

ARTICLE OPEN

Proneural and mesenchymal glioma stem cells display major differences in splicing and lncRNA profiles

Gabriela D. A. Guardia¹, Bruna R. Correa^{1,4}, Patricia Rosa Araujo², Mei Qiao², Suzanne Burns², Luiz O. F. Penalva^{2,3*} and Pedro A. F. Galante^{1*}

Therapy resistance and recurrence in high-grade gliomas are driven by their populations of glioma stem cells (GSCs). Thus, detailed molecular characterization of GSCs is needed to develop more effective therapies. We conducted a study to identify differences in the splicing profile and expression of long non-coding RNAs in proneural and mesenchymal GSC cell lines. Genes related to cell cycle, DNA repair, cilium assembly, and splicing showed the most differences between GSC subgroups. We also identified genes distinctly associated with survival among patients of mesenchymal or proneural subgroups. We determined that multiple long non-coding RNAs with increased expression in mesenchymal GSCs are associated with poor survival of glioblastoma patients. In summary, our study established critical differences between proneural and mesenchymal GSCs in splicing profiles and expression of long non-coding RNA. These splicing isoforms and lncRNA signatures may contribute to the uniqueness of GSC subgroups, thus contributing to cancer phenotypes and explaining differences in therapeutic responses.

npj Genomic Medicine (2020)5:2; <https://doi.org/10.1038/s41525-019-0108-5>

INTRODUCTION

High-grade (Grades III and IV) gliomas are the most common malignant brain tumors in adults. Glioblastoma (GBM, grade IV) in particular is highly invasive and refractory to conventional therapy; GBM patients have an average survival of 15 months.¹ Therapy resistance and relapse are driven by glioma stem cells (GSCs), which comprise a small subpopulation of tumorigenic cells displaying stem-like properties: self-renewal, persistent proliferation, and ability to generate progeny of multiple lineages.² Therefore, characterization of their biological properties, expression profile, and regulation is critical for creating new therapeutic strategies. GSCs are categorized based on molecular and phenotypic differences.³ For instance, mesenchymal (MES) GSCs have higher rates of proliferation in vitro,⁴ mice that received MES GSCs developed brain tumors at a much faster rate,⁴ and MES GSCs are more resistant to radiation than proneural (PN) GSCs.⁴ In addition, primary PN GBM, originally responsive to treatment, may relapse as MES tumors which become refractory to treatment. Two explanations for this change have been proposed: (i) PN-MES transition, in which PN GSCs are triggered to switch to a MES phenotype upon treatment; and (ii) tumor heterogeneity: MES GSCs already in primary PN tumors are more resistant to treatment and then take over, driving growth of secondary tumors.⁵

Transcriptomic analyses of GSCs have established a list of subtype-specific markers⁶ and defined changes in their expression levels in response to radiation.⁷ Additional studies have focused on mutation profiles and methylation status.⁸ Here we expand the characterization of GSCs by focusing on RNA-mediated mechanisms. Splicing profiles of MES and PN GSC lines show major differences affecting genes implicated in cell cycle regulation, DNA repair, cilium assembly, and RNA splicing. Additionally, we found long non-coding RNAs (lncRNAs) preferentially expressed in each GSC subgroup, with some exhibiting prognostic value.

RESULTS

Splicing profiles define GSC subgroups

To assess the contribution of alternative splicing to GSC phenotypes and identify relevant differences between MES and PN GSCs, we performed RNA sequencing of three mesenchymal (MES-83, MES-326, MES-1123) and three PN GSC (PN-19, PN-157, and PN-528) cell lines (Supplementary Table 1). RNA-seq analysis revealed that MES and PN GSCs showed differences in 4934 splicing events affecting 3253 genes ($|\Delta\text{PSI}| > 0.1$ and false discovery rate [FDR] < 0.05 , likelihood-ratio test); among these, 1793 events were not reported in the reference transcriptome (GENCODE version 26; Supplementary Table 2). Using only the splicing profiling, we correctly clustered samples according to their subgroups (Fig. 1a). MES and PN GSCs showed a similar number of genes presenting alternative splicing (AS) events of exon skipping (ES), mutually exclusive exons (MXE), intron retention (IR), and alternative 5'/3' splice sites (ASS; Fig. 1b). We randomly selected exon-skipping events to be validated by quantitative RT-PCR (Supplementary Table 3, Fig. 1c–f). All were in concordance (higher fold change, on average) with the ES “direction” indicated by RNA-seq data: ELMOD3, FDF1, OSBPL6, and RAB18 have skipped exons in PN GSCs (Fig. 1c, d; Supplementary Fig. 1a), while GSN, SLC9A5, and WASF3 have skipped exons in MES GSCs (Fig. 1e, f; Supplementary Fig. 1b).

Moreover, we used GBM samples from The Cancer Genome Atlas (TCGA) to determine which AS events found in PN GSC cell lines compared to MES are also confirmed in GBM. We found 78.7% (3,882/4,934; Supplementary Table 4) of AS isoforms from GSCs in GBM samples. Among the AS subtypes, IR and alternative splice sites (ASS) events were more highly match to GBM (>90%; Fig. 2a, b, classes in green plus yellow) than ES and MXE (~75%; Fig. 2a, b, classes in green plus yellow). Additionally, ~35% of AS events from MES GSC have agreement with GBM mesenchymal AS events (Fig. 2a, classes in yellow) and ~50% of AS events from PN GSC have

¹Centro de Oncologia Molecular, Hospital Sírio-Libanês, São Paulo, São Paulo 01309-060, Brazil. ²Children’s Cancer Research Institute, UT Health San Antonio, San Antonio, TX 78229, USA. ³Department of Cell Systems and Anatomy, UT Health San Antonio, San Antonio, TX 78229, USA; ⁴Present address: Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona 08003 Catalonia, Spain. *email: penalva@uthscsa.edu; pgalante@mochsl.org.br

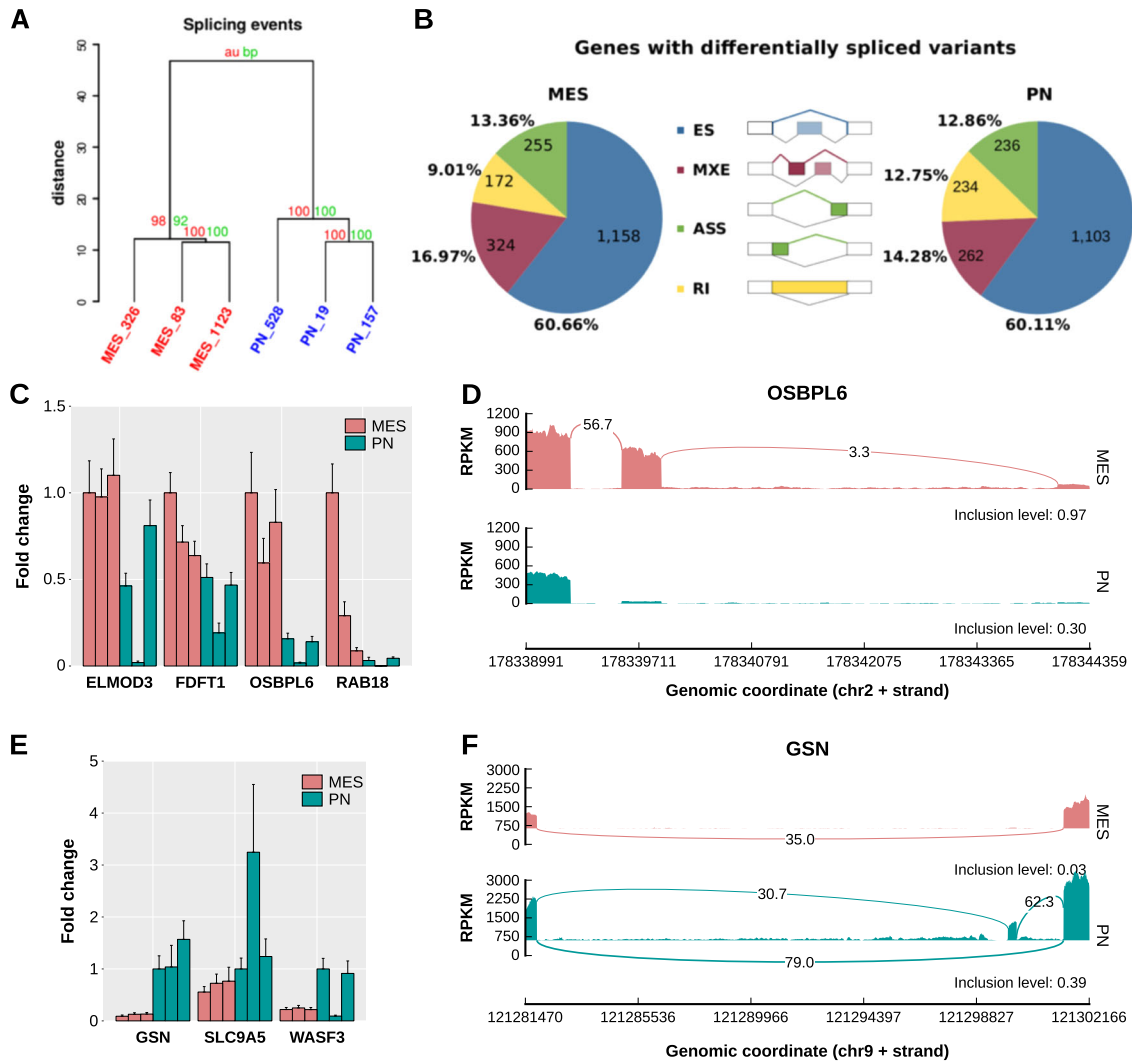


Fig. 1 Mesenchymal and proneural GSCs have distinct alternative splicing profiles, but a similar number of genes harboring alternative splicing events. **a** Hierarchical cluster showing that GSCs can be grouped based on their alternative splicing profile (PSI values). Bootstrap (bp) and approximately unbiased (AU) probability from pvclust.⁸⁹ **b** Percentage of events and number of genes harboring exon skipping (ES), mutually exclusive exons (MXE), alternative 3' or 3' splice site (ASS), and retained introns (RI) events in MES or PN GSCs. **c** Quantitative RT-PCR (qRT-PCR) validation of four exon-skipping events in PN and MES GSCs. ELMOD3, FDF1, OSBPL6, and RAB18 are official gene names. **d** An exon-skipping event (gene OSBPL6; second exon) in PN and MES GSCs. **e** qRT-PCR validation of three exon-skipping events in MES and PN GSCs. GSN, SLC9A5, and WASF3 are official gene names. **f** An exon-skipping event (gene GSN; second exon) in MES and PN GSCs.

agreement with GBM proneural AS events (Fig. 2b, classes in yellow). By performing gene ontology (GO) analysis with protein-coding genes sharing agreement in AS events ($|\Delta\text{PSI}| > 0.1$ and $\text{FDR} < 0.05$, likelihood-ratio test) between PN GSC and GBM versus MES GSC and GBM, several biological processes commonly involved in brain tumors (e.g., neurogenesis, cell division and epithelial-mesenchymal transition)⁹ were enriched ($\text{FDR} < 0.05$, Fisher's exact test) (Supplementary Fig. 2), indicating that AS events in MES and PN GSCs are reliable and potentially important to GBM maintenance.

Next, we evaluated if differences in splicing profile between PN and MES GSCs affect mRNA levels. First, we identified 6393 differentially expressed genes between MES and PN GSCs, 5,625 of which were multi-exonic: 2545 were upregulated in MES GSCs and 3,080 were upregulated in PN GSCs ($|\log_2\text{FC}| > 1$ and $\text{FDR} < 0.05$, Wald test; Supplementary Table 5). We also explored differences in gene expression observed between MES and PN GSCs in MES and PN GBM samples from TCGA. We confirmed higher expression of 756 and 924 genes in MES and PN GBM ($|\log_2\text{FC}| > 0$; $\text{FDR} < 0.05$, Wald test; Supplementary Table 6), respectively. Considering the heterogeneity in GBM and that GSCs represent only a subpopulation

of their cells, this low agreement (29.7%; $(756 + 924)/5625$) in gene sets with differential expression between PN or MES GSCs and PN or MES GBMs was expected. Among splicing variants prevalent in MES GSCs ($|\Delta\text{PSI}| > 0.1$ and $\text{FDR} < 0.05$, likelihood-ratio test), 195 genes (11.89%) had higher levels of expression, while 235 genes (14.33%) had lower levels of expression in MES compared to PN GSCs. In the case of splicing variants prevalent in PN GSCs ($|\Delta\text{PSI}| > 0.1$ and $\text{FDR} < 0.05$, likelihood-ratio test), 256 genes (15.87%) were upregulated and 156 genes (9.67%) downregulated in PN GSCs compared to MES GSCs (Fig. 2c, d, Supplementary Table 7). Overall, splicing variants prevalent in MES GSCs were more often associated with down-regulated genes in the same subgroup compared to splicing variants prevalent in PN GSCs (p -value = 0.0001; chi-square 28.83; d.f. = 1).

Genes displaying splicing differences between PN and MES GSCs are implicated in survival of GBM patients
We investigated whether expression levels of protein-coding genes with splicing differences in MES vs. PN GSCs influence survival of

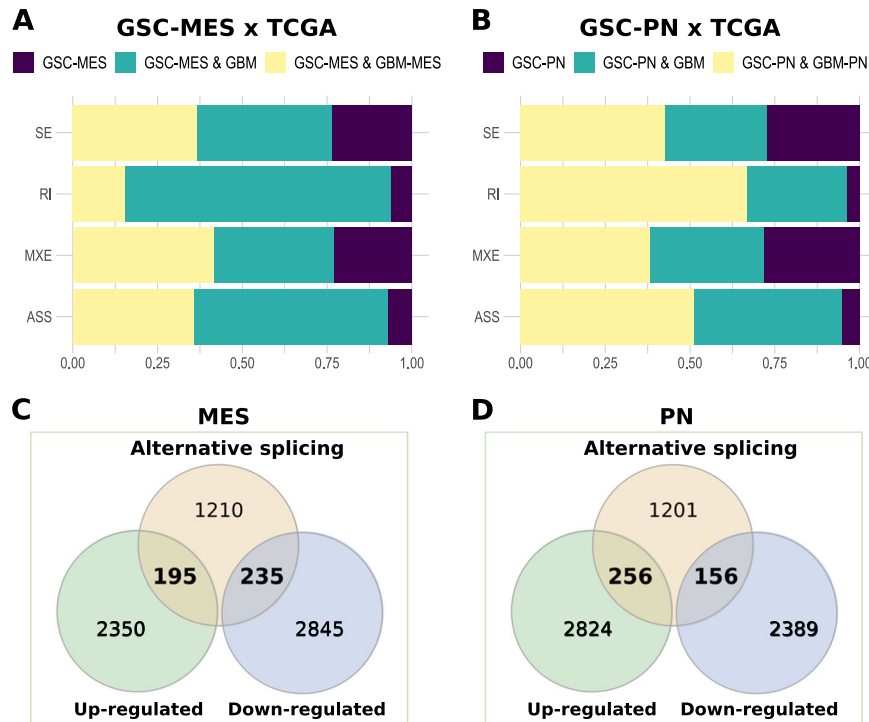


Fig. 2 Proneural and mesenchymal GSCs share most of their AS events with GBM samples. **a** Percentage of alternative splicing events from GSCs in GBM samples. GSC-MES & GBM-MES contains AS events found in GSC MES and GBM (molecular subtype) mesenchymal. GSC-PN & GBM-PN contains AS events found in GSC PN and GBM (molecular subtype) proneural. GSC-MES & GBM and GSC-PN & GBM contains AS events found in GSC-MES and GSC-PN and GBM samples (all molecular subtypes), respectively. GSC-MES and GSC-PN contain AS events not found in GBM samples. **b** Intersection between differentially spliced and differentially expressed genes (up- and downregulated) in the two GSC subtypes.

GBM patients with one type of GBM but not the other one. Using data from TCGA, we created Kaplan-Meier survival curves to separately explore correlations between expression levels of each gene and patient survival. Genes exclusively associated with prognosis of MES or PN GBM were then included in multivariate Cox proportional-hazards regression models to adjust for the effects of clinical variables. Several genes exclusively affected prognosis of patients with PN versus MES GBM (log-rank p -value < 0.05 and multivariate Cox p -value < 0.05 ; Supplementary Table 8, Fig. 3a–c). Most genes predicted prognostic outcomes exclusively in PN GBM and present more splicing events in PN GSCs (Fig. 3b). In particular, 11 genes that are upregulated in PN GSCs are also associated with prognostic outcome in patients with PN but not MES GBM. Six genes were associated with poor prognostic outcomes, i.e., samples with higher expression of these genes (based on their median expression values) were associated with decreased survival compared to samples exhibiting lower expression of that gene. The other five genes were associated with increased survival when highly expressed. Co-expression analysis in TCGA GBM samples revealed weak or no correlations among these genes ($|\text{Spearman's coefficient}| < 0.8$; Supplementary Table 9), suggesting that their impact on prognosis is independent. In the univariate analysis, only LRRFIP1, which is more highly expressed in MES GSCs, predicted prognosis in both GBM molecular subgroups. However, when considered with other clinical parameters in the multivariate Cox analysis, this gene remained significantly associated only with prognosis of MES GBM.

Genes displaying splicing differences in PN and MES GSCs are implicated in mRNA splicing, DNA repair, cell division, and cilium assembly

To elucidate the biological function of genes with differences in splicing profiles between PN and MES GSCs, we first focused on

cancer-related genes (Supplementary Fig. 3, Supplementary Table 10), using a consensus list of cancer driver genes identified in three studies.^{10–12} From this list, we identified 72 genes displaying splicing differences between PN and MES GSCs; 26 of these also showed differences in expression (Supplementary Table 11).

We also conducted GO analysis using the DAVID web tools¹³ and protein network analysis using the STRING database¹⁴ to determine if genes displaying differences in splicing levels are associated with particular biological processes. For genes displaying differences exclusively in splicing profiles and no significant changes in expression, the top enriched terms were cell cycle, cilium assembly, DNA repair, transcription, and mRNA splicing (FDR < 0.05 , Fisher's exact test; Fig. 4a, Supplementary Table 12). Genes showing differences in expression levels without significant splicing events were associated with a different set of GO terms, in particular cell adhesion and neuronal function (FDR < 0.05 , Fisher's exact test; Fig. 4a, Supplementary Table 13). Next, we conducted protein interaction analyses with gene sets associated with DNA repair, cell cycle, cilium assembly, and mRNA splicing; all showed a highly connected gene network (Fig. 4b–e).

Splicing regulators potentially driving isoform preference

To determine which splicing regulators are potentially driving different splicing profiles in PN versus MES GSCs, we analyzed the expression and splicing profiles of 388 splicing-related RBPs, comprising 66 components of the spliceosome¹⁵ (Supplementary Table 14). Among them, 50 (12.9%) were differentially expressed: 35 were upregulated in PN GSCs while 15 were upregulated in MES GSCs ($|\log_2\text{FC}| > 1$ and FDR < 0.05 , Wald test; Fig. 5a, b, Supplementary Table 15). From this group, 19 RBPs showed the same differences in expression in GBM MES vs. PN samples from TCGA (Supplementary Table 15). We also identified 188 events

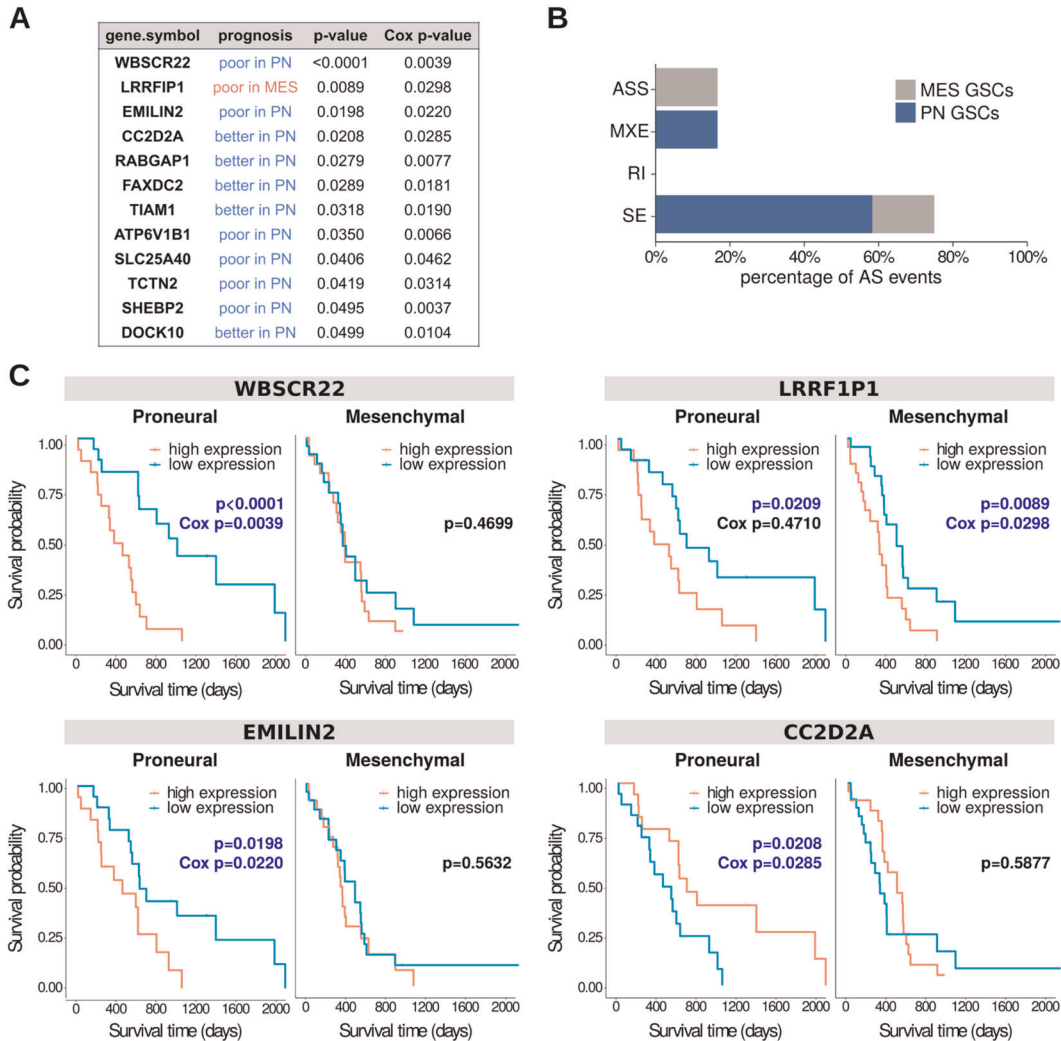


Fig. 3 Differentially expressed genes in mesenchymal or proneural GSCs are related to GBM patient prognosis of patients with GBM. **a** Protein-coding genes with higher expression in proneural GSCs (showed in blue) or mesenchymal GSCs (showed in red) are exclusively associated with better versus poor prognosis in GBM patients with tumors of molecular subtypes proneural or mesenchymal subtypes of GBM, respectively. **b** Genes associated with survival in GBM patients are listed by type of alternative splicing event detected in the GSCs: ES, exon skipping; IR, intron retention; MXE, mutually exclusive exons; ASS, alternative splice sites 3' or 5'. **c** Survival curves of four genes (presented in **a**) exclusively affecting patients with GBM proneural or mesenchymal GBM. Significant log-rank p -values and multivariate Cox p -values are shown in blue.

affecting 95 splicing-related RBPs that differed between PN and MES GSCs (Fig. 5a) ($|\Delta\text{PSI}| > 0.1$ and $\text{FDR} < 0.05$, likelihood-ratio test; Supplementary Table 16). Moreover, 39 RBPs from the group of splicing factors showing expression and/or splicing differences between PN and MES GSCs were identified in recent meta-studies that used TCGA data to map mutations affecting multiple tumor types^{16,17} (Fig. 5c, Tables S17 and S18).

We investigated co-expression patterns among differentially expressed RBPs in MES and PN GSCs using Spearman's rank correlation (Supplementary Fig. 4a, b, respectively). Correlation patterns were then confirmed in GBM samples from TCGA. We found two sets of positively correlated RBPs in MES GBM (Fig. 6a, b) and two other sets in PN GBM (Fig. 6c, d). Network analysis showed that these correlated RBPs also display functional interaction (Fig. 6e, f, respectively), indicating that these groups of related RBPs may act together to regulate distinct sets of splicing events and may regulate one another by modulating inclusion of some alternative exons.¹⁸

lncRNAs display differences in expression and isoform preferences in PN vs. MES GSCs

Aberrant expression of lncRNAs has been described in numerous cancer types, and a growing number of lncRNAs are implicated in malignant transformation.¹⁹ We investigated expression of lncRNAs in MES and PN GSCs. First, we found 1240 differentially expressed lncRNAs between MES and PN GSCs (Fig. 7a, Supplementary Table 19) and clustered GSCs based only on their lncRNA profile (Supplementary Fig. 5). Next, we investigated whether these differences are also observed in MES versus PN GBM tumors from TCGA ($|\log_2\text{FC}| > 0$ and $\text{FDR} < 0.05$, Wald test). We found 357 genes in concordance between GSCs and GBM tumors (i.e., upregulated in MES GSC and upregulated in MES GBM; downregulated in MES GSC and downregulated in MES GBM; same pattern as PN GSCs and PN GBM; Supplementary Table 20). The levels of agreement (28.7%; 357/1240) for differentially expressed lncRNAs were similar to those for coding genes (29.7%) and splicing isoforms between PN and MES GSCs and PN and MES GBM tumors.

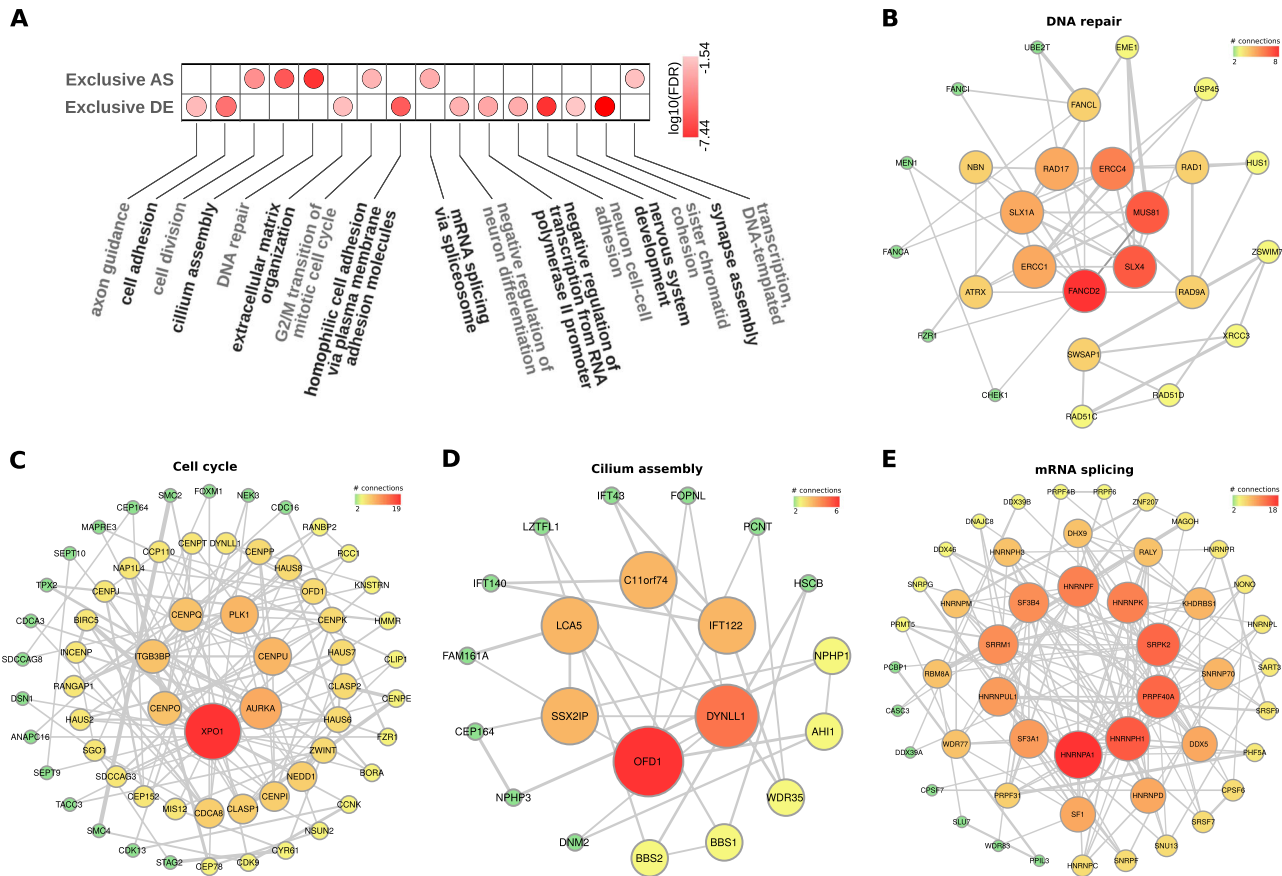


Fig. 4 Proneural and mesenchymal GSCs have genes with alternative splicing and differential expression related to key biological processes in tumorigenesis and sets of highly connected networks of genes. **a** Gene Ontology (Biological Process) enrichment analysis of genes harboring alternative splicing events and genes differentially expressed genes in proneural versus mesenchymal types of GSCs. The networks based on protein-protein interactions display genes associated with **b** DNA repair; **b** cell cycle; **c** cilium assembly; and **d** mRNA splicing.

We also queried all lncRNAs differentially expressed between MES and PN GSCs to determine if their expression levels might have prognostic value in GBM and other tumor types (Supplementary Table 21). We used univariate Cox proportional-hazards regression models to explore correlations between expression of lncRNAs and patient survival (Cox p -value < 0.05). Each lncRNA significantly correlated with GBM survival was then included in a multivariate Cox model with multiple covariates relevant to GBM prognosis. We created Kaplan-Meier survival curves for lncRNAs that remained significantly associated with patient survival in multivariate analyses (Cox p -value < 0.05). Six lncRNAs had significant association with prognosis of GBM (Cox p -value < 0.05; Supplementary Table 22, Supplementary Fig. 6). Of these, 5 (83.3%) displayed increased expression in MES GSCs and 5 (83.3%) were also linked to survival in other tumor types (Fig. 7b, c). In a second analysis, we evaluated the prognostic value of differentially expressed lncRNAs separately in MES and PN subgroups. Three and five lncRNAs were associated with survival rates in MES and PN GBM samples, respectively (multivariate Cox p -value < 0.05, log-rank p -value < 0.05; Supplementary Table 23, Figs. S7 and S8).

Despite having fewer exons per transcript on average compared to coding genes, alternative splicing occurs relatively often in lncRNAs.²⁰ We found 249 alternative splicing events (108 lncRNAs) that differed between MES and PN GSCs ($|\Delta\text{PSI}| > 0.1$ and FDR < 0.05, likelihood-ratio test; Fig. 7A; Supplementary Table 24). When we compared differentially expressed and differentially spliced lncRNAs in MES and PN GSCs, 34 appeared on both lists (13 lncRNAs upregulated in MES GSCs and 21 upregulated in PN GSCs;

Supplementary Table 25). A few lncRNAs displayed differences in multiple splicing events; these were GAS5 (23 events), PVT1 (5 events), TPT1-AS1 (14 events), PSMA3-AS1 (10 events), and CD27-AS1 (6 events).

Some lncRNAs are particularly complex. For example, GAS5 harbors several snoRNAs in its transcripts.²¹ The splicing differences observed between PN and MES (IR events) affect expression of snorD79, snorD80, and snorD47 (Fig. 7d). The latter has been described as a potential tumor suppressor in GBM, and associated with reduced proliferation, invasion and epithelial-mesenchymal transition, and induction of G2 cell-cycle arrest.²²

DISCUSSION

Alternative splicing affects all hallmarks of cancer, and splicing regulators have been shown to function as drivers of glioblastoma development.²³ Importantly, there are several examples of inhibitors of the splicing machinery and their use in cancer therapy has started to be explored.^{24,25} Here, we provide evidence that MES and PN GSCs display specific splicing profiles. Splicing differences preferentially affect genes implicated in cell cycle regulation, DNA repair, splicing, and cilium formation, which are all critical processes in therapy response and tumor relapse. We suggest that these sets of splicing events are important contributors to MES and PN features and differences in responding to radiation and temozolomide. In agreement, we observed that several genes whose splicing profiles differ between MES and PN GSCs show distinct association with survival in each subgroup.

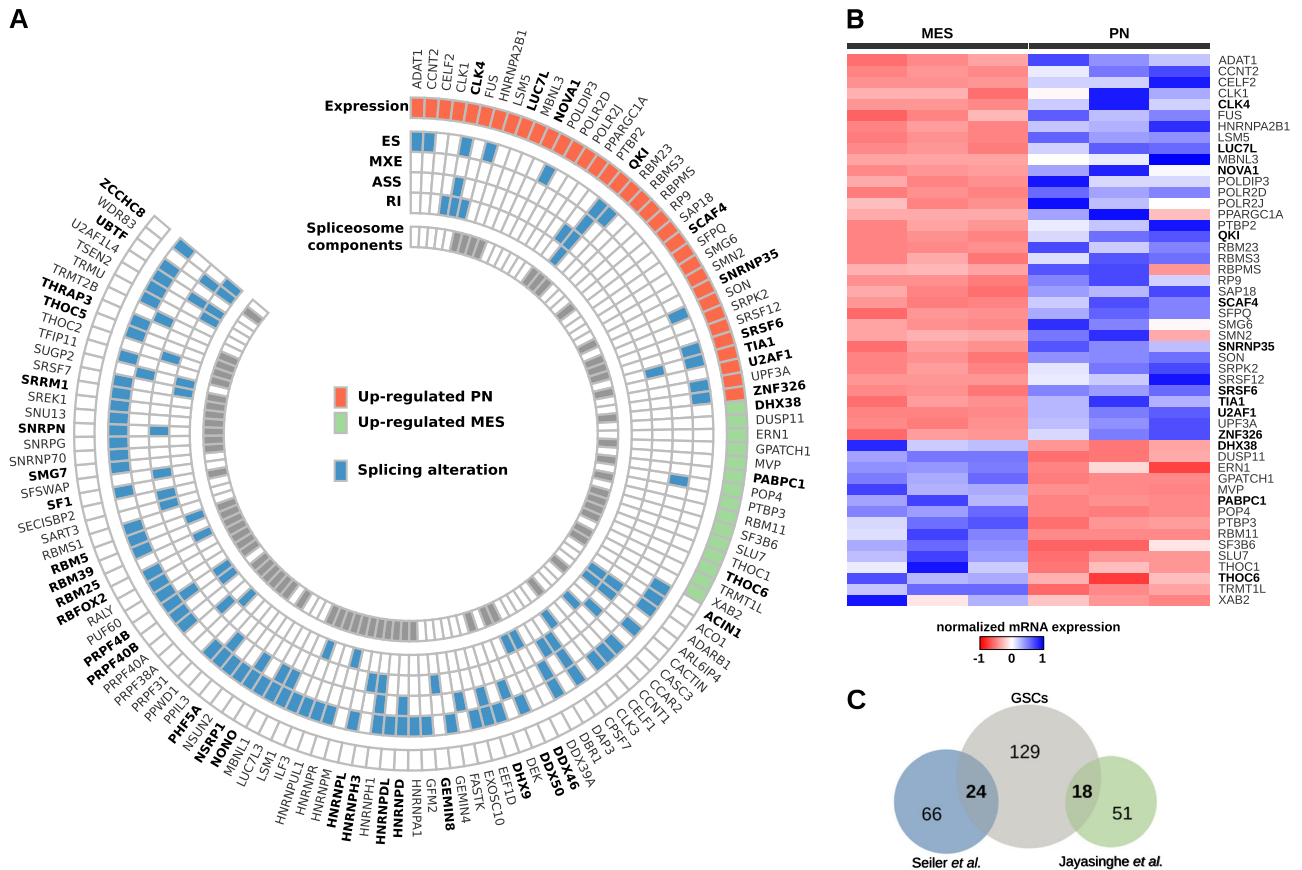


Fig. 5 Mesenchymal and proneural GSCs have splicing-related genes with distinct profile of alternative splicing and/or differentially expressed. **a** Splicing-related genes differentially expressed and/or harboring alternative splicing events in mesenchymal versus proneural GSCs. Mutated genes are shown in bold. **b** Differentially expressed splicing-related genes in MES and PN GSCs. Genes reported to harbor mutations are shown in bold. **c** Number of splicing-related genes in our study that harbor mutations according to studies by Seiler and colleagues¹⁶ and Jayasinghe and colleagues¹⁷.

GBM tumors contain radio-resistant GSCs that display increased pair capacity and, upon radiation, activate DNA damage checkpoint signals. MES GSCs are more resistant to radiation than PN GSCs, which can be explained based on differences in their expression profiles.^{26–29} A recent analysis with paired GBM samples (primary and recurrent) indicated that relapse was often associated with significant alterations in the expression of DNA repair and cell-cycle genes.³⁰ In concordance, our analysis showed that multiple DNA repair genes display splicing differences between MES and PN GSCs. In particular, genes in DNA damage and integrity checkpoints are enriched among those with differences in splicing profiles. Regarding major nodes in the DNA repair network (Fig. 4a), ERCC1, FANCD2, and RAD17 have been associated with therapy resistance in gliomas and other tumor types. In a preclinical study, inhibition of ERCC1 significantly decreased tumor growth and sensitized cells to chemotherapy with cisplatin and temozolomide, increasing cell death by up to 25%.³¹ Also, methylation of the ERCC1 promoter was associated with sensitivity to radiation in glioma cell lines.³² Expression levels of FANCD2, a member of the Fanconi anemia DNA repair pathway, were strongly associated with glioma tumor grade, and inhibition of FANCD2 improved sensitivity to temozolomide and carmustine in glioma cells.³³ Finally, certain splicing variants of gene RAD17 promote resistance to radiotherapy in cell lines.³⁴

GSCs employ cell cycle regulation mechanisms to circumvent the effects of chemotherapy and radiotherapy.³⁵ There are critical differences between PN and MES GBM regarding cell cycle regulation that could influence therapy response. For instance, CDK4/6 inhibition with palbociclib preferentially inhibited cell

proliferation of PN GSCs.³⁶ Alterations in cell cycle checkpoint can contribute to GSC resistance to radiation.³⁷ Higher expression of RAD51 after exposure to radiation was observed in radio-resistant GSCs. GSCs resistant and sensitive to radiation have different cell cycle checkpoint responses when radiation was combined with RAD51 inhibition.³⁰ RAD51, RAD51D, and RAD51AP1 show differences in their splicing profiles in PN vs. MES GSCs. The main nodes in the cell cycle network (Fig. 4b) contain critical players in glioma progression and response to therapy. For example, high expression of XPO1, also known as CRM1, correlates with malignancy and poor survival outcome in gliomas.³⁸ High expression of XPO1 provides a growth advantage to glioma cells by promoting nuclear export of p27, a known cell cycle regulator with reduced expression in various tumors.³⁸ Another main node, PLK1, has been associated with therapy resistance in gliomas and other brain tumors when highly expressed.³⁹ In a recent study, Koncar et al. established PLK1 as a potential therapeutic target in IDH1-mutated gliomas.⁴⁰ PLK1 bypasses the temozolomide-induced DNA damage checkpoint, limiting its effectiveness. Therefore, PLK1 inhibitors may improve temozolomide efficacy.

The protein kinase encoded by AURKA is a central regulator of mitotic processes, such as chromosomal segregation, chromatin condensation, and mitotic checkpoints. AURKA expression levels correlate with malignancy grade in gliomas. Its inhibition decreases cell proliferation, induces G2/M cell cycle arrest, and produces synergistic effects with radiation in GBM cell lines.⁴¹ Another study showed that AURKA regulates self-renewal and tumorigenicity of GSCs by activating the Wnt signaling pathway.⁴² Other nodes of the cell cycle network include centromere proteins

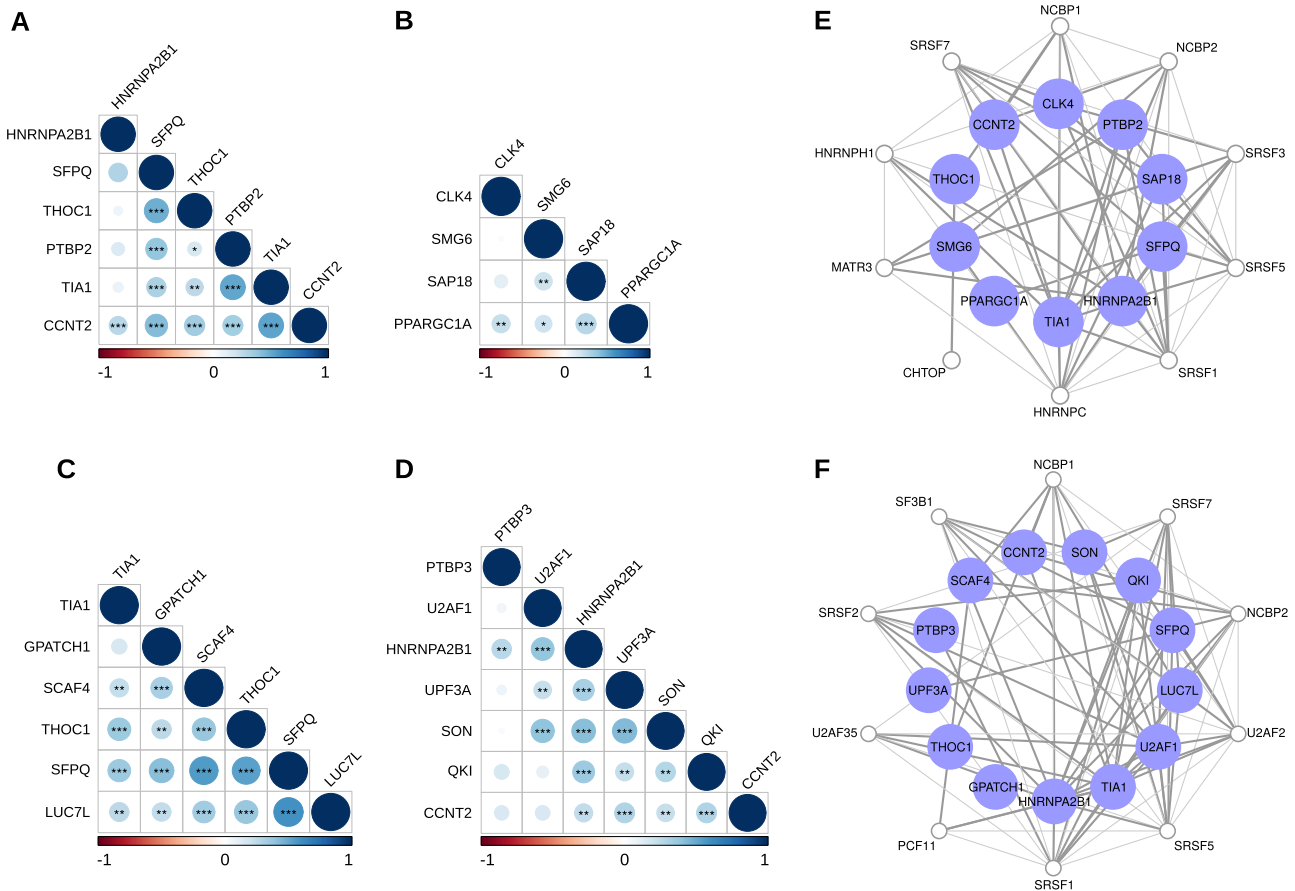


Fig. 6 RBPs co-expressed in mesenchymal and proneural GBM samples. Correlation plots display RBPs positively correlated in **(a, b)** mesenchymal GBM; and **(c, d)** proneural GBM (Spearman's significance: ** p -value < 0.01; *** p -value < 0.001). Interaction networks show RBPs functionally related from **(e)** mesenchymal GBM and **(f)** proneural GBM. Blue nodes represent selected RBPs, while white nodes were additionally included in the network to give a broader context of associations between the RBPs. Only experimentally validated associations are shown in gray.

(CENPU, CENPO, and CENPQ). Upregulation of CENPU has been reported in several malignancies, including GBM.⁴³

We found that multiple genes implicated in cilia formation show splicing differences between PN and MES GSCs. There is growing evidence that cilia participate in gliomagenesis, including cell signaling activation, cell proliferation, apoptosis, and participation in therapeutic resistance.^{44,45} The proposed function of primary cilia is to limit GBM proliferation, and its loss leads to increased proliferation.⁴⁶ However, others have proposed that cilia can induce or suppress tumorigenesis and is influenced by the oncogenic driver event.⁴⁴ Between 1 and 30% of cells in glioblastoma samples are ciliated, and study in GBM cell lines shows that they rarely gave rise to cilia.⁴⁷ Given that primary cilium has emerged as a key component in cancer development and alterations are observed during tumor development (including gliomas), the impact of splicing alterations on the function and expression of central nodes of the identified network (OFD1, DYNLL1, SSX2IP, LCA5, C11orf74, and IFT122) warrants further investigation.

We also determined that splicing differences between PN and MES GSCs preferentially affect genes implicated in splicing regulation. In the mRNA splicing network of genes with differences in splicing profiles (Fig. 4d), heterogeneous nuclear ribonucleoproteins (hnRNPs) occupy a central position. They regulate various post-transcriptional and translational processes, including alternative splicing and mRNA stabilization. hnRNPH is overexpressed in gliomas and regulates the splicing of RON and IG20, producing isoforms that promote survival, proliferation, and

migration and invasion of GBM cells.⁴⁸ hnRNPA1 has a central role in the let-7a/c-Myc/HNRNPA1/PKM2 signaling pathway via activating PKM2 expression and thus increasing aerobic glycolysis and cell proliferation in gliomas.⁴⁹ hnRNPC overexpression in more aggressive glioma cells correlates with increased migratory and invasive activities through regulation of PDCD4.⁵⁰ Similarly, hnRNPK has been associated with increased migration and invasion capabilities and MES transformation of GSCs through interactions with RTVP-1 and N-WASP.⁵¹ Finally, hnRNPM has been implicated in resistance to temozolomide in GBM.⁵²

Differences in expression levels of RBPs implicated in splicing regulation are likely the main drivers of PN and MES splicing profiles. As indicated by expression correlation and network analyses, specific groups of associated RBPs might coordinately regulate distinct groups of splicing events. We highlight a specific set of RBPs: CLK4, PHF5A, PRPF40B, QKI, THOC6, TIA1, and U2AF1. These RBPs are frequently mutated in GBM and show differential splicing and expression in MES versus PN GSCs, and in GBM samples from TCGA compared to normal brain samples from the Genotype-Tissue Expression dataset.⁵³ Expression variation of some of these RBPs has been associated with gliomas and other tumor types. In particular, CLK4 was included in a 6-gene signature that predicts cell proliferation of high-grade glioma cultures after in vitro treatment with the tyrosine kinase inhibitor sunitinib.⁵⁴ PHF5A, a component of the spliceosome machinery, maintains proper exon recognition of C-rich 3' splice sites in GSCs derived from GBM patients, and its knockdown leads to cell cycle arrest and loss of viability.⁵⁵ Depletion of PRPF40B modulates ASS

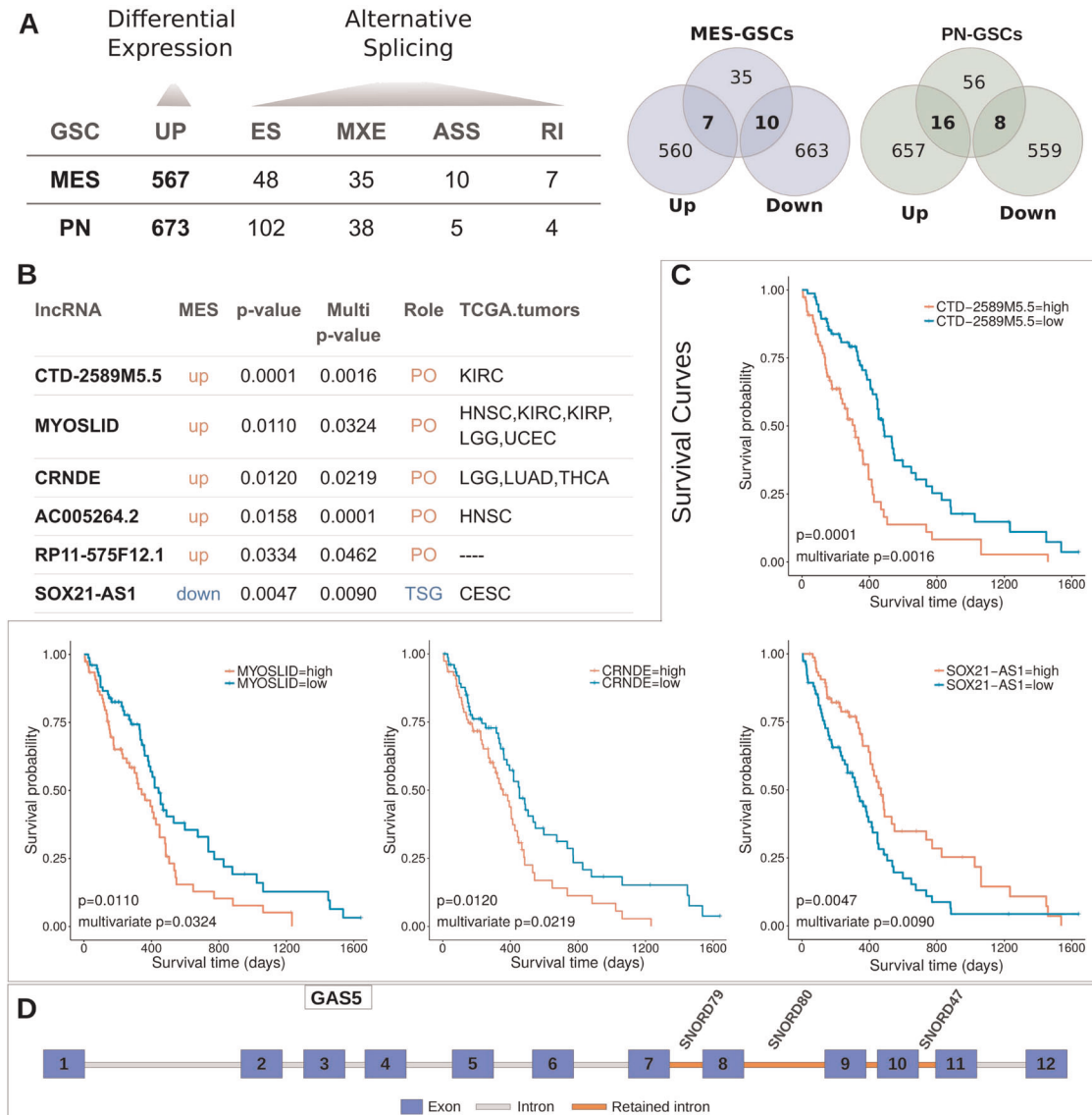


Fig. 7 Mesenchymal and proneural GSCs present lncRNAs differentially expressed, harboring alternative splicing events, and associated with prognosis in GBM patients. **a** Number of lncRNAs presenting alternative splicing events or differentially expressed (left side) and their intersections (right side) in MES or PN GSCs. **b** Differentially expressed lncRNAs in GSCs with prognostic value to GBM and other cancer patients ("TCGA tumors" follows TCGA nomenclature for cancer types (e.g., LGG low grade glioma)). PO Proto-oncogenes; TSG tumor-suppressor genes. **c** Survival curves based on median expression values of lncRNAs in GBM patients. Corresponding log-rank p -values and multivariate Cox p -values are also shown. **d** Representation of lncRNA GAS5 and its retained introns (in red) in MES GSC. Retained introns host three small nucleolar RNAs: SNORD79, SNORD80, and SNORD57.

selection of apoptotic genes through direct interactions with SF1 and U2AF2, leading to decreased cell survival.⁵⁶ Deletion of QKI, a known tumor suppressor, maintains stemness of GSCs and decreases differentiation in suboptimal environments.⁵⁷ QKI influences splicing in many solid tumors,⁵⁸ and is implicated in epithelial-to-mesenchymal transition.⁵⁹ Finally, one of the most frequently mutated RBPs, U2AF1, affects alternative splicing in different tumor types. For example, in lung adenocarcinoma, alterations in splicing driven by U2AF1 induce cell cycle dysregulation and mitotic stress.⁶⁰

Expression levels of splicing regulators preferentially affect genes implicated in RNA processing and other RNA-related processes.⁶¹ The spliceosome machinery may regulate itself by modulating inclusion of some alternative exons.¹⁸ Overall, our results indicate that alternative splicing signatures and the status of RNA processing components contribute to maintain the MES and PN phenotypes of GSCs. We have previously shown in a

functional screening in GBM that several splicing regulators are among highly expressed RBPs affecting cancer phenotypes.⁶¹ Importantly, a growing number of specific splicing factor inhibitors are being identified, and their use in cancer therapy is gaining momentum.²⁴

Differences in lncRNA expression and splicing profiles have been observed across tissues and during development, and impact several diseases, including cancer.²⁰ In the second part of our study, we identified differences in expression and splicing profiles of lncRNAs between the two GSC subgroups. We established that the lncRNA profile also defines MES and PN GSCs. Among lncRNAs displaying expression differences between PN and MES GSCs, we identified six lncRNAs associated with survival in GBM and other cancers: CTD-2589M5.5, MYOSLID, CRNDE, AC005264.2, SOX21-AS1 and RP11-575F12.1. MYOSLID, CRNDE, and SOX21-AS1 have been characterized in the context of tumorigenesis. MYOSLID was correlated with tumor size, stage,

invasion, and survival time in gastric cancer, and its knockdown inhibited tumorigenesis in mouse xenografts. MYOSLID acts as a ceRNA of miR-29c-3p, causing de-repression of anti-apoptotic gene MCL-1.⁶² MYOSLID was also defined as potential biomarker in head and neck squamous cell carcinoma, where it promotes invasion and metastasis by modulating the epithelial-mesenchymal transition.⁶³ SOX21-AS1 has been linked to tumor progression and defined as a prognostic marker in cervical cancer, nephroblastoma, hepatocellular carcinoma, and lung and oral cancer.^{64–69} In glioblastoma, SOX21-AS1 was defined along other four lncRNAs as a signature that predicts survival.⁷⁰ CRNDE is an important oncogenic lncRNA implicated in multiple malignancies including glioblastoma.^{71–73} Several relevant pathways, such as mTOR, EGFR and TLR3-NF- κ B-cytokine, are modulated by CRNDE.^{74–76} CRNDE also functions as a scaffold for chromatin-modifying complexes such as PRC2 and CoREST.⁷⁷

Altogether, our study provides a novel RNA map to be explored regarding GBM progression, PN to MES transition, and treatment response and resistance.

METHODS

Cell lines, cell culture and maintenance, RNA preparation and sequencing

We used six GSC lines previously described.⁶ Three were PN cell lines (PN-19, PN-157, and PN-528) and three were MES cell lines (MES-83, MES-326, MES-1123). PN and MES GSCs were maintained in DMEM/F12 supplemented with B27, heparin, bFGF, and EGF. Total RNA was: (i) extracted using TRIzol reagent (Life Technologies); (ii) purified with RNeasy (Qiagen), according to the manufacturer's instructions; (iii) prepared for RNA-Seq according to the manufacturer's instructions (Illumina); and (iv) sequenced using 101-bp paired-end chemistry on a HiSeq-2000 machine in the UTHSCSA Genomic Facility.

Alternative splicing analysis

To identify splicing events differentially represented between PN and MES GSC cell lines, we first mapped raw RNA sequencing reads from all samples against the human reference genome (hg38/GRCh38) and the reference transcriptome (GENCODE version 26; www.genencodegenes.org; accessed on 30 Nov 2018) using GSNAP version 2016-09-23⁷⁸ (parameters: -t 20; -B 4; -N 1; -E 1; -w 200000; -pairmax-rna 200000). Following this, reliable alignments against the genome (mapping quality score (Q) \geq 20) were selected using SAMtools.⁷⁹ To search for splicing differences between MES and PN cell lines, we used Replicate Multivariate Analysis of Transcript Splicing (rMATS),⁸⁰ which reports splicing events already reported in the reference transcriptome, and those absent in the reference transcriptome (novel events). First, we compared the two GSC subgroups and selected all differentially represented splicing events ($|\Delta\text{PSI}| > 0.1$ and FDR-adjusted p -value < 0.05 , likelihood-ratio test). Novel and known splicing variants were then classified as follows: ES, MXE, retained intron (RI), alternative donor site (A5SS) or alternative acceptor site (A3SS). Outputs from rMATS were further processed using a set of locally created Python and R scripts. A hierarchical cluster was built based on inclusion levels of splicing events (PSI values) using *dist* and *hclust* R functions (<https://www.r-project.org/>; accessed 30 Jan 2019). We sought alternative isoforms for lncRNAs, selected according to GENCODE annotations.

To differentiate splicing events between PN and MES subtypes of GBM, we first downloaded aligned reads from 49 MES and 38 PN GBM samples from TCGA. Next, reads were aligned against the human reference genome (hg38/GRCh38) and the reference transcriptome (GENCODE version 22; https://www.genencodegenes.org/human/release_22.html; accessed 1 Oct 2019) using STAR.⁸¹ For GSCs, we selected reliable alignments against the genome (mapping quality score (Q) \geq 20) using SAMtools, and searched for alternative splicing events using rMATS. Any batch-effect were detected or corrections were applied.

Gene expression analyses

Differential expression analyses were performed using DESeq2.⁸² To compare MES-GSCs and PN-GSCs, only genomic mapped reads presenting mapping quality (Q) \geq 20 (as described above) were considered. Read counts per gene were quantified using HTSeq-count,⁸³ and GENCODE v26

was used as the reference for the human transcriptome. For comparisons between GBM and normal brain samples, read counts of 156 GBM samples were obtained from TCGA (GENCODE v22) and read counts of 287 samples from frontal cortex were obtained from the Genotype-Tissue Expression project (GENCODE v19). To compare MES and PN GBM samples, we used read counts from 50 and 45 samples from TCGA, respectively. For all comparisons, we selected as differentially expressed those genes presenting with a Benjamini-Hochberg (FDR) adjusted p -value < 0.05 and $|\log_2\text{FoldChange}| \geq 1$ after the Wald test. Any batch effects were detected or corrections were applied.

RNA-binding proteins (RBPs) and splicing regulator genes

We used a catalogue of 1542 human genes encoding RBPs¹⁵ to build a list of splicing regulators. We selected all genes associated with splice-related functional annotations in the catalog (388 genes) and identified the differentially spliced ones among them. Then we evaluated whether these RBPs represent known components of the spliceosomal machinery according to a comprehensive list of 404 splicing factor genes.¹⁶ We also sought alternative isoforms for lncRNAs, selected according to GENCODE annotations.

Co-expression of RBP

We first analyzed co-expression patterns among differentially expressed RBPs in MES and PN GSCs using Spearman's rank correlation. To identify groups of positively correlated RBPs ($\rho > 0.8$), we performed hierarchical clustering of the correlation results. To confirm co-expression of the identified groups in a broader context, expression data (FPKM) from MES and PN GBM samples were downloaded from TCGA. Correlation analyses of selected groups were then performed separately for MES and PN GBM samples using Spearman's rank correlation. Correlation plots were built for groups of RBPs exhibiting significant correlation patterns in GBM samples (Spearman's rank correlation p -values < 0.01). Correlation analyses were performed using R packages *corrplot* (github.com/taiyun/corrplot) and *Hmisc* (github.com/harrelfe/Hmisc).

Survival analysis of protein-coding genes

Gene expression data (FPKM) and corresponding clinical data from GBM samples were downloaded from TCGA (<https://portal.gdc.cancer.gov/>). Expression data from TCGA were first filtered based on the list of protein-coding genes which were alternatively spliced and presented expression changes in PN versus MES GSCs. Based on the median value of gene expression levels, relevant survival differences between samples with high or low expression of each gene were determined using the log-rank test (p -value < 0.05). Associations of each gene with patient survival were separately assessed in MES and PN GBM samples, classified according to the Gliovis data portal.⁸⁴ Genes exclusively associated with prognosis of MES or PN GBM (log-rank p -value < 0.05) were further included in a multivariate Cox proportional hazards model with the following covariates: age at diagnosis, gender, CIMP status, IDH1 mutation, MGMT methylation, chromosome 19/20 co-gain, and chromosome 7 gain/chromosome 10 loss. After adjusting for effects of these clinical variables, Kaplan-Meier survival curves were then built for genes that remained significantly associated with patient survival (multivariate Cox proportional-hazards regression p -value < 0.05).

Functional annotation and interaction networks

GO categories and KEGG pathways were considered in the functional annotation, using the human genome as background in the DAVID web tool. Clusters of biological categories with FDR corrected p -values < 0.05 (Fisher's exact test) were considered enriched. Enriched clusters of GO categories were further processed to remove redundancy based on semantic similarities using the REVIGO web tool.⁸⁵ Interaction analyses were performed based on protein-protein interaction data from STRING. Interaction networks of proteins exhibiting at least two interactions were then built using Cytoscape.⁸⁶

Survival analysis of lncRNAs

Data for expression of lncRNAs and corresponding clinical data from all tumor types in the study were downloaded from TCGA (<https://portal.gdc.cancer.gov/>). Expression data from TCGA were first filtered based on the list of differentially expressed lncRNAs previously obtained from our data

(candidate lncRNAs). Expression data for each candidate lncRNA were then individually submitted to a separate survival analysis performed using univariate Cox proportional hazards models⁸⁷ with expression levels of lncRNAs as continuous variables. To adjust for other clinical variables, each lncRNA exhibiting significant correlation with patient survival in the univariate analysis (Cox proportional-hazards regression p -value < 0.05) was then included in a multivariate Cox proportional hazards model with the following covariates: age at diagnosis, gender, CIMP status, IDH1 mutation, MGMT methylation, chromosome 19/20 co-gain, and chromosome 7 gain/chromosome 10 loss. Kaplan-Meier survival curves were then built for lncRNAs that remained significantly associated with patient survival in the multivariate Cox proportional-hazards regression models (p -value < 0.05). Relevant survival differences between samples exhibiting high or low expression of each lncRNA (samples split based on the median value of lncRNA expression levels) were determined using the log-rank test (p -value < 0.05). We first considered all GBM samples in the survival analyses. Next, analyses were performed separately considering MES and PN subgroups. GBM samples were classified into these subgroups based on data from the Gliovis data portal.⁴⁸

Validation of splicing events by qRT-PCR

First, we selected a total of seven exon-skipping events to be tested by qRT-PCR. These candidates were randomly selected (using a random number generator) from a sorted list containing the top 500 exon-skipping events with the highest Δ PSI between GSC MES vs. PN. Next, total RNA was extracted using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Reverse transcription of messenger RNAs was performed using a high-capacity cDNA reverse transcription kit (Applied Biosystems) with random priming. For mRNA analysis, quantitative PCR was performed using the primers listed in Supplementary Table 26 and Power SYBR green PCR master mix (Applied Biosystems). Real-time PCR was performed on a ViiA™ STRING7 Real-Time PCR System (Applied Biosystems). Data were acquired using the ViiA 7 RUO software package (Applied Biosystems) and analyzed using an adapted 2- $\Delta\Delta$ CT method with B2M as an endogenous control.

Statistical analysis and figures

Statistical analyses were performed using R. Figures were built using R, Cytoscape,⁸⁶ Circos Plot,⁸⁸ Inkscape (<https://inkscape.org/>) and rstats2shimplot (<https://github.com/Xinglab/rstats2shimplot>).

Reporting summary

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

DATA AVAILABILITY

RNA sequencing data have been deposited in the European Nucleotide Archive [ENA: PRJEB27943].

Received: 22 June 2019; Accepted: 18 November 2019;

Published online: 16 January 2020

REFERENCES

- Paolillo, M., Boselli, C. & Schinelli, S. Glioblastoma under siege: an overview of current therapeutic strategies. *Brain Sci.* **8**, 1–13 (2018).
- Singh, S. K. et al. Identification of human brain tumour initiating cells. *Nature* **432**, 396–401 (2004).
- Safa, A. R., Saadatzaheh, M. R., Cohen-Gadol, A. A., Pollok, K. E. & Bijangi-Visharsaraei, K. Glioblastoma stem cells (GSCs) epigenetic plasticity and interconversion between differentiated non-GSCs and GSCs. *Genes Dis.* **2**, 152–163 (2015).
- Garnier, D., Renoult, O., Alves-Guerra, M.-C., Paris, F. & Pecqueur, C. Glioblastoma stem-like cells, metabolic strategy to kill a challenging target. *Front. Oncol.* **9**, 118 (2019).
- Bhat, K. P. L. et al. Mesenchymal differentiation mediated by NF- κ B promotes radiation resistance in glioblastoma. *Cancer Cell* **24**, 331–346 (2013).
- Mao, P. et al. Mesenchymal glioma stem cells are maintained by activated glycolytic metabolism involving aldehyde dehydrogenase 1A3. *Proc. Natl Acad. Sci. USA* **110**, 8644–8649 (2013).
- Kupp, R. et al. Lineage-restricted OLIG2-RTK signaling governs the molecular subtype of glioma stem-like cells. *Cell Rep.* **16**, 2838–2845 (2016).
- Pangeni, R. P. et al. Genome-wide methylomic and transcriptomic analyses identify subtype-specific epigenetic signatures commonly dysregulated in glioma stem cells and glioblastoma. *Epigenetics* **13**, 432–448 (2018).
- Sun, B. O., Fang, Y., Li, Z., Chen, Z. & Xiang, J. Role of cellular cytoskeleton in epithelial-mesenchymal transition process during cancer progression. *Biomed. Rep.* **3**, 603–610 (2015).
- Tamborero, D. et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* **3**, 2650 (2013).
- Lawrence, M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
- Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
- Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
- Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
- Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
- Seiler, M. et al. Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. *Cell Rep.* **23**, 282–296.e4 (2018).
- Jayasinghe, R. G. et al. Systematic analysis of splice-site-creating mutations in cancer. *Cell Rep.* **23**, 270–281.e3 (2018).
- Saltzman, A. L., Pan, Q. & Blencowe, B. J. Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev.* **25**, 373–384 (2011).
- Huarte, M. The emerging role of lncRNAs in cancer. *Nat. Med.* **21**, 1253–1261 (2015).
- Romero-Barrios, N., Legascue, M. F., Benhamed, M., Ariel, F. & Crespi, M. Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.* **46**, 2169–2184 (2018).
- Pickard, M. R. & Williams, G. T. Molecular and cellular mechanisms of action of tumour suppressor GAS5 lncRNA. *Genes* **6**, 484–499 (2015).
- Xu, B. et al. SNORD47, a box C/D snoRNA, suppresses tumorigenesis in glioblastoma. *Oncotarget* **8**, 43953–43966 (2017).
- Marcelino Meliso, F., Hubert, C. G., Favoretto Galante, P. A. & Penalva, L. O. RNA processing as an alternative route to attack glioblastoma. *Hum. Genet.* **136**, 1129–1141 (2017).
- Bonnal, S., Vigevani, L. & Valcárcel, J. The spliceosome as a target of novel anti-tumour drugs. *Nat. Rev. Drug Discov.* **11**, 847–859 (2012).
- Effenberger, K. A., Urabe, V. K. & Jurica, M. S. Modulating splicing with small molecular inhibitors of the spliceosome. *Wiley Interdiscip. Rev. RNA* **8**, 1–26 (2017).
- Bao, S. et al. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature* **444**, 756–760 (2006).
- Segerman, A. et al. Clonal variation in drug and radiation response among glioma-initiating cells is linked to proneural-mesenchymal transition. *Cell Rep.* **17**, 2994–3009 (2016).
- Pencheva, N. et al. Identification of a druggable pathway controlling glioblastoma invasiveness. *Cell Rep.* **20**, 48–60 (2017).
- Annovazzi, L., Mellai, M. & Schiffer, D. Chemotherapeutic drugs: DNA damage and repair in glioblastoma. *Cancers* **9**, 1–17 (2017).
- Gobin, M. et al. A DNA repair and cell-cycle gene expression signature in primary and recurrent glioblastoma: prognostic value and clinical implications. *Cancer Res.* **79**, 1226–1238 (2019).
- Boccard, S. G. et al. Inhibition of DNA-repair genes Ercc1 and Mgmt enhances temozolomide efficacy in gliomas treatment: a pre-clinical study. *Oncotarget* **6**, 29456–29468 (2015).
- Liu, Z.-G. et al. Relationship between methylation status of ERCC1 promoter and radiosensitivity in glioma cell lines. *Cell Biol. Int.* **33**, 1111–1117 (2009).
- Patil, A. A. et al. FANCD2 re-expression is associated with glioma grade and chemical inhibition of the Fanconi Anaemia pathway sensitises gliomas to chemotherapeutic agents. *Oncotarget* **5**, 6414–6424 (2014).
- Chen, M. S., Higashikubo, R., Laszlo, A. & Roti Roti, J. Multiple alternative splicing forms of human RAD17 and their differential response to ionizing radiation. *Gene* **277**, 145–152 (2001).
- Auffinger, B., Spencer, D., Pytel, P., Ahmed, A. U. & Lesniak, M. S. The role of glioma stem cells in chemotherapy resistance and glioblastoma multiforme recurrence. *Expert Rev. Neurother.* **15**, 741–752 (2015).
- Li, M. et al. CDK4/6 inhibition is more active against the glioblastoma proneural subtype. *Oncotarget* **8**, 55319–55331 (2017).
- Tachon, G. et al. Cell cycle changes after glioblastoma stem cell irradiation: the major role of RAD51. *Int. J. Mol. Sci.* **19**, 1–14 (2018).
- Shen, A. et al. Expression of CRM1 in human gliomas and its significance in p27 expression and clinical prognosis. *Neurosurgery* **65**, 153–159 (2009). discussion 159–60.

39. Amani, V. et al. Polo-like Kinase 1 as a potential therapeutic target in Diffuse Intrinsic Pontine Glioma. *BMC Cancer* **16**, 647 (2016).
40. Koncar, R. F. et al. PLK1 inhibition enhances temozolomide efficacy in IDH1 mutant gliomas. *Oncotarget* **8**, 15827–15837 (2017).
41. Barton, V. N. et al. Aurora kinase A as a rational target for therapy in glioblastoma. *J. Neurosurg. Pediatr.* **6**, 98–105 (2010).
42. Xia, Z. et al. AURKA governs self-renewal capacity in glioma-initiating cells via stabilization/activation of β -catenin/Wnt signaling. *Mol. Cancer Res.* **11**, 1101–1111 (2013).
43. Hanissian, S. H. et al. Regulation of myeloid leukemia factor-1 interacting protein (MLF1IP) expression in glioblastoma. *Brain Res.* **1047**, 56–64 (2005).
44. Álvarez-Satta, M. & Matheu, A. Primary cilium and glioblastoma. *Ther. Adv. Med. Oncol.* **10**, 1758835918801169 (2018).
45. Hoang-Minh, L. B. et al. PCM1 depletion inhibits glioblastoma cell ciliogenesis and increases cell death and sensitivity to temozolomide. *Transl. Oncol.* **9**, 392–402 (2016).
46. Sarkisian, M. R. & Semple-Rowland, S. L. Emerging roles of primary cilia in glioma. *Front. Cell. Neurosci.* **13**, 55 (2019).
47. Moser, J. J., Fritzier, M. J. & Rattner, J. B. Primary ciliogenesis defects are associated with human astrocytoma/glioblastoma cells. *BMC Cancer* **9**, 448 (2009).
48. Lefave, C. V. et al. Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. *EMBO J.* **30**, 4084–4097 (2011).
49. Luan, W. et al. PKM2 promotes glucose metabolism and cell growth in gliomas through a mechanism involving a let-7a/c-Myc/hnRNPA1 feedback loop. *Oncotarget* **6**, 13006–13018 (2015).
50. Park, Y. M. et al. Heterogeneous nuclear ribonucleoprotein C1/C2 controls the metastatic potential of glioblastoma by regulating PDCD4. *Mol. Cell. Biol.* **32**, 4237–4244 (2012).
51. Ziv-Av, A. et al. RTVP-1 regulates glioma cell migration and invasion via interaction with N-WASP and hnRNPK. *Oncotarget* **6**, 19826–19840 (2015).
52. Yang, S. H. et al. Metformin treatment reduces temozolomide resistance of glioblastoma cells. *Oncotarget* **7**, 78787–78803 (2016).
53. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
54. Moeckel, S. et al. Response-predictive gene expression profiling of glioma progenitor cells in vitro. *PLoS ONE* **9**, e108632 (2014).
55. Hubert, C. G. et al. Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. *Genes Dev.* **27**, 1032–1045 (2013).
56. Becerra, S., Montes, M., Hernández-Munain, C. & Suñé, C. Prp40 pre-mRNA processing factor 40 homolog B (PRPF40B) associates with SF1 and U2AF65 and modulates alternative pre-mRNA splicing in vivo. *RNA* **21**, 438–457 (2015).
57. Shingu, T. et al. Kqi deficiency maintains stemness of glioma stem cells in sub-optimal environment by downregulating endolysosomal degradation. *Nat. Genet.* **49**, 75–86 (2017).
58. Danan-Gotthold, M. et al. Identification of recurrent regulated alternative splicing events across human solid tumors. *Nucleic Acids Res.* **43**, 5130–5144 (2015).
59. Yang, Y. et al. Determination of a comprehensive alternative splicing regulatory network and combinatorial regulation by key factors during the epithelial-to-mesenchymal transition. *Mol. Cell. Biol.* **36**, 1704–1719 (2016).
60. Kim, S. et al. Integrative profiling of alternative splicing induced by U2AF1 S34F mutation in lung adenocarcinoma reveals a mechanistic link to mitotic stress. *Mol. Cells* **41**, 733–741 (2018).
61. Correa, B. R. et al. Functional genomics analyses of RNA-binding proteins reveal the splicing regulator SNRPB as an oncogenic candidate in glioblastoma. *Genome Biol.* **17**, 125 (2016).
62. Han, Y. et al. Long non-coding RNA MYOSLID functions as a competing endogenous RNA to regulate MCL-1 expression by sponging miR-29c-3p in gastric cancer. *Cell Prolif.* e12678 (2019).
63. Xiong, H.-G. et al. Long noncoding RNA MYOSLID promotes invasion and metastasis by modulating the partial epithelial-mesenchymal transition program in head and neck squamous cell carcinoma. *J. Exp. Clin. Cancer Res.* **38**, 278 (2019).
64. Wang, R. et al. Hypomethylation of the lncRNA SOX21-AS1 has clinical prognostic value in cervical cancer. *Life Sci.* **233**, 116708 (2019).
65. Zhang, X., Zhao, X., Li, Y., Zhou, Y. & Zhang, Z. Long noncoding RNA SOX21-AS1 promotes cervical cancer progression by competitively sponging miR-7/VDAC1. *J. Cell. Physiol.* **234**, 17494–17504 (2019).
66. Zhang, J., Hou, T., Qi, X., Wang, J. & Sun, X. SOX21-AS1 is associated with clinical stage and regulates cell proliferation in nephroblastoma. *Biosci. Rep.* **39**, 1–7 (2019).
67. Wei, C., Wang, H., Xu, F., Liu, Z. & Jiang, R. LncRNA SOX21-AS1 is associated with progression of hepatocellular carcinoma and predicts prognosis through epigenetically silencing p21. *Biomed. Pharmacother.* **104**, 137–144 (2018).
68. Lu, X. et al. A novel long non-coding RNA, SOX21-AS1, indicates a poor prognosis and promotes lung adenocarcinoma proliferation. *Cell. Physiol. Biochem.* **42**, 1857–1869 (2017).
69. Yang, C.-M. et al. Aberrant DNA hypermethylation-silenced SOX21-AS1 gene expression and its clinical importance in oral cancer. *Clin. Epigenetics* **8**, 129 (2016).
70. Paul, Y. et al. Genetic landscape of long noncoding RNA (lncRNAs) in glioblastoma: identification of complex lncRNA regulatory networks and clinically relevant lncRNAs in glioblastoma. *Oncotarget* **9**, 29548–29564 (2018).
71. Zhang, J., Yin, M., Peng, G. & Zhao, Y. CRNDE: an important oncogenic long non-coding RNA in human cancers. *Cell Prolif.* **51**, e12440 (2018).
72. Hongzhen, Z. et al. The diagnostic and prognostic significance of long non-coding RNA CRNDE in pan-cancer based on TCGA, GEO and comprehensive meta-analysis. *Pathol. Res. Pract.* **215**, 256–264 (2019).
73. Kiang, K. M.-Y., Zhang, X.-Q. & Leung, G. K.-K. Long non-coding RNAs: the key players in glioma pathogenesis. *Cancers* **7**, 1406–1424 (2015).
74. Kiang, K. M.-Y. et al. CRNDE expression positively correlates with EGFR activation and modulates glioma cell growth. *Target. Oncol.* **12**, 353–363 (2017).
75. Li, H. et al. LncRNA CRNDE triggers inflammation through the TLR3-NF- κ B-Cytokine signaling pathway. *Tumour Biol.* **39**, 1010428317703821 (2017).
76. Wang, Y. et al. CRNDE, a long-noncoding RNA, promotes glioma cell growth and invasion through mTOR signaling. *Cancer Lett.* **367**, 122–128 (2015).
77. Ellis, B. C., Molloy, P. L. & Graham, L. D. CRNDE: a long non-coding RNA involved in Cancer, Neurobiology, and Development. *Front. Genet.* **3**, 270 (2012).
78. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
79. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
80. Shen, S. et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl Acad. Sci. USA* **111**, E5593–601 (2014).
81. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
82. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
83. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
84. Bowman, R. L., Wang, Q., Carro, A., Verhaak, R. G. W. & Squatrito, M. GlioVis data portal for visualization and analysis of brain tumor expression datasets. *Neuro. Oncol.* **19**, 139–141 (2017).
85. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).
86. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
87. Cox, D. R. Regression models and life-tables. *J. R. Stat. Soc. Series B Stat. Methodol.* (1972).
88. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
89. Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).

ACKNOWLEDGEMENTS

We thank all members of Galante laboratory for helpful discussions. Daniel T. Ohara for technical assistance and Dr. Ichiro Nakano for sharing GSC lines. This study was supported by a grant from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil to PAFG and LOFP and by NIH 7R21CA175875-03. This study was partially supported by grants from Serrapilheira foundation and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; 2018/15579-8) to PAFG. GDAG and BRC were supported by fellowships from FAPESP (2017/19541-2) and (2013/25483-4 and 2013/07159-5), respectively. PRA was supported by CPRIT Training Grant - RP140105.

AUTHOR CONTRIBUTIONS

Study design and manuscript writing were done by G.D.A.G., B.R.C., L.O.P. and P.A.F.G. Computational data analyses were done by G.D.A.G. and B.R.C. Cell line cultures, RNA extraction, candidate validations, and preparation of material for sequencing were done by P.R.A., M.Q. and S.B., L.O.F.P. and P.A.F.G. led the project.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41525-019-0108-5>.

Correspondence and requests for materials should be addressed to L.O.F.P. or P.A.F.G.

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

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



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

© The Author(s) 2020