Cell Stem Cell. 2019;24:213–25.

660 40. Shlien A, Malkin D. Copy number variations and cancer. Genome Med. 2009;1:62.

661 41. Kenmochi N, Kawaguchi T, Rozen S, Davis E, Goodman N, Hudson TJ, et al. A map of 75
662 human ribosomal protein genes. Genome Res. 1998;8:509–23.

663 42. Pang B, Xu J, Hu J, Guo F, Wan L, Cheng M, et al. Single-cell RNA-seq reveals the invasive
664 trajectory and molecular cascades underlying glioblastoma progression. Mol Oncol.
665 2019;13:2588–603.

666 43. Liu X, Huang J, Chen T, Wang Y, Xin S, Li J, et al. Yamanaka factors critically regulate the
667 developmental signaling network in mouse embryonic stem cells. Cell Res. 2008;18:1177–89.

668 44. Sanchez CG, Teixeira FK, Czech B, Preall JB, Zamparini AL, Seifert JRK, et al. Regulation
669 of Ribosome Biogenesis and Protein Synthesis Controls Germline Stem Cell Differentiation. Cell
670 Stem Cell. 2016;18:276–90.

671 45. Urbán N, van den Berg DLC, Forget A, Andersen J, Demmers JAA, Hunt C, et al. Return to
672 quiescence of mouse neural stem cells by degradation of a proactivation protein. Science.
673 2016;353:292–5.

674 46. Rodgers JT, King KY, Brett JO, Cromie MJ, Charville GW, Maguire KK, et al. mTORC1
675 controls the adaptive transition of quiescent stem cells from G0 to GAlert. Nature.
676 2014;510:393–6.

677 47. Gan B, DePinho R. mTORC1 signaling governs hematopoietic stem cell quiescence. Cell
678 Cycle. 2009;8:1003–6.

679 48. Guo J, Zhang T, Guo Y, Sun T, Li H, Zhang X, et al. Oocyte stage-specific effects of MTOR
680 determine granulosa cell fate and oocyte quality in mice. PNAS. 2018;115:E5326–33.

681 49. mTORC1 controls the adaptive transition of quiescent stem cells from G0 to GAlert | Nature.
682 https://www.nature.com/articles/nature13255. Accessed 9 Dec 2021.

683 50. Zhang Q, Shalaby NA, Buszczak M. Changes in rRNA transcription influence proliferation
684 and cell fate within a stem cell lineage. Science. 2014;343:298–301.

685 51. Blair JD, Hockemeyer D, Doudna JA, Bateup HS, Floor SN. Widespread Translational
686 Remodeling during Human Neuronal Differentiation. Cell Rep. 2017;21:2005–16.

687 52. Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, et al. An Integrative Model of
688 Cellular States, Plasticity, and Genetics for Glioblastoma. Cell. 2019;178:835-849.e21.

689 53. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis
690 of multimodal single-cell data. Cell. 2021;184:3573-3587.e29.

691 54. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: a flexible
692 statistical framework for assessing transcriptional changes and characterizing heterogeneity in
693 single-cell RNA sequencing data. Genome Biology. 2015;16:278.

694 55. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-
695 sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol.
696 2018;36:421–7.

697 56. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung
698 single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol.
699 2019;20:163–72.

700 57. Zheng SC, Stein-O'Brien G, Augustin JJ, Slosberg J, Carosso GA, Winer B, et al. Universal
701 prediction of cell cycle position using transfer learning. 2021.

702 58. Müller S, Cho A, Liu SJ, Lim DA, Diaz A. CONICS integrates scRNA-seq with DNA
703 sequencing to map gene expression to tumor sub-clones. Bioinformatics. 2018;34:3217–9.

704

Figure 1

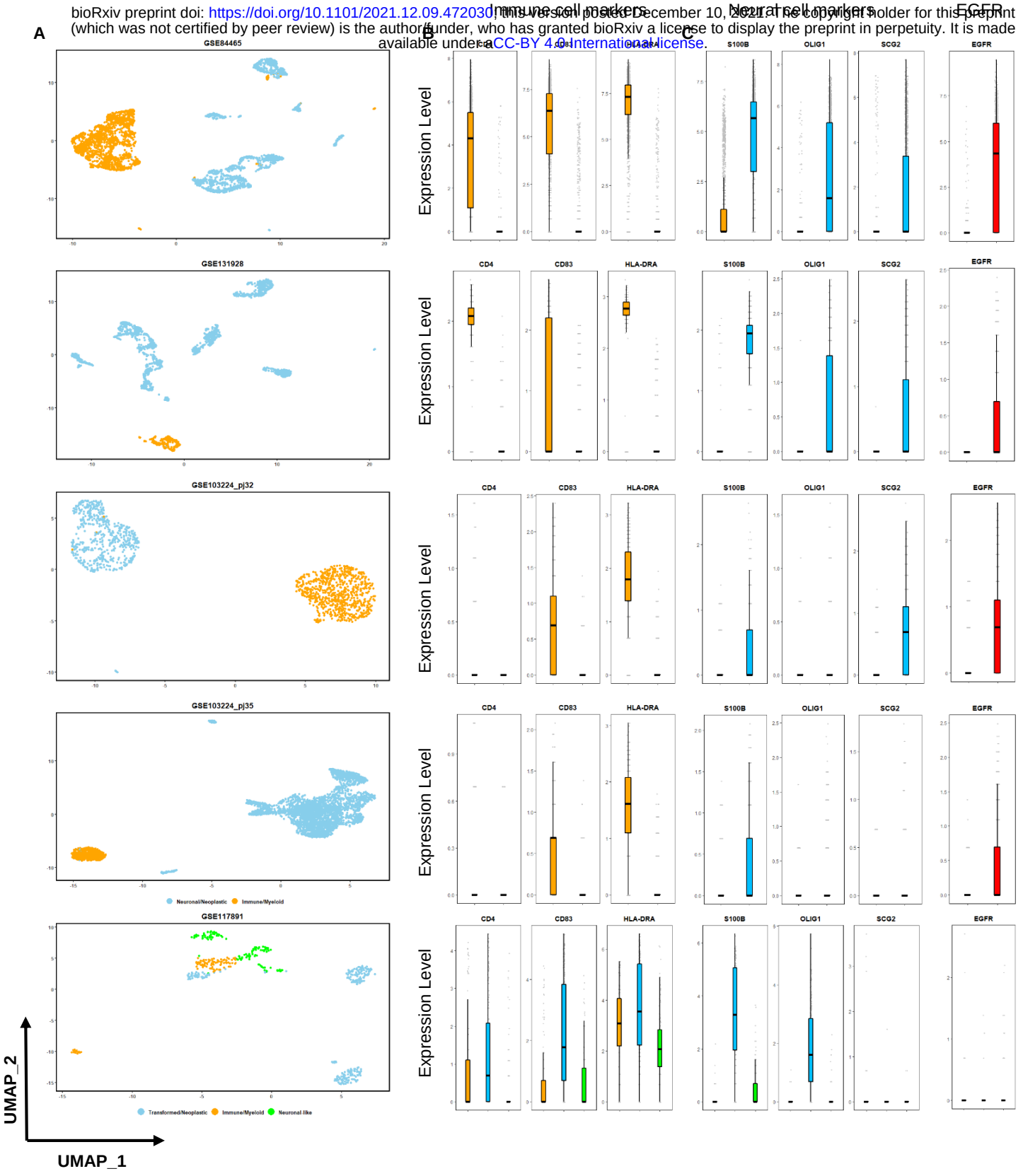


Figure 2

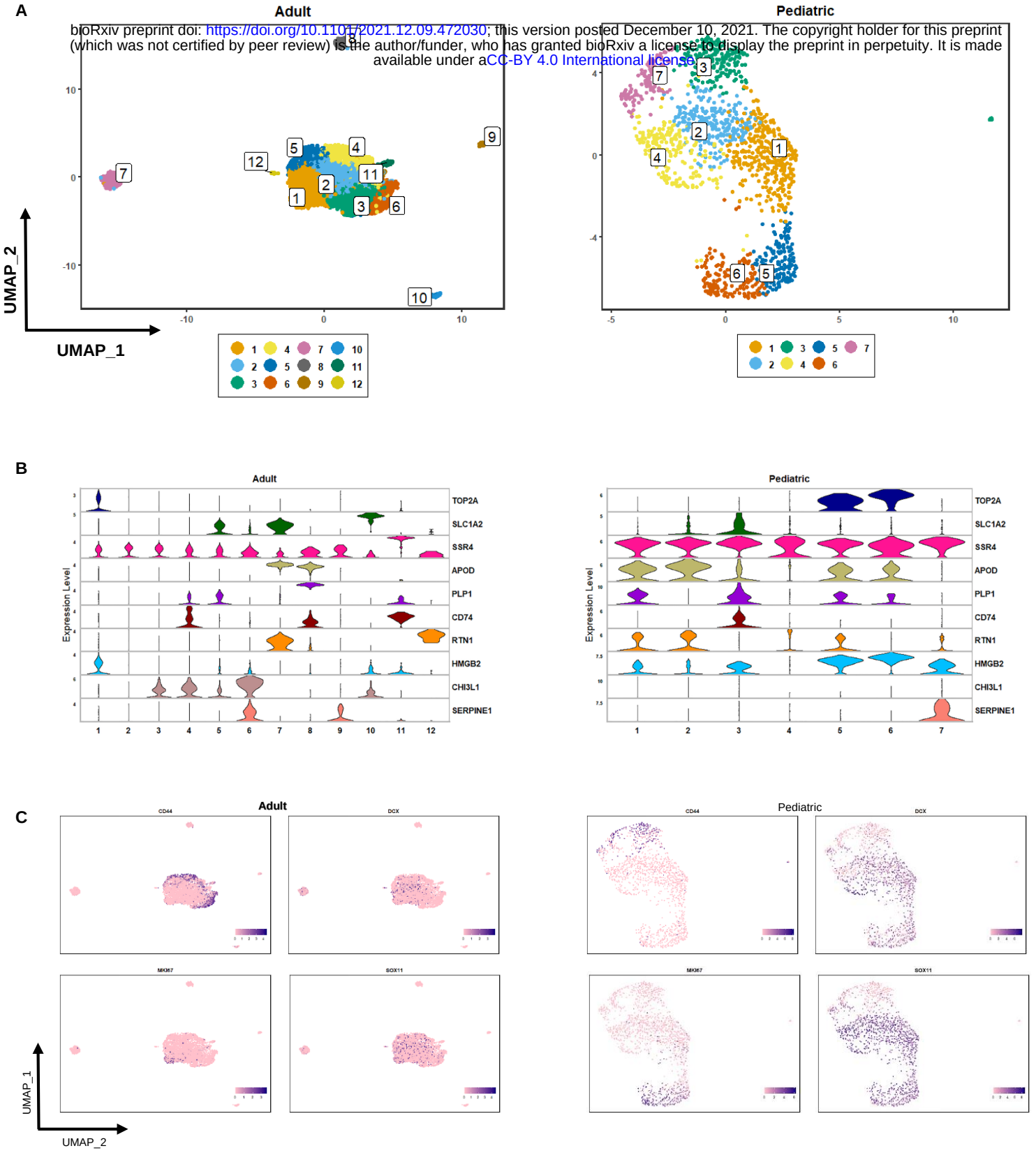


Figure 3

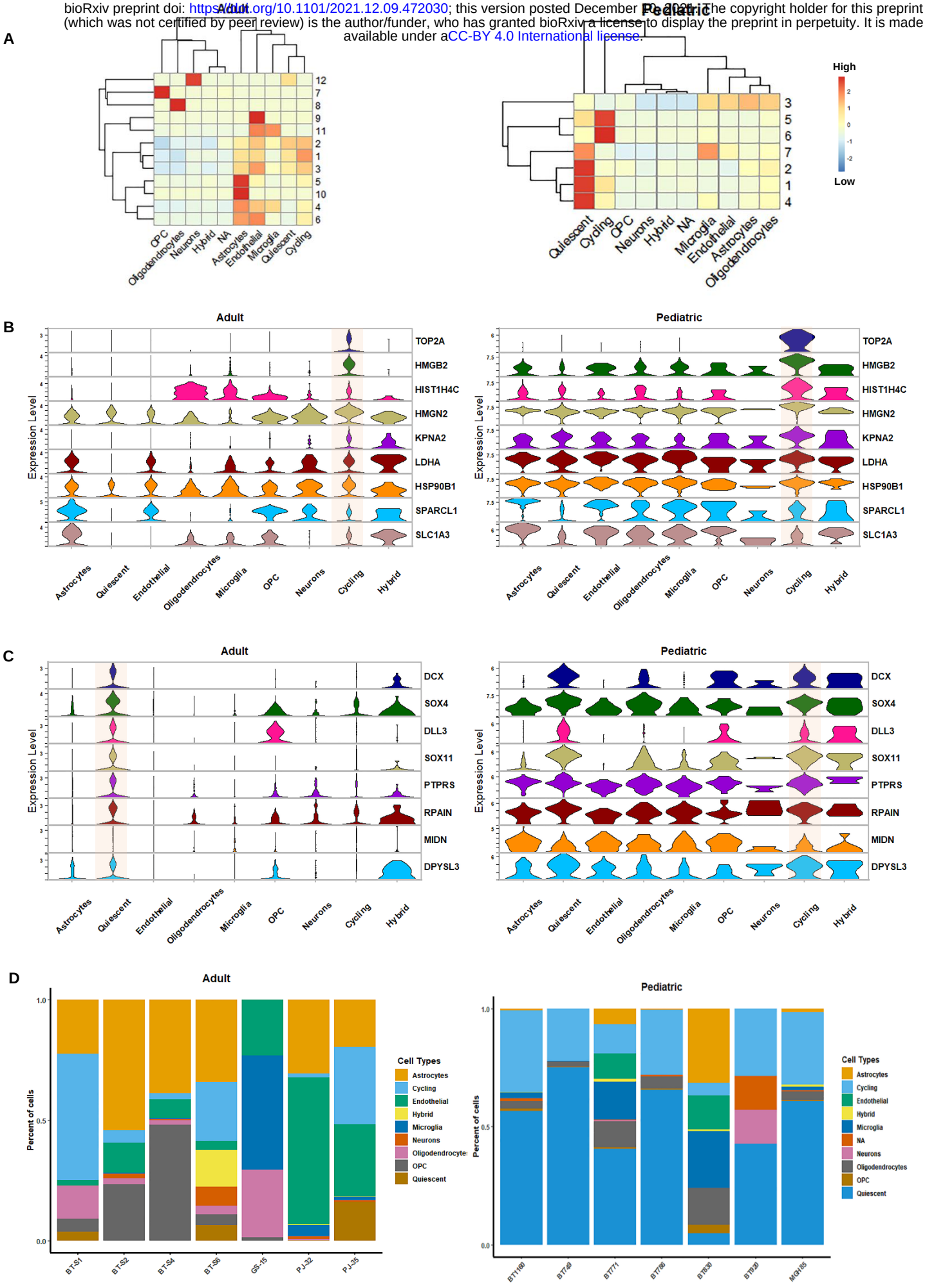


Figure 4

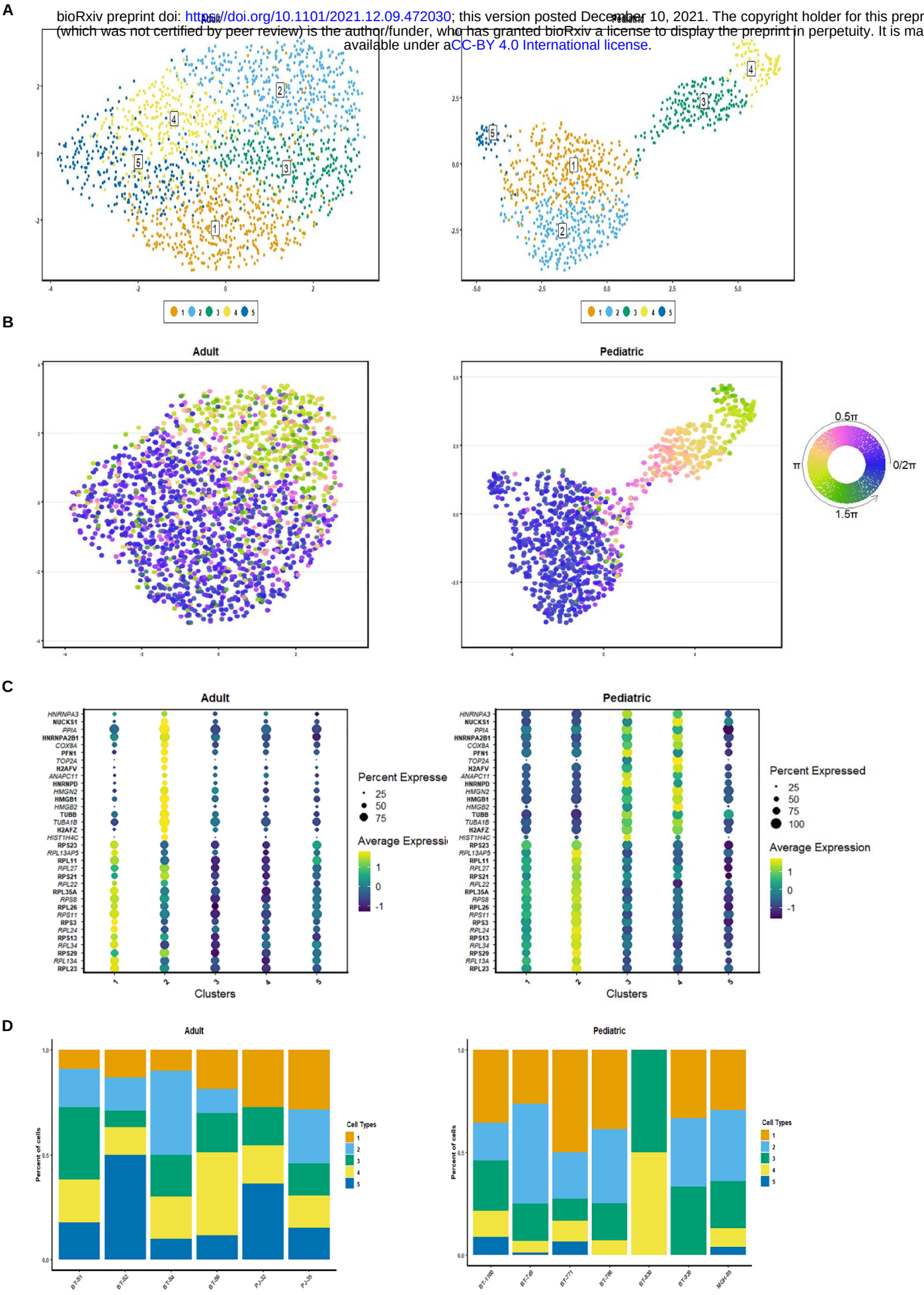
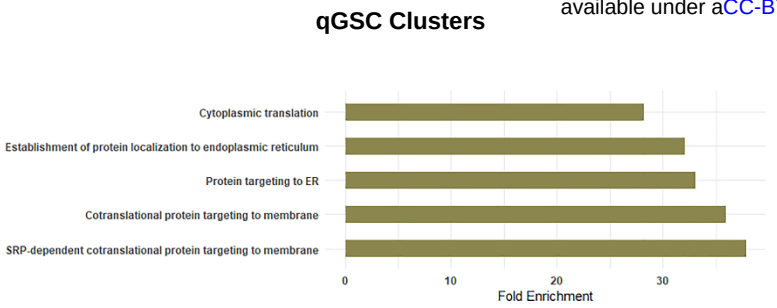
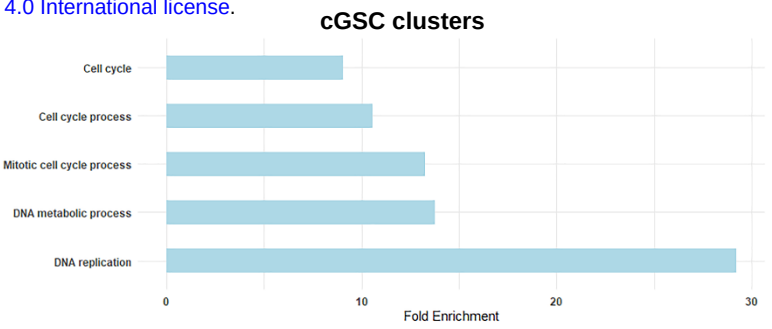
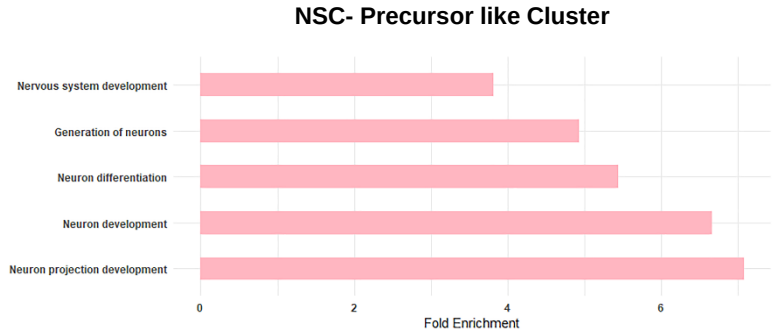


Figure 5

A







B