Title: Integrative genomics approach to primary ciliary gene regulation in glioma

Name: Sharlyn Mae Buendia Chua

Affiliation: Genome Evolution Laboratory, Department of Genetics, School of Life Science

Introduction:

Diffuse gliomas are classified as low-grade gliomas (LGG) and high-grade gliomas, also called glioblastomas (GBM). Diffuse gliomas share genetic mutations in *isocitrate dehydrogenases 1* and *2* (IDH^{mut}), which are among the earliest and most frequent molecular events in gliomagenesis. *Isocitrate dehydrogenases* (*IDH*) convert isocitrate into to alpha-ketoglutarate (α-KG) in the tricarboxylic acid (TCA) cycle; however, the mutation results in the oncometabolite 2-hydroxyglutarate (2-HG). IDH^{mut} is observed in approximately 80% of LGG cases and in approximately 10% of GBM cases originating from the LGG. Approximately 90% of cases are IDH wild-type (IDH^{wt}), corresponding to primary GBM, which develops de novo with no identifiable precursor tumor. GBM is the most aggressive and fatal form of brain cancer.

GBM cell lines, tumor biopsies, and patient-derived glioma stem cells (GSCs) show abnormal or absent primary cilium, or a reduced number of ciliated cells. The primary cilium functions as a prime sensor for extracellular signals and coordinates signaling pathways. It has been previously observed that GSCs with suppressed ciliogenesis do not undergo cilium checkpoint, consequently, evading cell cycle exit, maintaining their stemness, hindering differentiation, and proliferation. Conversely, the recovery of the primary cilium leads to differentiation and prevents invasion, demonstrating that ciliotherapy, a medical intervention that focuses on restoring ciliary function and changing its length, is a promising therapeutic option. In particular, patients with LGG have a seven-year average survival, while those with GBM have only 14.6 months median survival and only 26.5% survive for two years after

diagnosis, even after treatment. Nonetheless, there is limited knowledge regarding the molecular features of the primary cilium in gliomas, particularly regarding ciliary gene regulation.

This study aimed to uncover the regulation of ciliary genes in gliomas using publicly available databases to identify correlations between genomics, epigenomics, and transcriptomics. This study showed that Regulatory Factor X 1–3 (RFX) functions as a ciliary gene regulator in glioma. RFX motifs are frequently present in active regulatory regions in gliomas, and *RFX1-3* genes are differentially expressed in gliomas compared to normal brain samples.

Method:

In this study, I used LGG and GBM data from the Genomic Data Commons (GDC) repository. I analyzed the accessible chromatin regions and motifs enriched in these regions using ATAC-seq of LGG and GBM samples from The Cancer Genome Atlas (TCGA) project of the GDC portal. The peaks calling was done using Genrich and the genes containing motifs of RFX1, RFX2, and RFX3 in their regulatory regions were identified using motif enrichment analysis of HOMER. I identified the differentially expressed genes in glioma using the HTSeq-count from both primary and recurrent glioma tumors from TCGA, while normal brain tissues were from both TCGA and CPTAC (Clinical Proteomic Tumor Analysis Consortium), and were normalized and analyzed using DESeq2. I did survival analysis of data coming from "RTCGA.clinical" specifying GBM.clinical and LGG.clinical datasets. I categorized LGG and GBM samples into three groups based on their *RFX* expression levels and used the R package survival for Kaplan-Meier analysis to assess survival differences using the log-rank test and Cox proportional hazard to assess effects of multiple covariates including *RFX1* expression and *IDH* mutation. I also assessed the *RFX1-3* expression based on the *IDH* mutation status.

The somatic mutation data was from the UCSC Xena Platform of TCGA-LGG and TCGA-GBM projects, and Wilcoxon test was used to compare the significant difference in expression. I analyzed the relationship between the methylation of CpG sites in *RFX1-3*, expression, and *IDH* mutation status. The beta values were obtained from the UCSC Xena browser GDC TCGA-LGG and GDC TCGA-GBM datasets and were examined using the Wilcoxon test. For the functional enrichment of differentially expressed genes, gene ontology analysis was performed on genes having RFX1, RFX2, RFX3, and X-box motifs and were DEGs. Lastly, I identified ciliary genes whose expression was correlated with *RFX1*, *RFX2*, and *RFX3* using Spearman's correlation.

Results and Discussion:

There is limited understanding currently of how ciliary genes are controlled in diffuse gliomas. My study revealed the significant role played by RFX1-3 transcription factors in governing these genes in diffuse gliomas. The genes under their regulation could potentially serve as targets for restoring the primary cilium, a promising avenue since re-establishing the primary cilium in glioma stem cells (GSC) results to differentiation and hinders invasion.

RFX X-box motifs are highly enriched in the accessible chromatin regions of glioma samples

To understand the mechanisms governing primary ciliary genes in diffuse gliomas, I utilized ATAC-seq data, which offers comprehensive insights on the regulation of genome-wide active genes. Examining the highly enriched motifs in the accessible chromatin regions in LGG and GBM showed that RFX1-3 and X-box motifs were enriched. In these motifs, the Regulatory Factor X (RFX) family of transcription factors can bind. This indicated a significant role of the RFX family in regulating active genes in glioma samples. Moreover, among the

major ciliary transcription factors, only the RFX family and X-box motifs were highly ranked as enriched motifs in diffuse gliomas

RFX genes are differentially expressed in diffuse glioma and

We identified *RFX1* and *RFX2* were upregulated, while *RFX3* was downregulated in LGG and GBM samples compared to normal brain samples. I analyzed the survival of the patients in relation to *RFX1-3* expression and found that LGG patients with high *RFX1* had poorer prognosis after the adjustment for *IDH* mutation, age, and sex. This suggests that RFX1 expression has the potential to serve as a prognostic marker for LGG.

Interplay between RFX1-3 expression levels, IDH^{mut}, and methylation patterns

I analyzed the link between *RFX1-3* differential gene expression and *IDH* mutation in gliomas and found that there was an increased expression of *RFX1* and *RFX2* in patients with IDH^{wt} gliomas and decreased expression in patients with IDH^{mut} gliomas. Furthermore, these observations were found to be associated with specific CpG sites that were hypermethylated in the 7th intron and promoter of *RFX1* and the 1st intron of *RFX2* in IDH^{mut} patients. This aligns with previous research indicating that IDH^{mut} glioma patients typically display a hypermethylated phenotype, which generally results in inhibition of gene expression. This part of my study presents valuable insights into the epigenetic regulation of *RFX1* and *RFX2* expression in the context of IDH^{mut} glioma.

Functional analysis of RFX transcription factors

I analyzed the functional roles of *RFX* in glioma, as its previously known functions were mainly associated with the *Cluster of Differentiation 44* (*CD44*) and *Fibroblast Growth Factor 1* (*FGF1*) and no study showing its role in glioma primary cilium.

The downstream genes of RFX were analyzed to explore its regulatory roles in gliomas. Previous reports were limited to its regulation of Cluster of Differentiation 44 (CD44) and Fibroblast Growth Factor 1 (FGF1). My gene ontology analysis of genes with RFX1-3 and X-box motifs in their promoter regions that were differentially expressed in glioma compared to normal brain samples showed that the biological processes of these genes are in the assembly and formation of cilium and that they encode for genes that are part of the cilium and its related structures. To our knowledge, this study is the first evidence of RFX1-3 transcription factor regulating ciliary genes in gliomas, a finding that aligns with their known roles in various organisms and cell types. Furthermore, specific ciliary genes possibly regulated by RFX1-3 were identified by their correlation with *RFX1-3* expression in this study.

Conclusion:

This study suggests that RFX1, RFX2, and RFX3 TFs function as key regulators of ciliary genes, whose roles are essential in the primary cilium of LGG and GBM. I found the RFX X-box motifs were enriched in the active chromatin regions of patients with diffuse gliomas. *RFX1*, *RFX2*, and *RFX3* showed differential expression in patient tissues of diffuse gliomas compared to normal brain samples, and *RFX1* expression can serve as a prognostic biomarker and evaluate the risk for patients with LGG. In addition, hypermethylated CpG sites in the regulatory regions of *RFX1* and *RFX2* of IDH^{mut} gliomas and the expression level differences depending on *the IDH* mutation status. Gene ontology analysis revealed that the downstream genes of RFX1-3 were associated with primary cilium, and I further identified the ciliary genes possibly regulated by RFX1-3 TFs.

Ciliotherapy is an emerging field that seeks to manipulate the primary cilium and rescue normal signaling programs in diseases with abnormal cilium formation. This approach is particularly significant in glioma, and a recent study on GSCs demonstrated that reintroducing

the cilium leads to a loss of stemness, induction of differentiation, and a non-invasive phenotype. Our study suggests that the abundance of X-box motifs indicates that irregular expression of *RFX1-3* may affect numerous regulated ciliary genes, rendering them promising targets for inducing cilium formation.