Pan-glioma analyses reveal species- and tumor-specific regulation of neuron-glioma synapse genes by lncRNAs

Wei Xiong[1], Xuecong Zhang[1], Bin Peng[1], Hao Zhu[1]*, Lijin Huang[2]* and Sha He[1]*

[1]Bioinformatics Section, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China, [2]Neurosurgery Department, The Third Affiliated Hospital, Southern Medical University, Guangzhou, China

Gliomas are highly heterogeneous and aggressive. Malignant cells in gliomas can contact normal neurons through a synapse-like structure (called neuron-to-glioma synapse, NGS) to promote their proliferation, but it is unclear whether NGS gene expression and regulation show species- and tumor-specificity. This question is important in that many anti-cancer drugs are developed upon mouse models. To address this question, we conducted a pan-glioma analysis using nine scRNA-seq datasets from humans and mice. We also experimentally validated the key element of our methods and verified a key result using TCGA datasets of the same glioma types. Our analyses revealed that NGS gene expression and regulation by lncRNAs are highly species- and tumor-specific. Importantly, simian-specific lncRNAs are more involved in NGS gene regulation than lncRNAs conserved in mammals, and transgenic mouse gliomas have little in common with PDX mouse models and human gliomas in terms of NGS gene regulation. The analyses suggest that simian-specific lncRNAs are a new and rich class of potential targets for tumor-specific glioma treatment, and provide pertinent data for further experimentally and clinically exmining the targets.

KEYWORDS

gliomas, neuron-glioma synapse, pan-glioma analysis, lncRNA, species-specific

1 Introduction

In humans and mice, 95% of protein-coding genes are conserved, but substantial lncRNA genes are species-specific; moreover, lncRNA expression is highly tissue-specific (Derrien et al., 2012; Necsulea et al., 2014; Lin et al., 2019). Further, many simian-specific lncRNA genes are expressed only in the brain (Derrien et al., 2012). These three factors make gene regulation by lncRNAs in the brain highly species- and tissue-specific. Since most drugs developed upon mouse models fail to cure human patients, revealing the species- and tissue-

Abbreviations: AA, anaplastic astrocytoma; DBD, DNA binding domains; DBS, DNA binding site; DEG, differentially expressed gene; GBM, glioblastoma; GO, Gene Ontology; MB, medulloblastoma; NGS, neuron-glioma synapse; NGS gene, neuron-glioma synapse related gene; OD, 1p/19q codeleted oligodendroglioma; PDX, patient-derived xenografts model; p-GBM, pediatric GBM; TTS, triplex-targeting site.

specificities and the underlying implications upon big data is important for glioma diagnosis and therapy. However, to what extent gene expression regulation by lncRNAs is species- and tissue-specific is poorly understood, especially in the brain and brain tumors.

Gliomas are highly aggressive and heterogeneous, and the two properties greatly influence glioma therapy and survival. High-grade gliomas are especially aggressive, with a 5-year survival rate of < 5% (Ostrom et al., 2014). Currently, all therapies (including surgical resections, radiotherapy, and chemotherapy) have very limited effects (Amarouch and Mazeron, 2005; Reitman et al., 2018), making identifying new therapeutic targets a pressing need. Increasing evidence indicates that interactions between cells in the tumor microenvironment (TME) play vital roles in glioma generation, migration, and responses to chemotherapy and immunotherapy (Osswald et al., 2015; Wang et al., 2017; Wu and Dai, 2017). A specific kind of intercellular interactions is between glioma cells and neuronal cells. NLGN3, a protein normally on the neuron surface, can be secreted as a mitogen and promote the growth and infiltration of glioma cells through the PI3K-mTOR pathway (Venkatesh et al., 2015). Multiple studies have confirmed that glioma cells can contact neurons through a synapse-like structure called neuron-to-glioma synapse (NGS), by which malignant cells mimic normal neurons and hijack potassium currents to promote their aggression and proliferation (Venkataramani et al., 2019; Venkataramani et al., 2022b; Venkatesh et al., 2019). Such synapses also form between the brain-metastatic breast cancer cells and neural cells to facilitate breast-to-brain metastasis (Zeng et al., 2019). Studies also demonstrate that NGS and neuronal activity-dependent paracrine contribute to forming neuron-tumor networks in gliomas (Venkataramani et al., 2022a). These discoveries indicate that the NGS is an important mechanism underlying glioma aggression and should be a key target for glioma therapy.

According to experimental studies, NGS formation and function rely on a set of genes involved in microtube formation, synaptogenesis, neurotransmitter receptors, and neuronal paracrine (hereafter termed NGS genes) (Venkatesh et al., 2019). The knockout of NLGN3 (an NGS gene) can decrease the invasion of glioma in vivo in mouse models (Venkatesh et al., 2015), supporting that NGS genes can be therapeutic targets. However, researchers who use transgenic and patient-derived xenograft (PDX) mouse models to study tumors find that many drugs tested successfully in mice fail in human clinical trials (Mak et al., 2014; Gould et al., 2015). A previous study focusing on PDGF signaling in glioblastoma in humans, rats, mice, and dogs revealed that glioblastoma in these species shows highly distinct gene expression (Connolly et al., 2018). To explore NGS genes as targets for gliomas, more studies addressing the two questions are needed: whether NGS gene expression shows species- and tumor-specificity, and how the specificity is regulated, especially by lncRNAs (Yao et al., 2015; Alinejad-Rokny et al., 2020).

Pan-cancer analyses are highly valued in cancer studies. Many pan-cancer analyses have revealed shared features of different tumors (Ma et al., 2018; Campbell et al., 2020). Cross-species cancer analyses, which try to identify shared and species-specific features of tumors, have also drawn attention. Reported cross-species cancer analyses focused either on a specific aspect, such

as PDGF signaling (Connolly et al., 2018), cancer genome (Robles-Espinoza and Adams, 2014), and cancer resistance-associated genes (Ulhas Nair et al., 2022), or on a specific cancer, such as pancreatic cancer (Elyada et al., 2019). To make the most of published data and to address the above-mentioned two questions, cross-species pan-gliomas analyses are required. In this study, we examined NGS genes and their potential regulation by lncRNAs using nine single-cell RNA-sequencing (scRNA-seq) datasets covering 5 glioma types in humans and mice. The examined gliomas include oligodendroglioma (Tirosh et al., 2016), anaplastic astrocytoma (Venteicher et al., 2017), medulloblastoma (Hovestadt et al., 2019; Cheng et al., 2020), glioblastoma (Neftel et al., 2019; Weng et al., 2019), and diffuse H3K27M glioma (Filbin et al., 2018), which range from the WTO II grade to the WTO IV grade. We found that NGS gene expression is highly tumor-specific and that multiple lncRNAs, especially simian-specific ones, critically regulate tumor-specific NGS genes. Importantly, we found that mouse and human gliomas have very limit similarities in NGS regulation by lncRNAs, including that transgenic mouse gliomas have little in common with PDX mouse models and human gliomas in terms of NGS gene regulation. We experimentally validated the key element of our methods and verified a key result using TCGA datasets of three glioma types. The combined TCGA and scRNA-seq data analysis reveals that malignant cells with NGS are indeed associated with worse survival. These results help identify the mechanisms behind gliomas' species- and tumor-specificity and provide pertinent data for further experimentally and clinically identifying tumor-specific targets for glioma treatment.

2 Materials and methods

2.1 Data collection and pre-processing

Nine scRNA-seq datasets of gliomas (GSE131928, GSE89567, GSE70630, GSE102130, GSE119926, GSE122871, and GSE150752) were downloaded from the GEO website (Supplementary Table S1). GSE131928 contains two datasets that are adult and pediatric glioblastoma. GSE119926 contains two datasets that are human medulloblastoma and medulloblastoma PDX in mice. These datasets were generated by Drop-seq (GSE122871), 10X Genomics (GSE150752), and Smart-seq/Smart-seq2 (all others).

Batch effects may exist in samples of different studies and different sequencing protocols. Whether a researcher chooses to eliminate or not eliminate batch effects depends on the study's primary purpose and the data's property. Because gliomas are characterized by intra- and inter-tumoral heterogeneity, eliminating batch effects by integrating data using specific methods may risk removing real biological signals (including glioma heterogeneity). To preserve as much heterogeneity as possible between glioma samples, we did not remove the batch effect. However, we normalized each dataset using the Seurat (v3.6.1) package as follows (Satija et al., 2015).

For smart-seq datasets, transcripts per million (TPM) were first log-transformed using the equation $E_{ij} = \log_2\left(\frac{TPM_{ij}}{10} + 1\right)$ where "i" refers to gene and "j" refers to cell. After the transformation, we used the CreateSeuratObject function in Seurat to read the E_{ij} expression matrix. For datasets generated by Drop-seq or 10X Genomics, we

used the *read10X* function to read the UMI data directly. For all datasets generated by the three sequencing protocols, cells with *nFeature_RNA* < 3,000 and *nCount_RNA* < 4,000 were removed, all mitochondrial genes were removed, and each dataset was normalized using the *NormalizationData* function (with default parameters) to eliminate the effect of library size across cells. The numbers of remaining cells and genes in each dataset are listed in the (Supplementary Table S1). By default, the default parameters were used in the analyses described in the following sub-sections.

2.2 Cell clustering

For each dataset, the following steps were performed to cluster cells into clusters. First, we used the *FindVariableFeatures* function to identify the top 2000 highly variable genes, with the selection method set as "vst". Second, all cells were scaled using the *ScaleData* function with the identified highly variable genes. Third, we used the *RunPCA* function to perform a linear dimension reduction. Fourth, we constructed the shared nearest neighbor (SNN) graph using the *FindNeighbors* function, with the number of principal components used for the *FindNeighbors* determined by the *ElbowPlot* function. Finally, we clustered cells using the *FindClusters* function (with data-specific resolution settings). This clustering method is based on optimizing SNN modularity. The data-specific resolution was 0.3 for pediatric glioblastoma in GSE131928, 0.8 for GSE122871, 0.8 for human medulloblastoma in GSE119926, 0.25 for GSE150752, 0.3 for GSE70630, and 0.4 for all other datasets. Cell clusters were visualized using t-distributed stochastic neighbor embedding (t-SNE) (using the *RunTSNE* function).

2.3 Identifying normal cell clusters

For each dataset, after cell clustering, the following steps were performed to identify normal cell clusters. First, we used *FindMarkers* function in *Seurat* to identify differentially expressed genes (DEGs) in each cluster compared to all cells in all other clusters (with "only.pos" = TRUE, "min.pct" = 0, "logfc.threshold" = 0.5, "test.use" = "wilcox", and other parameters took the default values). Genes with logFC > 0.5 and adjusted *p*-value < 0.01 were defined as DEGs. Second, we applied gene set enrichment analysis to DEGs in each identified cell cluster to detect if any marker genes (i.e., DEGs) of normal cell types were enriched. Marker genes of normal cell types were collected from the *CellMarker* database (Zhang et al., 2019), and the gene set enrichment analysis was performed using *ClusterProfiler* (v3.14.3) (with default parameters) (Yu et al., 2012). If the DEGs of a cell cluster were significantly enriched in the marker genes of a cell type in the CellMarker database, this cell cluster was defined as this cell type (BH-adjusted *p*-value < 0.05).

2.4 Scoring cells using binary markers of normal cell clusters

After identifying normal cell clusters upon the enrichment of DEGs in annotated cell markers, we extracted the properties of each identified normal cell cluster to refine the identification of normal

cells. First, we followed the approach developed by Hodge et al. (2019) to detect "binary markers" for each normal cell cluster. This method computed a binary score ranging from 0 to 1 to assess the exclusivity of a DEG between clusters. A higher binary score indicates greater exclusivity of gene expression in a particular cell cluster, and a DEG in a normal cell cluster with a binary score above 0.7 was defined as a binary marker of that normal cell cluster. Then, upon each set of binary markers, all cells in a dataset were scored using the *AddModuleScore* function. If a cell was scored above a dataset-specific cutoff, the cell was classified into the corresponding normal cell cluster.

2.5 Copy number analysis of cells

For each dataset generated by Smart-seq/Smart-seq2 and Drop-seq, we used a reported copy number analysis (CNA) method to evaluate whether a cell underwent genome duplication (Tirosh et al., 2016). First, we first sorted genes by chromosome and gene start position. Second, we used cells defined as normal cells upon the score of *AddModuleScore* as the reference. Third, we computed the aggregate expression of each gene using the equation $E_a(i) = log_2(average(\sum_{j=1}^{N} TPM_{ij}) + 1)$, where "i", "j" and "N" refer to gene, cell, and the total number of cells, respectively. Only genes with $E_a > 4$ (which indicates very high expression levels) were selected for further analysis. Fourth, for each selected gene, its relative expression in cell j was calculated by $Er_{ij} = E_{ij} - average(\sum_{j=1}^{N} E_{ij})$. Fifth, the initial CNA of each gene was averaged over the Er_{ij} values of this gene, its 50 genes upstream, and its 50 genes downstream, and the CNA was denoted as $CNA_{ij} = \sum_{i-50}^{i+50} Er_{ij}/101$. Sixth, the baseline CNA of each gene was defined as the average of CNA_{ij} of the reference cells and denoted as $Baseline\_CNA_i$. Seventh, the relative CNA of each gene was calculated as $CNA_{ij} - Baseline\_CNA_i$. Eighth, the CNA signal of each cell was defined as the averaged square of the relative CNA of all genes, that is, $CNA\_Signal_j = average(\sum_{i=1}^{M}(CNA_{ij} - Baseline\_CNA_i)^2)$, where "M" specifies the total number of genes. Finally, the CNA correlation of each cell was defined as the Pearson correlation between the relative CNA and the average of the relative CNA of non-reference cells from the same sample. The above CNA analysis was not applied to GSE119926 because the analysis needs the normal cells as the reference but no normal cell clusters were identified in GSE119926.

2.6 Identifying malignant cells

For datasets generated by Smart-seq/Smart-seq2 and Drop-seq, malignant cells were identified jointly upon three criteria: (a) A cell does not belong to normal cell clusters according to the cell cluster-based classification (see Method 2.3), (b) a cell is not defined as a normal cell when scored using the binary markers of normal cell clusters (see Method 2.4), (c) a cell has CNA signal > 0.02 and CNA correlation > 0.4 based on CNA analysis (see Method 2.5). For datasets generated by 10X genomics, malignant cells were identified based on the criteria (a) and (b) mentioned above. A cell was classified as malignant only when all the corresponding criteria gave concordant assignment. Since normal cell clusters were not identified in GSE119926, all cells in this dataset were defined as malignant cells.

2.7 Detection of NGS-negative and NGS-positive cells

We collected NGS genes from related studies (Osswald et al., 2015; Venkataramani et al., 2019; Venkatesh et al., 2019). Using the NGS genes as a signature, we first used *addmoduleScore* (with "nbin" = 30) in *Seurat* to score each cell to measure its potential to form NGS with neurons. Malignant cells with a score > 0 were identified as NGS-positive cells, which indicates they probably form NGS with neurons. Malignant cells with a score < 0 were identified as NGS-negative cells, which indicates they probably do not form NGS with neurons. By doing so, we divided malignant cells further into NGS-positive (NGS+) cells and NGS-negative (NGS−) cells.

2.8 Identifying malignant cells related to the worse survival time of glioma patients

We used the *Scissor* (v2.0) package to identify malignant cells associated with survival time (Sun et al., 2021). We conducted Scissor analysis for GSE131928, GSE70630, and GSE89567 by integrating the scRNA-seq datasets with the TCGA datasets of GBM, oligodendroglioma, and astrocytoma as follows (other scRNA-seq datasets lack the corresponding TCGA datasets). **First**, expression files and sample information of TCGA GBM and LGG were downloaded through *TCGAbiolinks* (v2.25.3). The oligodendroglioma and astrocytoma TCGA datasets are referred to as LGG because they belong to grade II low-grade gliomas. **Second**, TCGA samples without survival time and status were excluded. **Third**, we matched TCGA samples with scRNA-seq samples by age based on the following two considerations. The first one is that age is an important prognostic factor in glioma (Corell et al., 2018; Kim et al., 2021; Jia et al., 2022; Ostrom et al., 2022). For example, the age-adjusted incidence rate of GBM increases from 1.25 per 100,000 people among adults 35–34 years to 8.05, 12.99, and 15.13 among adults 55–65, 65–74, and 75–84 years, respectively (Ostrom et al., 2022); a recent finding revealed that the risk of mortality increases with age in a J-shaped pattern (Jia et al., 2022). The second consideration is that due to the limited samples in the three glioma scRNA-seq datasets, the age distribution in these datasets exhibits considerable variation and differs significantly from that of the TCGA samples. Specifically, the 20 GBM samples were between 52 and 78 years old; the 7 anaplastic astrocytoma samples were between 24 and 47 years old; the 3 oligodendroglioma samples were aged 31, 35, and 67 years old, respectively (Supplementary Table S1). In order to minimize the confounding effects of the age and tumor subtype, only TCGA samples matching corresponding scRNA-seq samples were included in the following analysis. Thus, only TCGA GBM samples aged between 52 and 78 were included for analyzing GSE131928; only TCGA oligodendroglioma samples with 1p/19q codeletion and aged between 30 and 40 or between 60 and 70 were included for analyzing GSE70630; only TCGA anaplastic astrocytoma samples with IDH1 mutation and aged between 24 and 47 were included for analyzing GSE89567. **Fourth**, for each of the three glioma types, the *Scissor* program was used to identify Scissor+ cells and Scissor− cells, with the parameter settings "cutoff" = 0.2, "family" = "cox", and "alpha" = 0.001/0.002/0.08 for astrocytoma/GBM/oligodendroglioma. These parameters ensured that the ratio of selected cells did not exceed 50% of the total analyzed cells. Scissor+ cells are associated with worse survival, while Scissor− cells are associated with good survival (Sun et al., 2021). **Finally**, we used the one-sided Fisher's exact test to investigate whether Scissor+ or Scissor− cells were enriched with NGS+ cells, and p-value < 0.01 was defined as the significant level.

2.9 Detection of differentially expressed genes in NGS+ cells

We used the *FindMarkers* function in *Seurat* to identify differentially expressed genes (DEGs) in NGS+ cells against NGS− cells with the parameter settings "min.pct" = 0, "logfc.thresholds" = 0.5, "test.use" = "wilcox", and other parameters adopting the default values). Genes with logFC > 0.5 and adjusted p-value < 0.01 were defined as DEGs. To reveal what the *Biological Process* gene ontology (GO) terms DEGs are related to, gene set enrichment analysis was performed using the *ClusterProfiler* (v3.14.3) R package (with default parameters, BH-adjusted p-value < 0.05) (Yu et al., 2012).

2.10 Prediction of transcriptional regulation of NGS genes by lncRNAs

To analyze NGS gene regulation by lncRNAs, we jointly used two methods on the transcriptome and genome levels, respectively. The first was the computation of Bayesian correlation between each pair of lncRNAs and NGS genes using the *Baco. R* function (with default parameters) (https://github.com/dsancheztaltavull/Bayesian-Correlations/blob/master/BaCo.R). Bayesian correlation is a robust measure for correlated genes in scRNA-seq data (Sanchez-Taltavull et al., 2020). The second was the prediction of lncRNA/DNA binding using the *LongTarget* program (with default parameters, all DBSs > 50 bp) (He et al., 2015; Lin et al., 2019). Many lncRNAs can bind to specific DNA sequences and recruit DNA and histone modification enzymes to the DNA binding sites. *LongTarget* predicts DNA binding domains (DBDs) in lncRNAs and DNA binding sites (DBSs) in DNA sequences simultaneously. We predicted the DBSs of lncRNAs in the promoter regions (+3,500 bp ~ −1,500 bp upstream and downstream transcription start site) of NGS genes. If a lncRNA and an NGS gene have a Bayesian correlation > 0.4 (adjusted p-value < 0.05) and the lncRNA has a DBS with binding affinity > 100 (about > 150 bp) in the promoter region of the NGA gene, the lncRNA was assumed to be able to regulate the NGS gene. The networked regulations of multiple NGS genes by multiple lncRNAs were visualized using the *Cytoscape* program (3.9.1).

2.11 Validation of the analysis methods and results

We validated the lncRNA/DNA binding prediction using two methods. First, we identified human-specific lncRNAs from the 13562 GENCODE-annotated human lncRNAs, predicted their DBSs genome-wide using the *LongTarget* program (He et al., 2015; Lin et al., 2019), used the CRISPR/Cas9 technique to knock out the predicted DBDs in three human-specific lncRNA genes in the HeLa cell line, RKO cell line, and SK-MES-1 cell line, and performed RNA-seq before and

after the DBD knockout, then analyzed DEGs in each cell line. Second, the genome-wide DNA binding sites of NEAT1, MALAT1, and MEG3 were experimentally identified using ChIRP-seq (West et al., 2014; Mondal et al., 2015), we predicted the genome-wide DBSs of the three lncRNAs (NEAT1, MALAT1, and MEG3, compared predicted DBSs with the experimentally identified ones and examined whether the two groups overlap significantly (Wen et al., 2022).

We validated the contribution of NGS+ cells to the poor survival of glioma patients by performing two survival analyses. First, expression files and sample information of TCGA GBM and LGG were downloaded through *TCGAbiolinks* (v2.25.3). TCGA samples without survival information were discarded. To remove the confounding effect of sex and age, we set these factors as fixed effects in the *coxph* function. Using the *Scissor* program (Sun et al., 2021), we then identified malignant cells associated with patient survival time. We used Fisher's exact test to examine whether the Scissor + cell group was predominantly composed of NGS + cells. Second, to verify that lncRNAs involved in regulating NGS genes influence survival time, we used the Cox proportional hazards model (in the R package *survival* v3.2.7) to examine whether lncRNAs regulating NGS genes correlate with survival time.

XIST is an important lncRNA and regulates many genes genome-wide (Statello et al., 2021). We observed that XIST has DBSs in 12 out of 15 NGS genes in the oligodendroglioma dataset. To examine whether these DBSs could be determined by chance, we conducted a permutation test. First, we predicted XIST's DBSs genome-wide. Of all the 45,515 annotated genes examined, 13,504 genes have DBSs. Second, we randomly sampled 15 genes from the whole genome and calculated the ratio of genes having DBS of XIST, repeated the process 1,000,000 times, and obtained an empirical distribution of this ratio. Finally, we compared 12 out of the 15 NGS genes with this distribution and obtained the *p*-value.

3 Results

3.1 NGS gene expression characterizes glioma types, cell clusters, and patients

We downloaded nine scRNA-seq datasets of five gliomas from the GEO website, including glioblastoma (GBM), diffuse H3K27M glioma (H3K27M glioma), medulloblastoma (MB), 1p/19q codeleted oligodendroglioma (oligodendrocyte or OD), and IDH1-mutant anaplastic astrocytoma (astrocytoma or IDH1-mutant AA) (Supplementary Table S1). We integrated NGS genes in these datasets (Supplementary Table S2). Upon gene annotations, these NGS genes are involved in microtube assembly, synaptogenesis, neurotransmitter receptors, and neuronal paracrine based on the experimental studies (Venkataramani et al., 2019; Venkatesh et al., 2019; Zeng et al., 2019).

For each dataset, we first identified normal cells using two methods. The first method includes clustering cells into clusters, identifying DEGs in each cluster, and checking whether DEGs in a cluster are enriched in annotated marker genes of cell types collected in the *CellMarker* database (Zhang et al., 2019). If DEGs in a cell cluster were significantly enriched in the annotated marker genes of a cell type, this cell cluster was defined as the cell type and identified

as a normal cell type (all cell types in the *CellMarker* database are normal cells) (Figures 1A, B). In the medulloblastoma and medulloblastoma PDX mouse models, no cell cluster was identified as normal cells. The second method includes extracting the properties of each identified normal cell cluster to refine normal cell identification. We used a method to detect "binary markers" for each normal cell cluster (Hodge et al., 2019). In this method, a binary score ranging from 0 to 1 was computed to assess the exclusivity of a DEG between clusters. A DEG in a cell cluster with a binary score > 0.7 was defined as a binary marker of that cell cluster. Upon each set of binary markers, all cells in a dataset (but not in any identified cluster) were scored. If a cell was scored above a dataset-specific cutoff, the cell was classified into the corresponding normal cell cluster.

Then, we determined whether a cell was a malignant cell jointly upon three criteria: a) It did not belong to any normal cell cluster, b) it could not be defined as a normal cell when scored by binary markers of normal cell clusters, and c) copy number analysis (CNA) revealed that it underwent abnormal genome duplication. Since no normal cells were identified in the medulloblastoma and medulloblastoma PDX mouse models, all cell clusters were malignant cells. In these gliomas, malignant cells can be divided into multiple clusters (Figures 1A, B), and some malignant cell clusters are present only in very few patients (Supplementary Table S3). For example, malignant cells in GBM were divided into 14 clusters, with clusters 7, 8, 9, and 12 comprising cells from a single patient. Thus, gliomas have high inter- and intra-tumor heterogeneity, as previously reported (Patel et al., 2014; Neftel et al., 2019).

Since gliomas are highly heterogenous, we next examined whether NGS gene expression is highly heterogeneous. We found that, generally, NGS genes were expressed tumor-specifically, with distinct expression in different tumors and even in patients with the same tumor. Nevertheless, the NGS genes *PTPRS*, *MAP2*, and *CADM1* were expressed in most malignant cells across these gliomas (Figures 2A, B), consistent with the observed neurological associations of these genes (Zhou et al., 2015; Hendriks et al., 2018; Pan et al., 2020). To examine whether NGS genes characterize malignant cells, we scored malignant cells upon NGS gene expression and divided malignant cells into NGS-negative (NGS−) and NGS-positive (NGS+) ones (Figure 1C). In each glioma, the NGS score in malignant cells is significantly higher than the NGS score in non-malignant cells (*p*-value = 3.664e-06), suggesting reasonable identification of malignant and normal cells. The ratio of NGS+ malignant cells varies not only among different types of gliomas (Figure 2C) but also among different cell clusters (Figure 2D). For example, the ratio of NGS+ cells in GBM in clusters 4, 9, and 12 is much lower than the ratio in other clusters (Figure 2D). Thus, different subsets of NGS genes characterize glioma types, cell clusters, and aggression ability.

3.2 Dysregulated genes in malignant NGS+ cells are enriched in neurodevelopment- and synapse-related GO terms

Based on our definition, the critical difference between NGS+ malignant and NGS− malignant cells is whether a malignant cell can
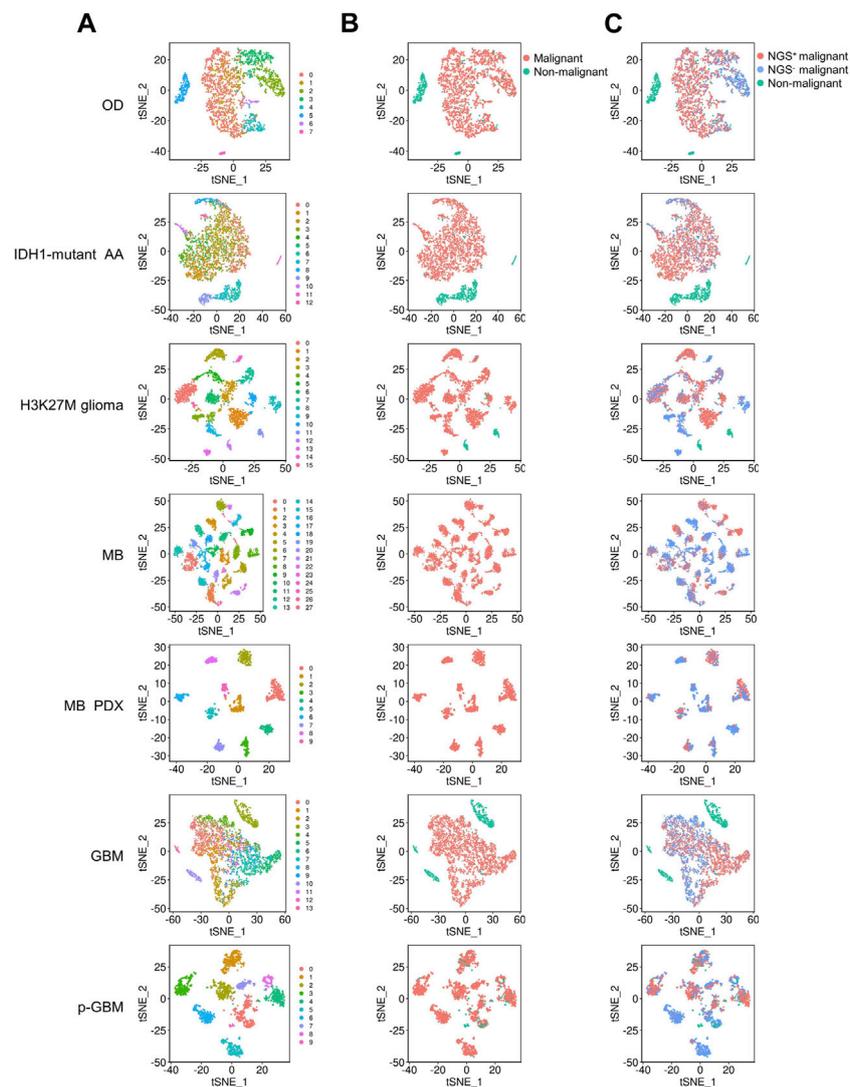
FIGURE 1
tSNE plots show the clustering of cells in the 9 glioma datasets. OD, IDH1-mutant AA, H3K27M glioma, MB, MB PDX, GBM, and p-GBM indicates 1p/19q codeleted oligodendroglioma, IDH1-mutant anaplastic astrocytoma, diffuse H3K27M glioma, medulloblastoma, patient-derived xenografts model of medulloblastoma, glioblastoma, and pediatric GBM, respectively. Each glioma consists of multiple cell clusters. **(A)** The distribution of cell clusters (numbers indicate different clusters). **(B)** The distribution of malignant and non-malignant cells. **(C)** The distribution of NGS⁺ malignant cells, NGS⁻ malignant cells, and non-malignant cells.

form synapses with neurons. Thus, if our method of classifying malignant cells into NGS⁺ and NGS⁻ groups is valid, DEGs between NGS⁺ and NGS⁻ malignant cells should have an association with synapse formation. To verify this, we first identified DEGs between NGS⁺ and NGS⁻ malignant cells in each glioma (Supplementary Table S4), then applied gene set enrichment analysis to DEGs using the *ClusterProfiler* program and the Gene Ontology (GO) database. DEGs in all gliomas are significantly enriched in neurodevelopmental-related GO terms, and DEGs in many gliomas are enriched in synapse-related GO terms (Figure 3), supporting a strong association between neurons and NGS⁺ malignant cells. In addition, different gliomas show somewhat different GO terms and levels of enrichment, suggesting differences in NGS formation in different gliomas. These results also support the validity of classifying malignant cells into NGS⁺ and NGS⁻ cells.

3.3 LncRNAs transcriptionally regulate NGS gene expression

As abundant lncRNA genes are expressed in the human brain, it is sensible to conjecture that they also contribute to tumorigenesis in the brain. We identified differentially expressed lncRNA genes between NGS+ and NGS- malignant cells to reveal the contribution. Multiple top differentially expressed lncRNA genes (Supplementary Table S4), including LINC00599, MEG3, and TTTY15, were identified in NGS⁺ malignant cells. We next examined the tumor-specificity of lncRNA expression because lncRNA gene expression is highly tissue-specific. Distinctively dysregulated lncRNAs in gliomas include LOC150622 and LINC00645 in glioblastoma, LOC100216479 in diffuse H3K27M glioma, FLJ41350 and FTX in medulloblastoma, TPTEP1 and
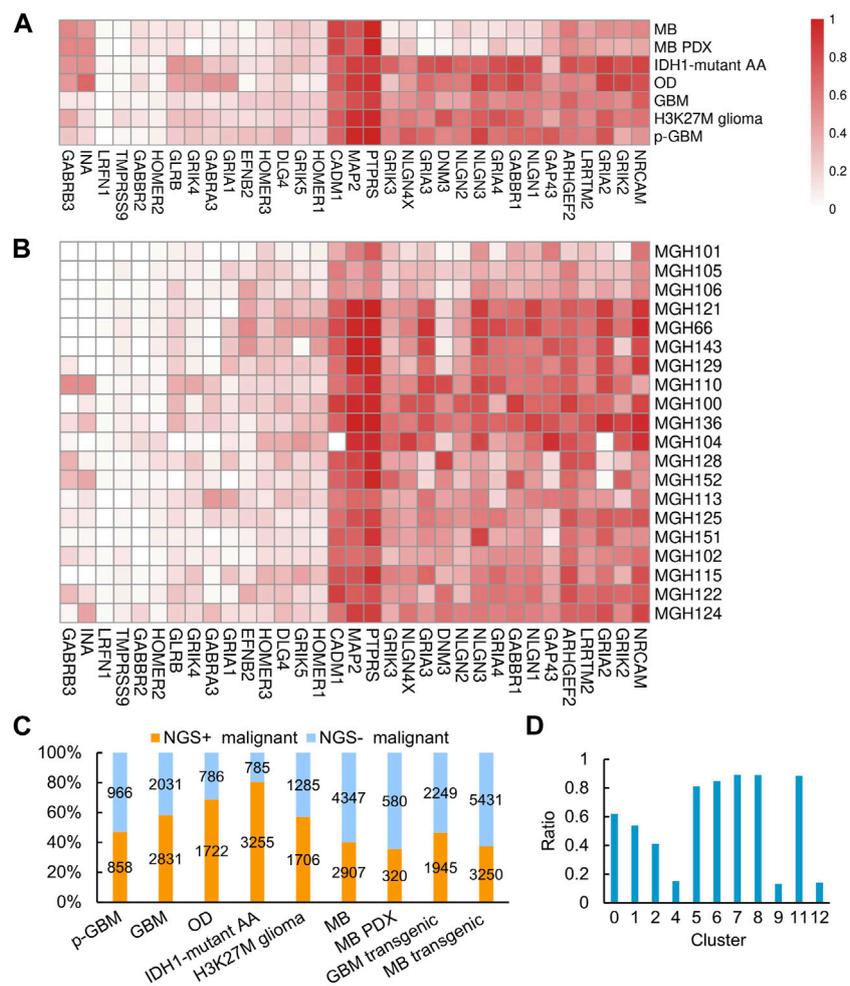
FIGURE 2
NGS gene expression in different gliomas, GBM samples, and cell clusters. **(A)** NGS gene expression in different gliomas. **(B)** NGS gene expression in different GBM samples. In **(A, B)**, the color bar indicates the percentage of expressed cells. **(C)** The percentages and numbers of NGS+ and NGS− malignant cells in different gliomas. **(D)** The percentages of NGS+ cells in the 13 clusters of malignant GBM cells.

LOC100144595 in astrocytoma, and LOC100216545, FAM66B, and LOC100499405 in oligodendroglioma.

To obtain evidence supporting that dysregulated lncRNAs regulate NGS gene expression, we used two methods to examine the relationships between lncRNAs and NGS genes. On the transcriptome level, we calculated the pairwise Bayes correlation between all differentially expressed NGS genes and all differentially expressed lncRNAs. On the genome level, we used the *LongTarget* program to predict differentially expressed lncRNAs' DNA binding sites (DBS) in the promoter regions of differentially expressed NGS genes (He et al., 2015; Lin et al., 2019). Jointly upon the two methods, 28 lncRNAs were identified to regulate NGS genes in gliomas (Supplementary Table S5). The most notable lncRNA is LINC00599, which has DBSs in many NGS genes in most gliomas, including *GABBR1, NRCAM, PTPRS, GLRB, GRIA4, GRIK3*, and *MAP2* (Figures 4A–G), and the most notable NGS gene is *MAP2*, which is associated with microtube assembly and regulated by multiple lncRNAs through the same DBS (Figure 4H) (Osswald et al., 2015).

3.4 NGS gene regulation by lncRNAs is species- and tumor-specific

Many studies reported tissue-specific expression of lncRNA genes in tumors (Mattioli et al., 2019; Liu et al., 2021)ꞏ suggesting their tumor-specific regulatory functions. To examine whether tumor-specific lncRNA gene expression causes tumor-specific NGS gene expression, we generated the regulatory networks of lncRNAs and NGS genes in different gliomas based jointly on the Bayes correlation and predicted lncRNA-target NGS gene relationship (Figure 5). **First**, 61% of lncRNAs regulate NGS genes in one specific glioma (Supplementary Table S5), suggesting that most lncRNAs regulate NGS genes tumor-specifically. **Second**, some essential NGS genes such as *MAP2* are regulated by different lncRNAs in different gliomas; meanwhile, important lncRNAs such as LINC00599 regulate different NGS genes in different gliomas (Figure 5). **Third**, an NGS gene can be regulated by different lncRNAs in humans and mice. Notable examples are *MAP2*, *PTPRS*, and *INA* which are expressed in MB and mouse MB PDX; these genes are regulated

FIGURE 3
The enriched gene sets of DEGs in different gliomas **(A−F)**. DEGs are enriched in neurodevelopmental-related GO terms in all gliomas **(A−F)** and synapse-related GO terms in four gliomas **(B,C,D,F)**. These results support the importance of dysregulated NGS genes for gliomas.
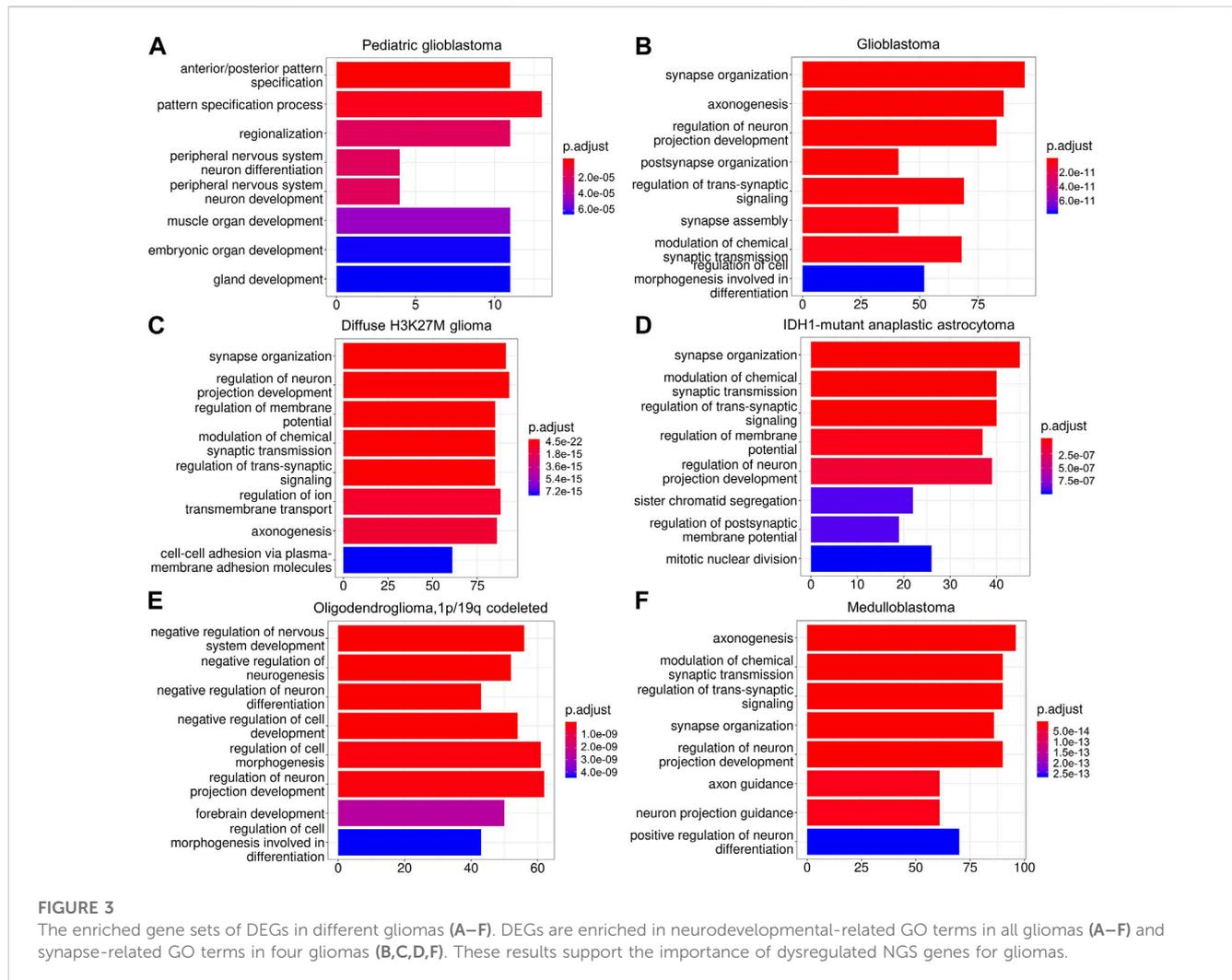
by multiple lncRNAs in human MB but only by LINC00461 in mouse MB PDX. Similarly, a lncRNA can target different NGS genes in humans and mice. We found that LINC00461 regulates NGS genes in mouse MB PDX but does not regulate any NGS gene in human gliomas. These suggest that NGS gene regulation by lncRNAs relies on specific transcriptional landscapes, and such landscapes may critically characterize patterns of NGS gene regulation. In support of this interpretation, Miat was found to regulate the NGS gene *Gria2* in transgenic mouse MB, but MIAT (the human orthologue of Miat) was found not to regulate any NGS genes in these glioma datasets (Supplementary Table S5). **Fourth**, NGS genes in mouse MB PDX showed similar expression levels to NGS genes in human MB; however, NGS genes in mouse transgenic MB showed different expression levels from NGS genes in human GBM (Supplementary Table S6). **Fifth**, the identified networks indicate that LINC00599 is a critical lncRNA regulating NGS genes in multiple gliomas, with LINC00599-MAP2 and LINC00599-NRCAM being the core lncRNA-NGS gene pairs. These results suggest that NGS gene regulation by lncRNAs is highly species- and tumor-specific, and these lncRNAs provide a new class of targets for glioma treatment.

3.5 Simian-specific lncRNAs are more involved in NGS gene regulation

Clade-specific lncRNAs, such as simian- and rodent-specific ones, are evolutionarily younger than lnRNAs conserved in mammals. Since simian-specific lncRNAs in the human genome have contributed specifically to the evolution of the human brain (Derrien et al., 2012; Briggs et al., 2015; Rashighi and Harris, 2017; Zimmer-Bensch, 2019), it is sensible to conjecture that these lncRNAs may also greatly affect brain tumorigenesis compared with lncRNAs conserved in mammals. To test this conjecture, we examined the conservation of the 28 lncRNAs that regulate NGS genes in these gliomas (Supplementary Table S5). Upon the whole-genome alignments in the UCSC Genome Browser (https://genome.ucsc.edu) and the simian-specific lncRNAs in the lncRNA database *LongMan* (http://www.gaemons.net/LongMan) (Lin et al., 2019), we found that 10 (35.7%) of the 28 lncRNAs are conserved in mammals (i.e., having orthologues in non-primate mammals) but 18 (64.3%) of the 28 lncRNAs are simian-specific (i.e., having orthologues only in marmoset, macaque, and chimpanzee) (Figure 6A). Additionally, compared with the
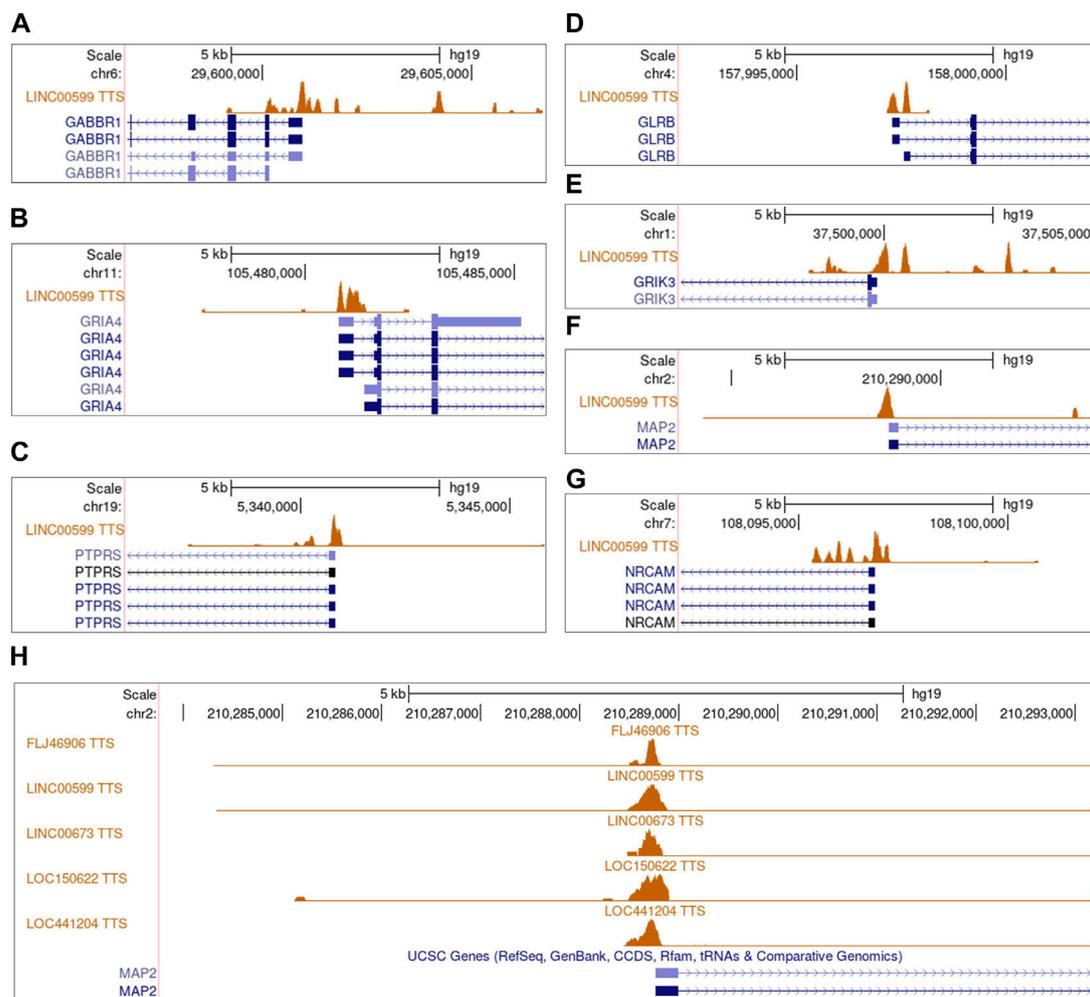
FIGURE 4
The predicted DNA binding sites (DBSs) of lncRNAs in the promoter regions of differentially expressed NGS genes. Panels are snapshots of the UCSC Genome Browser wherein predicted DBSs were uploaded as a custom track. "TTS" means triplex-targeting sites, and a set of overlapping TTSs forming a peak was a DBS. **(A–G)** The DBSs of LINC00599 in the promoter region of multiple differentially expressed NGS genes. The peaks in each panel's "LINC00599 TTS" track indicate LINC00599s DBSs. **(H)** The DBSs of FLJ46906, LINC00599, LIC00673, LOC150622, and LOC441204 in the promoter region of *MAP2*.

ratio of all simian-specific lncRNAs to all ENCODE-annotated human lncRNAs, the ratio of NGS gene regulation-related simian-specific lncRNAs to all NGS gene regulation-related lncRNAs is significantly higher (*p*-value = 0.047, Fisher's exact test). This suggests that simian-specific lncRNAs not only affect tumorigenesis in the brain but also determine the differences of gliomas in humans and mice. This lends an explanation for the finding that NGS gene regulation by lncRNAs is highly species-specific. In accordance, in the 17 (61%) lncRNAs that regulate NGS genes in one glioma (Figure 6B), 11 are simian-specific.

Many human diseases have distinct evolutionary roots in the genome (Maxwell et al., 2014; Benton et al., 2021), and the same may be true for brain tumors. By checking lncRNAs regulating NGS genes in different glioma, we found that most lncRNAs in diffuse H3K27M glioma are simian-specific and most lncRNAs in OD are conserved in mammals (Figure 6C). This probably highlights different evolutionary origins of epigenetic regulation of genes in different types of gliomas.

3.6 Validation of analysis methods and results

A key component of the above human/mouse comparative pan-gliomas analyses is analyzing the regulation of NGS genes by lncRNAs through predicting DBDs in lncRNAs and their DBSs in NGSs genes. Thus, we validated DBD/DBS prediction in three ways. **First**, in a related study, we identified human-specific lncRNAs from the 13562 GENCODE-annotated human lncRNAs (http://www.gaemons.net/LongMan) and predicted the genome-wide DBSs of these lncRNAs (He et al., 2015; Lin et al., 2019). We used the CRISPR/Cas9 technique to knock out the predicted DBDs in three human-specific lncRNA genes in the HeLa, RKO, and SK-MES-1 cell lines and performed RNA-seq before and after the DBD knockout. Differential gene expression analysis revealed that the |fold change| of target genes (DEGs whose promoter regions contain DBSs of these lncRNAs) was significantly larger than the |fold change| of non-target genes
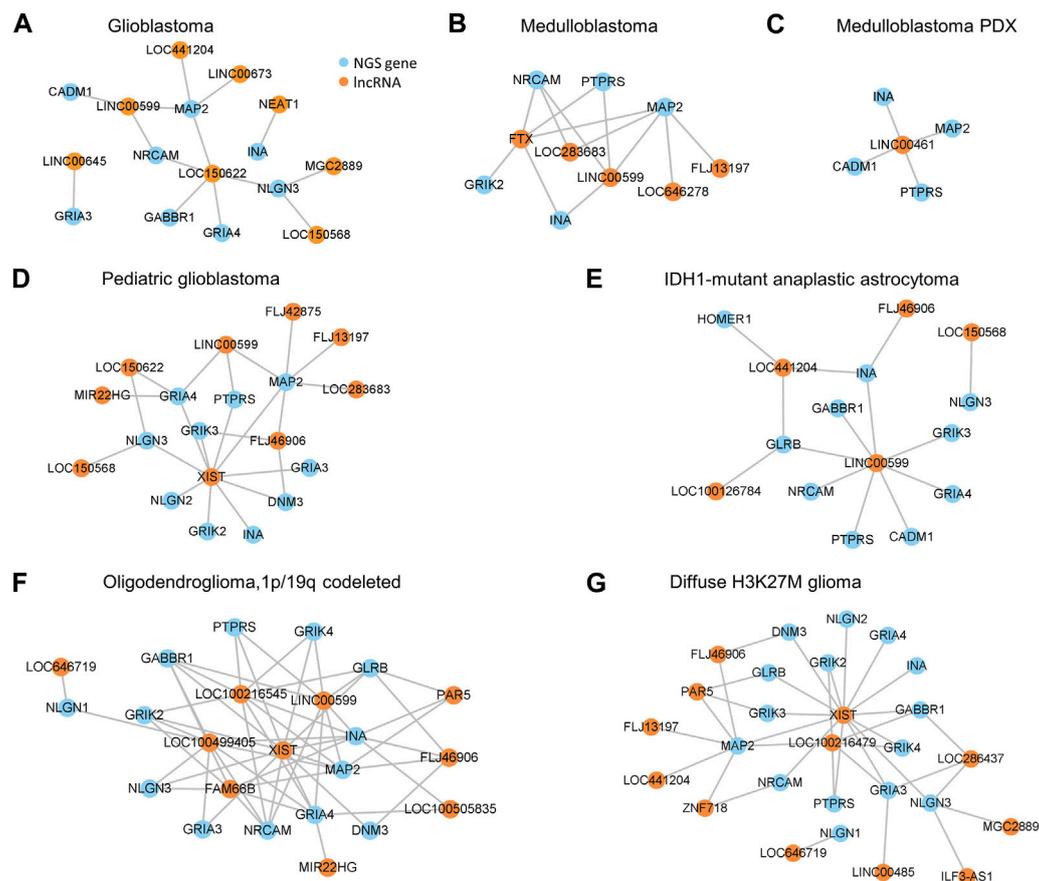
FIGURE 5
Networks of NGS gene regulation by lncRNAs based jointly on Bayes correlation and lncRNA-target gene relationship. **(A–G)** The networks in GBM **(A)**, MB **(B)**, MB PDX **(C)**, p-GBM **(D)**, IDH1-mutant AA **(E)**, OD **(F)**, and H3K27M glioma **(G)**.

(DEGs whose promoter regions do not contain DBSs of these lncRNAs) (one-sided Mann-Whitney test, *p*-value = 3.1e-72, 1.49e-114, and 1.12e-206 for RP13-516M14.1, RP11-426L16.8, and SNORA59B) (see *Data availability statement*). These results suggest that the knockout of DBDs caused the changed expression of target genes. **Second**, since the genome-wide DNA binding sites of NEAT1, MALAT1, and MEG3 were experimentally identified using ChIRP-seq (West et al., 2014; Mondal et al., 2015), we predicted genome-wide DBSs of NEAT1, MALAT1, and MEG3, compared predicted DBSs with the experimentally identified one, and found that the two groups of DBSs overlap well (Wen et al., 2022). **Third**, the *LongTarget* program uses a variant of the Smith-Waterman algorithm to identify all local alignments between a pair of RNA and DNA sequences, thus identifying DBDs and DBSs simultaneously. We computed the likelihood that an alignment could be generated by chance. According to the alignment algorithm, a DBS of >140 bp is extremely unlikely to be generated by chance (*p* < 8.2e-19 to 1.5e-48). These results support the analysis of NGS gene regulation by lncRNAs.

Malignant cells in gliomas hijack potassium currents to promote their aggression and proliferation, suggesting that glioma patients with substantial NGS formation should have a worse prognosis. Therefore, we performed two survival analyses

of glioma patients. **First**, we downloaded the TCGA data of GBM, oligodendroglioma, and astrocytoma (the three gliomas have TCGA datasets) and identified malignant cells associated with patient survival time using the *Scissor* program (Sun et al., 2021). *Scissor* was developed to associate cells in scRNA-seq data with some trait (here, the survival time) in the corresponding TCGA data. According to the usage of *Scissor*, Scissor+ cells were defined as cells associated with worse survival, while Scissor− cells were defined as cells associated with good survival. We used Fisher's exact test to examine whether the Scissor+ cell group was predominantly composed of NGS+ cells. Compared with the malignant cell group as the background, the Scissor+ cell group is significantly enriched in NGS+ cells, while the Scissor− cell group is significantly lacking in NGS+ cells in GBM, oligodendroglioma, and astrocytoma (Figures 6D–F; *p*-value = 1.6e-4, 4.4e-19, and 9.9e-4, respectively). Specifically, the *Scissor* program identified 189 Scissor+ cells (including 55 NGS− cells and 134 NGS+ cells) and 2064 Scissor− cells (including 1102 NGS− cells and 962 NGS+ cells) in GBM, 713 Scissor+ cells and 5 Scissor− cells in oligodendroglioma, and 730 Scissor+ cells and 189 Scissor− cells in astrocytoma (Figures 6D–F). In line with the relationship between NGS formation and survival time, these results support that NGS+ cells contribute to the worse survival of
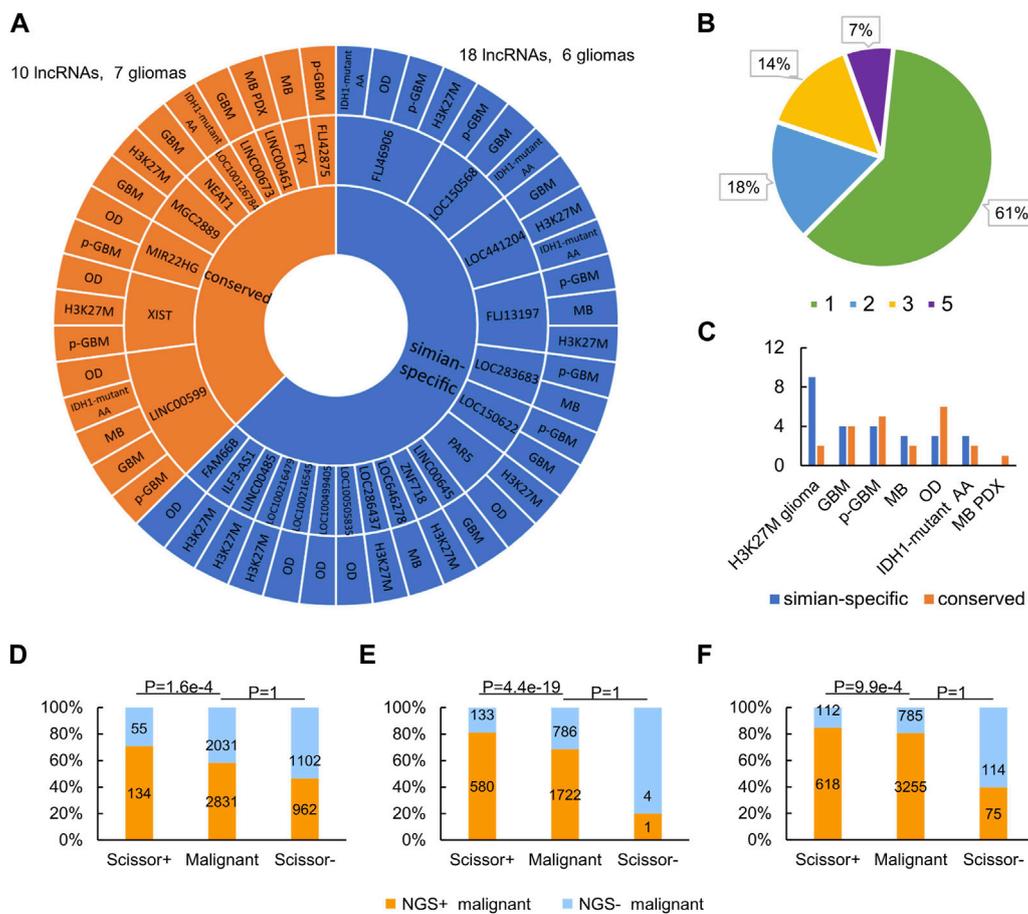
FIGURE 6
NGS regulation by lncRNAs has significant cross-species differences. **(A)** Conserved and simian-specific lncRNAs that regulate NGS genes in gliomas. **(B)** The percentages of lncRNAs that regulate NGS genes in 1, 2, 3, and 5 gliomas. **(C)** Numbers of conserved and simian-specific lncRNAs in each glioma. **(D–F)** Compared with the malignant cell group (as the background), the Scissor+ cell group is significantly enriched in NGS+ cells, while the Scissor− cell group is significantly lacking in NGS+ cells, in GBM **(D)**, oligodendroglioma **(E)**, and IDH1-mutant anaplastic astrocytoma **(F)** (p-value = 1.6e-4, 4.4e-19, and 9.9e-4, respectively, Fisher's exact test).

glioma patients. **Second**, to verify that lncRNAs involved in regulating NGS genes influence survival time, we used the Cox proportional hazards model (in the R package *survival* v3.2.7) and lncRNAs to perform survival analysis. The analysis unveiled that 5 of 28 lncRNA potentially regulating NGS genes were correlated with survival. Among them, the lower LINC00599 expression and higher XIST expression were significantly correlated with worse survival of low-grade glioma patients (Supplementary Table S7). This result is supported by a related study (Fu et al., 2019).

Finally, we conducted a permutation test using the lncRNA XIST and its target NGS genes in the oligodendroglioma dataset to check whether NGS gene regulation by lncRNAs could occur by chance. XIST has DBSs in 12 out of 15 NGS genes in the oligodendroglioma dataset and has DBSs in 13,504 genes out of the 45,515 Ensembl-annotated genes. We randomly sampled 15 genes from the whole genome, calculated the ratio of genes having DBS of XIST, repeated the process 1,000,000 times, obtained an empirical distribution of this ratio, and compared 12 out of the 15 NGS genes with the obtained distribution. The result suggests

that NGS gene regulation by XIST unlikely occurs by chance (p-value = 8e-5, permutation test).

4 Discussion

Gliomas are the most lethal tumors in humans. NGS in gliomas and brain-metastatic breast cancers were recently found to promote tumor progression (Venkataramani et al., 2019; Venkataramani et al., 2022b; Venkatesh et al., 2019; Zeng et al., 2019). This finding greatly improves the understanding of why gliomas are highly aggressive and highlights the potential of targeting NGS as a treatment for gliomas. While NGS genes have drawn increasing attention (Ji et al., 2020; Lin et al., 2021), some questions remain unclear, including whether NGS gene expression shows species- and tumor-specificity and whether lncRNAs regulate the specificity in a species-specific manner. Since targeting NGS genes (which have wide functions) may cause undesired side effects, and researchers use PDX and transgenic mouse models to study gliomas, answering the two questions and exploring targetable lncRNAs are important.

This study addresses these questions by checking NGS gene expression and regulation by lncRNAs in multiple gliomas in humans and mice. The main results are that lncRNAs regulate NGS genes highly species- and tumor-specifically and that simian-specific lncRNAs in humans are more involved in NGS gene regulation than conserved lncRNAs. These findings, with Supplementary Material, suggest that lncRNAs are potential tumor-specific targets for human gliomas. To verify that lncRNAs involved in regulating NGS genes influence the survival time of glioma patients, we combined scRNA-seq data and TCGA data to perform survival analyses using two methods. These analyses generated consistent results.

Two studies examined 17 and 231 synapse-related genes in human gliomas (Ji et al., 2020; Lin et al., 2021). Compared with them, our cross-species pan-gliomas analysis focused on 34 NGS genes collected from the literature. Despite using different genes, all of these studies demonstrate the prognostic values of NGS genes. Previous studies have also revealed that thousands of new lncRNAs have emerged during primate nervous system evolution and that nearly half of the identified lncRNAs are specifically expressed in the human brain (Derrien et al., 2012; Rashighi and Harris, 2017; Zimmer-Bensch, 2019). Of interest, we found that simian-specific lncRNAs are more involved in NGS gene regulation than lncRNAs conserved in mammals. Thus, simian-specific lncRNAs contribute not only to the evolution of the brain but also to the development of brain tumors. We also observed that different types of gliomas may have different evolutionary origins. For instance, most lncRNAs regulating NGS genes in diffuse H3K27M glioma are simian-specific, while most lncRNAs regulating NGS genes in OD are conserved in mammals. This finding supports the recent viewpoint that many human diseases have distinct evolutionary roots in the genome (Maxwell et al., 2014; Benton et al., 2021).

Some comparative cancer studies suggested conservation between human and mouse tumors (Aytes et al., 2014; Onken et al., 2014; Woo et al., 2021). However, many drugs tested successfully in mice failed in human clinical trials (Gould et al., 2015). We observed that, although NGS genes show a similar expression pattern in PDX mouse models, which is consistent with the finding that more than 80% of glioma cells in PDX mouse models had tumor microtubes (Drumm et al., 2020), the patterns of NGS gene expression in transgenic mouse glioma models and human gliomas are poorly comparable. When it comes to the regulation of NGS gene by lncRNAs, neither PDX mouse models nor transgenic mouse models share commonalities with human gliomas. This agrees with the finding that phenotypic differences between human and mouse orthologous genes are mainly caused by the evolutionary rewiring of regulatory networks (Ha et al., 2022). In brief, this study reveals that lncRNAs regulate NGS genes highly species- and tumor-specifically and that simian-specific lncRNAs are substantially involved in human gliomas. These findings help rational design of mouse models for glioma studies (Landgraf et al., 2018), and suggest that caution is needed when interpreting discoveries from glioma studies using PDX and transgenic mouse models.

5 Conclusion

NGS genes are regulated by lncRNAs and this regulation is highly species- and tumor-specific. NGS gene expression and regulation in human gliomas are quite different from those in PDX and transgenic mouse models. Simian-specific lncRNAs with tumor-specific expression and function provide a new class of potential targets for glioma treatment.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Human lncRNAs' orthologues in 16 mammals and the *LongTarget* program are available at the *LongTarget/LongMan* website (http://www.gaemons.net/). The RNA-seq data of the three HS lncRNAs before and after the CRISPR/Cas9 experiments have been deposited in the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo) under the accession number GSE213231. Further inquiries can be directed to the corresponding authors.

Author contributions

SH and HZ designed the study. WX, SH, BP, and XZ performed the data analysis. SH and WX drafted the manuscript. HZ and LH revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1218408/full#supplementary-material

References

Alinejad-Rokny, H., Heng, J. I. T., and Forrest, A. R. R. (2020). Brain-enriched coding and long non-coding RNA genes are overrepresented in recurrent neurodevelopmental disorder CNVs. *Cell Rep.* 33, 108307. doi:10.1016/j.celrep.2020.108307

Amarouch, A., and Mazeron, J. J. (2005). Radiotherapy plus concomitant and adjuvant Temozolomide for glioblastoma. *Cancer/Radiotherapie* 9, 196–197. doi:10.1016/j.canrad.2005.05.001

Aytes, A., Mitrofanova, A., Lefebvre, C., Alvarez, M. J., Castillo-Martin, M., Zheng, T., et al. (2014). Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* 25, 638–651. doi:10.1016/j.ccr.2014.03.017

Benton, M. L., Abraham, A., LaBella, A. L., Abbot, P., Rokas, A., and Capra, J. A. (2021). The influence of evolutionary history on human health and disease. *Nat. Rev. Genet.* 22, 269–283. doi:10.1038/s41576-020-00305-9

Briggs, J. A., Wolvetang, E. J., Mattick, J. S., Rinn, J. L., and Barry, G. (2015). Mechanisms of long non-coding RNAs in mammalian nervous system development, plasticity, disease, and evolution. *Neuron* 88, 861–877. doi:10.1016/j.neuron.2015.09.045

Campbell, P. J., Getz, G., Korbel, J. O., Stuart, J. M., Jennings, J. L., Stein, L. D., et al. (2020). Pan-cancer analysis of whole genomes. *Nature* 578, 82–93. doi:10.1038/s41586-020-1969-6

Cheng, Y., Liao, S., Xu, G., Hu, J., Guo, D., Du, F., et al. (2020). NeuroD1 dictates tumor cell differentiation in medulloblastoma. *Cell Rep.* 31, 107782. doi:10.1016/j.celrep.2020.107782

Connolly, N. P., Shetty, A. C., Stokum, J. A., Hoeschele, I., Siegel, M. B., Miller, C. R., et al. (2018). Cross-species transcriptional analysis reveals conserved and host-specific neoplastic processes in mammalian glioma. *Sci. Rep.* 8, 1180–1215. doi:10.1038/s41598-018-19451-6

Corell, A., Carstam, L., Smits, A., Henriksson, R., and Jakola, A. S. (2018). Age and surgical outcome of low-grade glioma in Sweden. *Acta Neurol. Scand.* 138, 359–368. doi:10.1111/ane.12973

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi:10.1101/gr.132159.111

Drumm, M. R., Dixit, K. S., Grimm, S., Kumthekar, P., Lukas, R. V., Raizer, J. J., et al. (2020). Extensive brainstem infiltration, not mass effect, is a common feature of end-stage cerebral glioblastomas. *Neuro. Oncol.* 22, 470–479. doi:10.1093/neuonc/noz216

Elyada, E., Bolisetty, M., Laise, P., Flynn, W. F., Courtois, E. T., Burkhart, R. A., et al. (2019). Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov.* 9, 1102–1123. doi:10.1158/2159-8290.CD-19-0094

Filbin, M. G., Tirosh, I., Hovestadt, V., Shaw, M. L., Escalante, L. E., Mathewson, N. D., et al. (2018). Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Sci. (80-.)* 360, 331–335. doi:10.1126/science.aao4750

Fu, Q., Li, S., Zhou, Q., Yalikun, K., Yisireyili, D., and Xia, M. (2019). Low LINC00599 expression is a poor prognostic factor in glioma. *Biosci. Rep.* 29, BSR20190232–9. doi:10.1042/BSR20190232

Gould, S. E., Junttila, M. R., and De Sauvage, F. J. (2015). Translational value of mouse models in oncology drug development. *Nat. Med.* 21, 431–439. doi:10.1038/nm.3853

Ha, D., Kim, D., Kim, I., Oh, Y., Kong, J. H., Han, S. K., et al. (2022). Evolutionary rewiring of regulatory networks contributes to phenotypic differences between human and mouse orthologous genes. *Nucleic Acids Res.* 50, 1849–1863. doi:10.1093/nar/gkac050

He, S., Zhang, H., Liu, H., and Zhu, H. (2015). LongTarget: a tool to predict lncRNA DNA-binding motifs and binding sites via hoogsteen base-pairing analysis. *Bioinformatics* 31, 178–186. doi:10.1093/bioinformatics/btu643

Hendriks, W., Bourgonje, A., Leenders, W., and Pulido, R. (2018). Proteinaceous regulators and inhibitors of protein tyrosine phosphatases. *Molecules* 23, 395–422. doi:10.3390/molecules23020395

Hodge, R. D., Bakken, T. E., Miller, J. A., Smith, K. A., Barkan, E. R., Graybuck, L. T., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. *Nature* 573, 61–68. doi:10.1038/s41586-019-1506-7

Hovestadt, V., Smith, K. S., Bihannic, L., Filbin, M. G., Shaw, M. K. L., Baumgartner, A., et al. (2019). Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* 572, 74–79. doi:10.1038/s41586-019-1434-6

Ji, X., Zhang, H., and Cui, Q. (2020). A panel of synapse-related genes as a biomarker for gliomas. *Front. Neurosci.* 14, 822–828. doi:10.3389/fnins.2020.00822

Jia, Z., Li, X., Yan, Y., Shen, X., Wang, J., Yang, H., et al. (2022). Exploring the relationship between age and prognosis in glioma: rethinking current age stratification. *BMC Neurol.* 22, 350. doi:10.1186/s12883-022-02879-9

Kim, M., Ladomersky, E., Mozny, A., Kocherginsky, M., O'shea, K., Reinstein, Z. Z., et al. (2021). Glioblastoma as an age-related neurological disorder in adults. *Neuro-Oncology Adv.* 3, vdab125–13. doi:10.1093/noajnl/vdab125

Landgraf, M., McGovern, J. A., Friedl, P., and Hutmacher, D. W. (2018). Rational design of mouse models for cancer research. *Trends Biotechnol.* 36, 242–251. doi:10.1016/J.TIBTECH.2017.12.001

Lin, H., Yang, Y., Hou, C., Huang, Y., Zhou, L., Zheng, J., et al. (2021). Validation of the functions and prognostic values of synapse-associated proteins in lower-grade glioma. *Biosci. Rep.* 41, BSR20210391–16. doi:10.1042/BSR20210391

Lin, J., Wen, Y., He, S., Yang, X., Zhang, H., and Zhu, H. (2019). Pipelines for cross-species and genome-wide prediction of long noncoding RNA binding. *Nat. Protoc.* 14, 795–818. doi:10.1038/s41596-018-0115-5

Liu, S. J., Dang, H. X., Lim, D. A., Feng, F. Y., and Maher, C. A. (2021). Long noncoding RNAs in cancer metastasis. *Nat. Rev. Cancer* 21, 446–460. doi:10.1038/s41568-021-00353-1

Ma, X., Liu, Y., Liu, Y., Alexandrov, L. B., Edmonson, M. N., Gawad, C., et al. (2018). Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nat* 5557696, 371–376. doi:10.1038/nature25795

Mak, I. W. Y., Evaniew, N., and Ghert, M. (2014). Lost in translation: animal models and clinical trials in cancer treatment. *Am. J. Transl. Res.* 6, 114–118.

Mattioli, K., Volders, P. J., Gerhardinger, C., Lee, J. C., Maass, P. G., Melé, M., et al. (2019). High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity. *Genome Res.* 29, 344–355. doi:10.1101/gr.242222.118

Maxwell, E. K., Schnitzler, C. E., Havlak, P., Putnam, N. H., Nguyen, A. D., Moreland, R. T., et al. (2014). Evolutionary profiling reveals the heterogeneous origins of classes of human disease genes: implications for modeling disease genetics in animals. *BMC Evol. Biol.* 14, 212. doi:10.1186/s12862-014-0212-1

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., et al. (2015). MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. *Nat. Commun.* 6, 7743–7817. doi:10.1038/ncomms8743

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., et al. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505, 635–640. doi:10.1038/nature12943

Neftel, C., Laffy, J., Filbin, M. G., Hara, T., Shore, M. E., Rahme, G. J., et al. (2019). An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* 178, 835–849. doi:10.1016/j.cell.2019.06.024

Onken, M. D., Winkler, A. E., Kanchi, K., Chalivendra, V., Law, J. H., Rickert, C. G., et al. (2014). A surprising cross-species conservation in the genomic landscape of mouse and human oral cancer identiﬁes a transcriptional signature predicting metastatic disease. *Clin. Cancer Res.* 20, 2873–2884. doi:10.1158/1078-0432.CCR-14-0205

Osswald, M., Jung, E., Sahm, F., Solecki, G., Venkataramani, V., Blaes, J., et al. (2015). Brain tumour cells interconnect to a functional and resistant network. *Nature* 528, 93–98. doi:10.1038/nature16071

Ostrom, Q. T., Gittleman, H., Liao, P., Rouse, C., Chen, Y., Dowling, J., et al. (2014). CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. *Neuro. Oncol.* 16, iv1–i63. doi:10.1093/neuonc/nou223

Ostrom, Q. T., Price, M., Neff, C., Cioffi, G., Waite, K. A., Kruchko, C., et al. (2022). CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2015–2019. *Neuro. Oncol.* 24, v1–v95. doi:10.1093/NEUONC/NOAC202

Pan, Y., Zhang, J. H., Zhao, L., Guo, J. C., Wang, S., Zhao, Y., et al. (2020). A robust two-gene signature for glioblastoma survival prediction. *J. Cell. Biochem.* 121, 3593–3605. doi:10.1002/jcb.29653

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Sci. (80-)* 344, 1396–1401. doi:10.1126/science.1254257

Rashighi, M., and Harris, J. E. (2017). Understanding the role of lncRNAs in nervous system development. *Physiol. Behav.* 176, 139–148. doi:10.1053/j.gastro.2016.08.014

Reitman, Z. J., Winkler, F., and Elia, A. E. H. (2018). New directions in the treatment of glioblastoma. *Semin. Neurol.* 38, 50–61. doi:10.1055/s-0038-1623534

Robles-Espinoza, C. D., and Adams, D. J. (2014). Cross-species analysis of mouse and human cancer genomes. *Cold Spring Harb. Protoc.* 2014, 350–358. doi:10.1101/pdb.top078824

Sanchez-Taltavull, D., Perkins, T. J., Dommann, N., Melin, N., Keogh, A., Candinas, D., et al. (2020). Bayesian correlation is a robust gene similarity measure for single-cell RNA-seq data. *Nar. Genomics Bioinforma.* 2, lqaa002. doi:10.1093/nargab/lqaa002

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502. doi:10.1038/nbt.3192

Statello, L., Guo, C. J., Chen, L. L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22, 96–118. doi:10.1038/s41580-020-00315-9

Sun, D., Guan, X., Moran, A. E., Wu, L. Y., Qian, D. Z., Schedin, P., et al. (2021). Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat. Biotechnol.* 40, 527–538. doi:10.1038/s41587-021-01091-3

Tirosh, I., Venteicher, A. S., Hebert, C., Escalante, L. E., Patel, A. P., Yizhak, K., et al. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* 539, 309–313. doi:10.1038/nature20123

Ulhas Nair, N., Cheng, K., Naddaf, L., Sharon, E., Pal, L. R., Rajagopal, P. S., et al. (2022). Cross-species identification of cancer resistance-associated genes that may mediate human cancer risk. *Sci. Adv.* 8, 1–13. doi:10.1126/sciadv.abj7176

Venkataramani, V., Schneider, M., Giordano, F. A., Kuner, T., Wick, W., Herrlinger, U., et al. (2022a). Disconnecting multicellular networks in brain tumours. *Nat. Rev. Cancer* 22, 481–491. doi:10.1038/s41568-022-00475-0

Venkataramani, V., Tanev, D. I., Strahle, C., Studier-Fischer, A., Fankhauser, L., Kessler, T., et al. (2019). Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature* 573, 532–538. doi:10.1038/s41586-019-1564-x

Venkataramani, V., Yang, Y., Schubert, M. C., Reyhan, E., Tetzlaff, S. K., Wißmann, N., et al. (2022b). Glioblastoma hijacks neuronal mechanisms for brain invasion. *Cell* 185, 2899–2917.e31. doi:10.1016/j.cell.2022.06.054

Venkatesh, H. S., Johung, T. B., Mallick, P., Monje, M., Venkatesh, H. S., Johung, T. B., et al. (2015). Neuronal activity promotes glioma growth through neuroligin-3 secretion. *Cell* 161, 803–816. doi:10.1016/j.cell.2015.04.012

Venkatesh, H. S., Morishita, W., Geraghty, A. C., Silverbush, D., Gillespie, S. M., Arzt, M., et al. (2019). Electrical and synaptic integration of glioma into neural circuits. *Nature* 573, 539–545. doi:10.1038/s41586-019-1563-y

Venteicher, A. S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M. G., et al. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* 355, eaai8478. doi:10.1126/science.aai8478

Wang, M., Zhao, J., Zhang, L., Wei, F., Lian, Y., Wu, Y., et al. (2017). Role of tumor microenvironment in tumorigenesis. *J. Cancer* 8, 761–773. doi:10.7150/jca.17648

Wen, Y., Wu, Y., Xu, B., Lin, J., and Zhu, H. (2022). Fasim-LongTarget enables fast and accurate genome-wide lncRNA/DNA binding prediction. *Comput. Struct. Biotechnol. J.* 20, 3347–3350. doi:10.1016/j.csbj.2022.06.017

Weng, Q., Wang, J., Wang, J., He, D., Cheng, Z., Zhang, F., et al. (2019). Single-cell transcriptomics uncovers glial progenitor diversity and cell fate determinants during development and gliomagenesis. *Cell Stem Cell* 24, 707–723. doi:10.1016/j.stem.2019.03.006

West, J. A., Davis, C. P., Sunwoo, H., Simon, M. D., Sadreyev, R. I., Wang, P. I., et al. (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* 55, 791–802. doi:10.1016/J.MOLCEL.2014.07.012

Woo, X. Y., Giordano, J., Srivastava, A., Zhao, Z. M., Lloyd, M. W., de Bruijn, R., et al. (2021). Conservation of copy number profiles during engraftment and passaging of patient-derived cancer xenografts. *Nat. Genet.* 53, 86–99. doi:10.1038/s41588-020-00750-6

Wu, T., and Dai, Y. (2017). Tumor microenvironment and therapeutic response. *Cancer Lett.* 387, 61–68. doi:10.1016/j.canlet.2016.01.043

Yao, Y., Ma, J., Xue, Y., Wang, P., Li, Z., Liu, J., et al. (2015). Knockdown of long non-coding RNA XIST exerts tumor-suppressive functions in human glioblastoma stem cells by up-regulating miR-152. *Cancer Lett.* 359, 75–86. doi:10.1016/j.canlet.2014.12.051

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: an R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Zeng, Q., Michael, I. P., Zhang, P., Saghafinia, S., Knott, G., Jiao, W., et al. (2019). Synaptic proximity enables NMDAR signalling to promote brain metastasis. *Nature* 573, 526–531. doi:10.1038/s41586-019-1576-6

Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., et al. (2019). CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* 47, D721-D728–D728. doi:10.1093/nar/gky900

Zhou, Y., Wu, S., Liang, C., Lin, Y., Zou, Y., Li, K., et al. (2015). Transcriptional upregulation of microtubule-associated protein 2 is involved in the protein kinase A-induced decrease in the invasiveness of glioma cells. *Neuro. Oncol.* 17, 1578–1588. doi:10.1093/neuonc/nov060

Zimmer-Bensch, G. (2019). Emerging roles of long non-coding RNAs as drivers of brain evolution. *Cells* 8, 1399. doi:10.3390/cells8111399