bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mapping chromatin accessibility and active regulatory elements reveals new pathological mechanisms in human gliomas

Authors/Affiliations

Karolina Stępniak[1]*, Magdalena A. Machnicka[2]*, Jakub Mieczkowski[1]*, Anna Macioszek[2], Bartosz Wojtaś[1], Bartłomiej Gielniewski[1], Sylwia K. Król[1], Rafał Guzik[1], Michał J. Dąbrowski[3], Michał Dramiński[3], Marta Jardanowska[3], Ilona Grabowicz[3], Agata Dziedzic[3], Hanna Kranas[2a], Karolina Sienkiewicz[2], Klev Diamanti[4], Katarzyna Kotulska[5], Wiesława Grajkowska[5], Marcin Roszkowski[5], Tomasz Czernicki[6], Andrzej Marchel[6], Jan Komorowski[4], Bozena Kaminska[1]#, Bartek Wilczyński[2]#

[1] Nencki Institute of Experimental Biology of the Polish Academy of Sciences; Warsaw, 02-093, Poland
[2] Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Warsaw, 02-097, Poland
[3] Institute of Computer Science of the Polish Academy of Sciences; Warsaw, 01-248 Poland
[4] Department of Cell and Molecular Biology, Uppsala University, Uppsala, 75124, Sweden
[5] Deptartments of Neurology, Neurosurgery, Neuropathology, The Children's Memorial Health Institute, Warsaw, 02-093, Poland
[6] Department of Neurosurgery, Medical University of Warsaw, Warsaw, 02-097, Poland.

**Equally contributing authors*
 # Corresponding authors

Contact Info: b.wilczynski@mimuw.edu.pl; b.kaminska@nencki.gov.pl

Summary

Chromatin structure and accessibility, and combinatorial binding of transcription factors to regulatory elements in genomic DNA control transcription. Genetic variations in genes encoding histones, epigenetics-related enzymes or modifiers affect chromatin structure/dynamics and result in alterations in gene expression contributing to cancer development or progression. Gliomas are brain tumors frequently associated with epigenetics-related gene deregulation. We performed whole-genome mapping of chromatin accessibility, histone modifications, DNA methylation patterns and transcriptome analysis simultaneously in multiple tumor samples to unravel novel epigenetic dysfunctions driving gliomagenesis. Based on the results of the integrative analysis of the acquired profiles, we created an **atlas of active enhancers and promoters in benign and malignant gliomas.** We explored these elements and intersected with Hi-C data to uncover molecular

a Current address: Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac, 10, 08028 Barcelona, Spain

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mechanisms instructing gene expression in gliomas.

Keywords

Gliomas; chromatin accessibility; histone modifications; ATAC-sequencing; RNA sequencing; topological domains; gliomagenesis

Significance

Epigenetics-driven deregulation of gene expression accompanies cancer development, but its comprehensive characterization in cancer patients is fragmentary. We performed whole-genome profiling of **gene expression, open chromatin, histone modifications and DNA-methylation profiles in the same samples from benign and malignant gliomas.** Our study provides a first comprehensive atlas of active regulatory elements in gliomas, which allowed identification of the functional enhancers and promoters in patient samples. This comprehensive approach revealed epigenetic patterns influencing gene expression in benign gliomas and a new pathogenic mechanism involving FOXM1-driven network in glioblastomas. This atlas **provides a common set of elements for cross-comparisons** of existing and new datasets, prompting novel discoveries and better understanding of gliomagenesis.

Highlights

- We provide an atlas of *cis*-regulatory elements active in human gliomas
- Enhancer-promoter contacts operating in gliomas are revealed
- Diverse enhancer activation is pronounced in malignant gliomas
- Chromatin loop activates FOXM1-ANXA2R pathological network in glioblastomas.

INTRODUCTION

Gliomas are most common primary tumors of the central nervous system (CNS). According to a World Health Organization (WHO) classification, gliomas are divided based on molecular dysfunctions and histopathology into WHO grade I gliomas (benign tumors with infrequent genomic alterations and longer-term survival), grades II and III diffuse gliomas (characterized by the presence or absence of mutations in *IDH, TERT, ATRX* and copy number alterations of 1p, 19q, 7, and 10q) and grade IV glioblastoma, the most malignant tumor with numerous genomic alterations (Ceccarelli et al., 2016; Louis et al., 2016). Glioma development is a multistep process, involving genetic and epigenetic alterations (Reifenberger et al., 2016). Comprehensive sequencing studies have revealed numerous aberrations in epigenetics- and chromatin remodeling-related genes in high grade gliomas. Due to aberrant activation or inactivation of epigenetics-related enzymes the epigenetic landscape is highly dysregulated leading to aberrant gene expression (Fontebasso et al., 2013; Maleszewska and Kaminska, 2013).

Chromatin acts as a dynamic platform of signal integration and a long-term maintenance of gene expression profiles. Modifications of histones, particularly their N-terminal tails,

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

influence chromatin accessibility and gene expression, and can be activating or repressive (Kouzarides, 2007). Acetylation of N-terminal lysine residues of histones H3 and H4, trimethylation of H3 lysine 4 (H3K4me3) and lysine 79 (H3K79me), and trimethylation of H3 lysine 36 (H3K36m3), are typically associated with an active chromatin, while methylation of lysines 9 (H3K9me3) and 27 (H3K27me3) of histone H3 are hallmarks of a condensed chromatin at silent loci (Tessarz and Kouzarides, 2014). H3 acetylation and H3K4me1 marks outside of promoter regions have been correlated with functional enhancers in different cell types (Bogdanovic et al., 2012; Bonn et al., 2012). The deceptively opposite modifications, H3K4me3 and H3K27me3 co-localize in regions termed “bivalent domains”, mostly in embryonic stem cells (Bernstein et al., 2006). Extensive characterization of histone methylations has been completed at a limited genome coverage and resolution in selected human cells (Bernstein et al., 2007). There are only a few reports regarding profiles of histone modifications or chromatin accessibility in freshly resected human specimens. This includes high-resolution Hi-C maps of chromatin contacts in human embryonic brain (Won et al., 2016), ATAC-seq on postmortem brain (Fullard et al., 2017) or glioma stem like cells from glioblastoma (Tome-Garcia et al., 2018), ChIP-seq analysis for H3K27me3 modification on 11 glioblastoma and 2 diffuse astrocytoma samples (Sharma et al., 2017). Only recently researchers began integrating H3K27ac profiles with gene expression, DNA methylomes, copy number variations, and whole exomes demonstrating distinct chromatin and epigenetic profiles in glioblastomas, distinguishing this entity from other brain tumors (Mack et al., 2019). Combination of those features in a single patient and intersections of acquired profiles in gliomas of different grades have not yet been achieved.

We performed a comprehensive analysis of whole genome profiles of open chromatin, histone modifications, DNA methylation and gene expression in a given patient using freshly resected gliomas of various malignancy grades. Intersecting the acquired data led to more precise mapping of active promoters and enhancers in gliomas, some reflecting brain-specific sites. We intersected our data with Hi-C maps from a human embryonic brain to identify topologically associated domains and we searched for co-regulated genes, regulatory binding motifs and potential master regulators. This approach revealed a new regulatory connection between FOXM1 and ANXA2R implicated in gliomagenesis. We provide an atlas of brain specific enhancers and promoters which could be a valuable resource for further exploration. The data are available at http://regulomics.mimuw.edu.pl/GliomaAtlas/ in raw and processed forms ready to visualize in genome browsers.

Results

Cohort description and experimental workflow

We focused on groups of patients representing major malignancy groups: pilocytic astrocytomas (WHO grade I, PA, n=11), diffuse astrocytomas (WHO grade II and III, DA, n=7) and glioblastomas (WHO grade IV, GBM, n=15). Clinical data about each patient have

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

been collected and are presented in Table S1. Freshly resected glioma tissues were immediately homogenized to a single cell suspension and experimentally defined numbers of cells were split into portions that were directly processed for ATAC-seq, ChIP-seq, RNA-seq or DNA methylation experiments (Figure 1). Tissue homogenization was performed to avoid confounding factors related to tumor heterogeneity. We analyzed the profiles of DNA methylation, open chromatin and histone modifications characteristic for active (H3K4me3, H3K27ac) and repressed chromatin (H3K27me3) along with RNA-sequencing on the individual patients tissues. The IDH1/2 mutational status has been evaluated in all samples by Sanger sequencing of PCR amplicons (Table S1). All experiments and sequencing were carried out in the same laboratory, therefore, we could avoid possible sequencing artifacts. All datasets were subjected to rigorous quality control according to Encyclopedia of DNA elements (ENCODE) best practices, and only datasets passing stringent quality control metrics are presented.

Creating the atlas of regulatory sites in benign and malignant gliomas

For each individual, we performed transposase-accessible chromatin with sequencing (ATAC-seq) and chromatin immunoprecipitation followed by sequencing (ChIP-seq) to detect histone modification marks characteristic for active and poised enhancers, and active promoters. Figure 2A summarizes numbers of samples for which ATAC-seq and ChIP-seq were collected excluding the samples that did not meet our stringent quality criteria. For a substantial number of glioma samples we collected complete datasets, except for a lack of H3K27me3 profiles from PAs, due to a low size of the resected tumors. Representative ATAC-seq, ChIP-seq and RNA-seq peaks for the *NRXN2* gene in two samples, one selected from pilocytic astrocytomas (PA04) and one from glioblastomas (GB01) are shown in Figure 2B. The peak enrichments are comparable across samples with expected overrepresentation of peaks in promoter regions. Numbers of detected peaks are presented in Figure 2C.

By performing ChIP-seq experiments for two active and one repressive histone marks (H3K4me3, H3K27ac and H3K27me3), ATAC-seq and RNA-seq we have collected a comprehensive dataset describing chromatin state and gene activity across different glioma grades. We identified regulatory regions whose activity is shared between samples (“common”) or specific to a small subset of samples (“variable”) for both H3K4me3 and H3K27ac ChIP-seq experiments (Figure 2D). We identify many more common than variable H3K4me3 peaks, suggesting those commonly active promoters represent genes expressed in the brain, while the variable ones may be more related to gliomagenesis.

Genome-wide chromatin conformation capture (Hi-C) allows identification of interactions between DNA regions including promoter-enhancer interactions. To better describe enhancers, we used Hi-C data from high-resolution 3D maps of chromatin contacts from developing human brain (Won et al., 2016) for identification of enhancer-promoter contacts that might function in gliomas. First, using Hi-C data we searched for contacts of all putative active enhancers identified in our samples with any region within 2Mbp range

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Figure 2E). We found significant (p<0.001) contacts for 21,141 putative enhancers out of 45,548 (including 5,530 common enhancers out of 10,673). Target genomic intervals which were predicted to be in contact with common putative enhancers are enriched in promoter regions, supporting biological relevance of these predictions (Suppl. Figure 2A).

Identified, non-promoter regulatory regions show high diversity between patients

Active promoters, defined by the presence of either H3K4me3 or H3K27ac, are highly consistent between patients. Similarly, gene expression profiles show significant overlap between patients (Figure 3A, Figure S3). Expectedly, sets of promoters identified as active based on the presence of either one of the active histone marks or open chromatin peaks, are highly similar; 8007 common active promoters display ATAC-seq signals, H3K4me3 and H3K27ac marks (Figure 3B). Active promoters exhibit also evolutionary conservation in the context of 100 vertebrate genomes, compared to random genomic intervals (Figure 3C). In contrast to promoter regions, putative enhancers show high variability between patients (Figure 3A, Figure S3). Their sequence conservation is also weaker than sequence conservation of the promoters, but still very significant when compared to random genomic intervals (Figure 3C). By comparing these common and variable promoter regions to the published studies of promoters active in normal brain (Bernstein et al. 2012) and in non-brain cell lines, we can estimate that approximately 50% of common promoters are also active in normal brain and 30% are also active in other tissues (Figure S3B). The values for variable elements are considerably lower, with approximately 90% of them not appearing in brain or cell-type derived datasets, making them likely to be specific to gliomas.

Having paired ChIP-seq and RNA-seq data from the same samples we investigated whether the expected correlation between the level of active histone marks and gene expression levels (Karlic et al., 2010) is visible in our data. Both H3K4me3 and H3K27ac ChIP-seq coverages at promoter regions correlate with the expression of the associated transcripts (Spearman R = 0.62 and 0.66, respectively; Figure 3D). However, our data demonstrate that dependency between occurrence of H3K4me3 peaks and transcript expression is quite non-linear, with a pronounced difference in coverage between transcripts whose expression is within 1st or 2nd quantile. In contrast, in case of H3K27ac the increase in the transcript expression is associated with the steady increase in the ChIP-seq signal (Figure 3D).

Hierarchical clustering on the mRNA expression profiles of the samples revealed that malignancy groups correspond very well to the obtained clusters (Figure 3E). All but one PA samples formed a distinct cluster showing a stronger within-group similarity than groups related to other malignancies. As expected due to the presence of PA-like glioma subtype within LGr4 RNA cluster (Ceccarelli et al. 2016), majority of the PA samples were classified to LGr4 Pan Glioma RNA Expression Cluster. DA samples also exhibited a high similarity of

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RNA expression profiles. Majority of them were assigned to LGr1/LGr3 RNA clusters. In concordance with Pan Glioma RNA Expression Cluster specificity, all IDH1 mutant samples were assigned to the LGr1 or LGr3 RNA clusters, even though they were split on the dendrogram into DA and GB specific clusters. The GB mRNA expression profiles exhibited the largest variability, although the majority of them were found to be of Mesenchymal glioblastoma subtype assigned to LGr4 RNA cluster.

To further assess biological relevance of enhancer-promoter contacts predicted in glioma samples, we used Hi-C data from a fetal human brain (Won et al., 2016). We computed the Spearman correlation of H3K27ac coverage on enhancers with transcript expression for enhancer-promoter pairs defined either based on Hi-C data or by assigning the closest transcript to each analyzed enhancer. We found that both for closest enhancer-transcript pairs and pairs identified from Hi-C data the correlation between the histone mark level and transcript expression is relatively low (around 0.1 to 0.25) (Figure 3F). Our results suggest that closest enhancer-promoter pairs are expected to be correlated if the distance between them is at most 20 kb. However, in cases when no transcript can be found within 200kbp from the putative enhancer, Hi-C data can provide significantly more information about expected enhancer-promoter interactions than assignment based on the shortest promoter-enhancer distance.

Altogether, the presented results on integrative analysis of ATAC-seq, ChIP-seq and RNA-seq allowed us to identify non-coding, regulatory sites that are active in gliomas and represent a common brain specific sites. We predicted the presence of active promoters and putative enhancers, and predicted enhancer-promoter contacts in our glioma samples by exploiting the published Hi-C data dataset (Won et al., 2016). The data allowed us to create an atlas of non-coding, regulatory sites in those brain tumors and we made the data easily available in raw and processed formats in a genome browser on our website (http://regulomics.mimuw.edu.pl/GliomaAtlas)

H3K27ac at the transcription start sites marks genes related to neuronal signaling in pilocytic astrocytomas

We explored the atlas of active promoters and putative enhancers to identify tumor–specific gene regulation. We compared chromatin activity around transcription start sites (TSS) between benign (PA) and malignant (DA/GB) tumors using H3K4me3 and H3K27ac profiles (Figure 4AB). The analysis was limited to protein coding genes. We found that the abundance of the selected histone marks discriminated the two glioma groups. While H3K4me3 marks at TSS were significantly higher in PA samples (Figure 4A), H3K27ac signals were significantly more abundant in DA/GB samples (Figure 4B, Figure S4A).

Interestingly, besides the global increase of H3K27ac marks in DA/GB, we found a set of regions with higher H3K27ac abundance in PA than in DA/GB samples (Figure 4C, we

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

will refer to them as the selected regions in the following paragraph). The detailed H3K27ac profiles centered at the corresponding TSS confirmed that the observed increase in their abundance was not an artifact resulting from data transformation, and was spread over the selected region (TSS+/-2kb) (Figure 4D). Also, ATAC-seq profiles obtained for four PA and four DA/GB samples showed differences in chromatin accessibility (Figure 4E). To explore those changes in the genome-wide context, we compared the H3K27ac profiles in the selected regions with the remaining regions. We found that in comparison to the other regions, H3K27ac profiles in the selected regions were substantially lower in DA/GB samples (Figure 4F, right panel), while in PA samples, the H3K27ac levels remained unchanged (Figure 4F, left panel). This shows that H3K27ac mark in the selected regions was diminished in samples of the malignant tumors, which may result from the absence of some regulatory modes or cell types in DA/GB tumors.

To further characterize PA-specific H3K27ac profiles in terms of gene functionality, we analyzed expression levels of the genes corresponding to the selected TSS proximal regions. The average expression levels of these genes were significantly higher (P=2.8e-167, Wilcoxon rank-sum test) in PA tumors, which supported previous observations (Figure 4G). Gene Ontology (GO) functional enrichment analysis of the selected genes indicated their participation in synaptic signaling and nervous system development (Figure 4H, Table S2). A search for transcription factor binding site motifs within the selected regions revealed the enrichment of factors involved in neuronal processes, including NFR1 (Figure 4I, see Methods for details). NRF1 is involved in neurite growth and its deletion leads to neuronal dysfunctions (Chang et al., 2005; Piao et al., 2012). Another identified factor, MXI1, was reported as a tumor suppressor negatively regulating c-MYC and an inhibitor of glioma proliferation(Lahoz et al., 1994; Manni et al., 2002). Altogether, these findings indicate higher expression of genes related to neuronal signaling in PA tumors when compared to DA/GB gliomas.

Pilocytic astrocytomas show H3K4me3 hypermethylation at the promoters targeted by PRC2

Hierarchical clustering of samples with H3K4me3 marks in promoters shows a noticeable difference between PA and DA/GB gliomas (Figure 5A). The majority of the PA samples exhibited large pairwise overlap of H3K4me3 marked promoters (0.93-0.95), while the pairwise overlap between PA samples and DA/GB gliomas was lower. Additionally, the group of DA/GB tumors showed relatively lower similarity among them. We identified 359 genes with the promoters differentially marked with H3K4me3 betweenthe five PA samples forming the tight cluster, and DA/GB gliomas. The vast majority of them (338 out of 359) had the active chromatin mark in their promoters only in PA samples (Table S3) and were characterized by a significant increase in the mRNA levels (Figure 5B), H3K27ac and ATAC-seq signal intensity in PA samples in comparison to DA/GB samples (Figure S5A-B). Interestingly, this promoter set is rich in binding sites of the PRC2 complex components:

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

nearly half of them contain the binding sites for EZH2 and nearly one sixth contains the binding sites for SUZ12, based on ChIP-seq data from the ENCODE project (Gerstein et al., 2012). When compared to promoters marked by H3K4me3 in all our samples, the promoters specifically active in PA are enriched in EZH2 and SUZ12 binding sites 7.2 and 2.5 times, respectively (Figure 5C). Additionally, transcription factor motifs enrichment analysis showed that the promoters specifically active in PA are characterized by the presence of GC rich sequences, while no strongly enriched motifs were identified (Figure S5C). This observation is in agreement with an expected lack of strong sequence specificity in the PRC2 complex binding. Moreover, when we compared the motifs that are enriched with respect to other active promoters, we found much fewer motifs significantly enriched in the PA specific promoters than when compared to other inactive promoters. Also, comparison to motifs enriched in DA/GB-specific promoters, supports lack of PA-specificity of the GC-rich motifs (Figure S5D).

We acquired the whole-genome bisulfite sequencing data from the same tumor samples as discussed above. Interestingly, we found DNA methylation levels at the promoters with EZH2 binding sites to be lower in PA compared to DA/GB gliomas (Figure 5D). This observation was corroborated by the data from the independent PA and GB dataset analyzed with HumanMethylation450 BeadChip array (Figure 5E). Overall, these promoters may represent an epigenetic signature which differentiates PA from DA/GB gliomas and indicate that mechanisms related to chromatin repression through the PRC2 complex together with DNA methylation may play leading roles in regulating a set of genes.

Using PANTHER Overrepresentation Test (Mi et al., 2019), we searched for enriched Gene Ontology (GO) terms in a set of the PA-specific genes with EZH2 binding sites in promoters. This analysis did not reveal strong enrichments, except for terms related with developmental processes, cell differentiation, transmembrane transport processes (Figure 5F, S5G). For genes specifically active in PA, but without EZH2 binding site in promoters, no significantly enriched GO terms were identified. We speculate that altered expression of these genes may arise as a secondary effect, since representatives of several diverse transcription factor families (including HES, GATA, NR2E, VSX and SIM1) can be identified among those genes that have EZH2 binding site in their promoters.

Pilocytic astrocytomas are usually located in cerebellum, while higher grade gliomas more often develop in cortical areas (Larjavaara et al., 2007), we set out to verify, whether the observed set of promoters, specifically activated in PA tumors may be connected to the common developmental origin of these tumors. We collected DNA methylation levels of the promoters specifically active in PA and having an EZH2 binding site in previously published samples from different brain regions (Lambert et al., 2013; Pidsley et al., 2014). In agreement with our prediction, these promoters have increased DNA methylation levels in normal cortex compared to cerebellum (Figure S5F), however the increase is not as

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

significant as in our PA samples. Similarly, PA located in infratentorial brain regions are hypermethylated compared to the control cerebellum (Figure S5G). These differences in DNA methylation are accompanied by changed transcript expression levels (Figure S5H) supporting the functional nature of the hypermethylation. These results suggest that the epigenetic mechanism of activation of the genes in question is a combination of their developmental background and further tumor-related dysregulation of the PRC2 complex. At the same time, these genes exhibit even higher methylation levels in GB samples, consistent with expected dysregulation of Polycomb complexes in high grade tumors.

Differential activation of regulatory regions coupled with chromatin contacts uncovers the FOXM1-ANXA2R network operating in glioblastoma and impacting patient survival

Chromatin conformation may influence regulation of gene expression (Lupiáñez et al., 2015, Donaldson-Collier et al., 2019) and topologically associating domains (TADs) segregate the genome into megabase-scale regions. Enhancers and their cognate promoters are typically located in the same TAD, even if their genomic distance between them is as large as a mega-base (Racko et al., 2018). Since we were most interested in promoter-enhancer contacts that require high-resolution maps, we used published high resolution fetal brain Hi-C data (Won et al., 2016) to putative enhancers locations to search for potential interactions that may elicit pathogenic mechanisms. In order to validate that these data are applicable to putative glioma enhancers, we verified that the TAD boundaries in the fetal brain data are not significantly different from those derived from low-resolution glioma Hi-C data (Harewood et al. 2017)

We identified 109 enhancer-promoter pairs with significant contact frequency based on Hi-C data and a high correlation between the enhancer coverage in H3K27ac ChIP-seq data and transcript expression (Spearman R > 0.7; Figure 6A). The identified contacts corresponded to 85 genes among which 17 (FDR < 0.01) were differentially expressed in tumors of different malignancy (Figure 6B). We focused on genes whose expression levels were the highest in GB samples and the lowest in PA samples. We obtained a list of 5 genes, and for all of them we found a significant correlation between promoter and putative enhancer activity levels (Figure 6C). One of the identified genes, *ANXA2R*, encodes a receptor for Annexin A2 (ANXA2), an element of ANXA2-ANXA2R axis known to play a role in cancerogenesis by promoting cell invasion and migration (D'Souza et al., 2012; Shiozawa et al., 2008). Annexin 2 is the most abundant protein in breast cancer-derived exosomes and enhances angiogenesis (Maji et al., 2017).

The analysis of the putative enhancer sequence revealed a presence of several transcription factor binding motifs, including motifs for the FOXM1 transcription factor (Figure S6]. *FOXM1* expression was significantly increased in malignant gliomas: DA and GB (Figure 6D). Comparison of *FOXM1* and *ANXA2R* RNA-seq profiles showed a highly significant positive correlation between their expression levels in 33 samples from our cohort (Figure 6E) and in 299 TCGA glioma samples (Figure 6F). This high level of co-expression

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between *FOXM1* and *ANXA2R* indicates a presence of a regulatory network involving these two genes. To further explore the association between their expression levels and the patient's predicted outcome, we performed survival analyses stratifying patients according to *ANXA2R* or *FOXM1* expression levels (Figure 6G) using TCGA datasets. For both genes, patients with higher expression levels were associated with shorter survival, even when patients with DA and GB were not combined (Figure S6). This shows that the identified *ANXA2R* regulation may be clinically relevant.

To validate experimentally the discovered regulatory link, we evaluated *ANXA2R* and *FOXM1* expression levels in different glioma cells: LN18, LN229, U87 established glioma cell lines, patient-derived WG4, IPIN glioma cell cultures (Ciechomska et al., 2016) and normal human astrocytes (NHA) using quantitative PCR (qPCR). Relatively to normal human astrocytes, the expression of *ANXA2R* in those cells was elevated only in patient-derived WG4 glioma cells; while *FOXM1* expression was significantly elevated in 4 out of 5 cell lines (Figure 6H). For further studies we have selected WG4 glioma cells due to the highest *ANXA2R* and *FOXM1* expression in those cells.

In order to test whether the predicted enhancer-promoter contact occurs in the context of glioma cells, we performed a Capture-C assay on WG4 cells, retrieving the gene promoter cleavage fragment with all interacting DNA fragments. The profile obtained by plotting the data as a function of chromosome 5 genomic location (Figure 6I) shows a pattern similar to the one that can be obtained with 4C-seq. The coverage plot shows enrichment in reads around the studied enhancer denoted by a purple arrow (Figure 6I), confirming occurrence of the enhancer-promoter contact in glioma cells. As illustrated in Figure 6J, FOXM1 binds to the *CCNB1* gene promoter in WG4 cells (*CCNB1* encodes CYCLINB1 and is a known FOXM1 target in cancer cells) as demonstrated by ChIP-qPCR. FOXM1 binds to both the promoter and distal regulatory region of the *ANXA2R* gene, although only binding to the promoter is statistically significant. Binding of FOXM1 to the *ANXA2R* gene was corroborated with two antibodies from independent suppliers. These results confirm that the predicted chromosomal contact in the *ANXA2R* locus is consistent with the interaction network obtained with Capture-C and validated by ChIP-qPCR.

DISCUSSION

Accessible chromatin across the genome reflects a network of permitted interactions through which enhancers, promoters, insulators and chromatin-binding factors cooperatively regulate gene expression. We combined several methods using massively parallel DNA sequencing to assay chromatin accessibility, active or repressive modification of histones to map regulatory sites instructive for gene expression in gliomas. Gliomas arise from neural stem cells, progenitors or dedifferentiated nervous system cells (Azzarelli et al, 2018). It is expected to find active chromatin regions characteristic of the tissue from which tumor originated as well as tumor specific activity in the data obtained from the bulk tumor material.

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

By combining acquired data, we create the Atlas regulatory sites in gliomas encompassing brain-specific regulatory sites. Within the presented Atlas that describes glioma epigenome features we discriminated two categories of identified regulatory regions: common and variable. The dataset integrating cis-acting regulatory elements which are commonly present in our cohort tend to exhibit the brain-specific signature supported by evolutionary conservation analysis and intersection with publicly available data for non-brain cells. Taking this into account, our Atlas can serve as a tool not only for assessment of gliomagenesis related events, it can also be beneficial in the field of neuropsychiatric disorders research giving the evidence of biologically relevant role of sites overlapping with GWAS-identified SNPs (Liu et al., 2019).

While the utility of the mapped common promoters and enhancers is broader than just its application to the glioma samples, our study has uncovered also a large universe of distal regions that are surprisingly variable in their activity between samples from different glioma grades. As we have shown in this study, many of these, previously unannotated regions can be linked to their likely target genes using Hi-C data from publicly available datasets, obtained from both glioma and fetal brain samples. Such mapping can be very useful in studying biological mechanisms operating in gliomas of different grades, as exemplified in our analysis of the ANXA2R enhancer. The putative enhancer sequence showed a presence of several transcription factor binding motifs, including motifs for a FOXM1 transcription factor. The expression of *FOXM1* and *ANXA2R* RNA-seq profiles showed a strong correlation between our samples and in 299 independent TCGA glioma samples. Importantly, their higher expression levels are associated with the shorter patient survival. We validated the interactions using ChIP-qPCR and Capture-C assay. ANXA2-ANXA2R axis known to play a role in cancerogenesis by promoting cell invasion and migration (D'Souza et al., 2012; Shiozawa et al., 2008) While its role on glioblastomas needs to be elucidated, an inverse correlation of its expression with patient survival indicates clinical relevance.

We interrogated the genome-wide maps of open chromatin and modified histones correlating with transcriptional activity to find regulatory networks distinct in tumors of various grades. We showed significant differences between the H3K27ac and H3K4me3 profiles in PA and DA/GB samples. We found significantly higher H3K27ac levels in DA/GBs versus PA globally, but a detailed analysis uncovered a group of genes with H3K27ac levels higher in the PAs, contrary to the global trend. Closer inspection of these genes showed that they are characterized by neuronal functions (synaptic transmission and signaling, nervous system development, AMPA receptor activity), reflecting higher contribution of non-transformed neural compartments in PAs. A set of transcription factor binding site motifs indicates potential regulation of these genes by transcription factors such as NRF1, REST, E2F1/4 CTCF, MAX. *NRF1* encodes Nuclear Respiratory Factor 1, a transcription factor that activates the expression of mitochondrial oxidative metabolism genes (Johar et al., 2013) that are critical for the maintenance of neuronal homeostasis. E2F2, and E2F4 are

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transcription factors that control numerous target genes, playing a role in cell-cycle progression and apoptosis, with E2F4 acting as a transcriptional repressor (Crosby and Almasan, 2004). The Myc/Max/Mad network depending on composition of Myc-Max or Mad-Max complexes directs gene-specific transcriptional activation or repression (Grandori et al., 2000). Unlike higher-grade gliomas, which usually exhibit multiple driver mutations (Network et al., 2008), most PAs exhibit a single driver somatic genetic alteration, leading to activation of the MAPK pathway related to FGFR or NTRK2 (Jones et al., 2013) or with rearrangements generating the KIAA1549-BRAF fusion oncogene accounting for ~70% of PAs (Jones et al., 2008). A recent single cell sequencing of PAs indicates a higher proportion of mature glia-like cells to progenitor-like cells in PAs as compared to other gliomas (Reitman et al., 2019).

In majority of PA samples we identified a set of H3K4me3 marked promoters which are characterized by a significant increase in the transcript levels, H3K27ac and ATAC-seq signal intensity in comparison to DA/GB samples (Figure 5B, Figure S5). This promoter set is strongly enriched in binding sites for the PRC2 complex components: EZH2 and SUZ12 proteins according to ChIP-seq data from the ENCODE project (Gerstein et al., 2012). These are likely to be differentiation related genes, consistently with the known dysregulation of EZH2 and its regulatory networks in high grade gliomas that are characterized by abundance of dedifferentiated glioma stem cells (Jin et al., 2017).

Altogether, we present here the results that provide the first resource integrating so many different types of data on epigenetic and regulatory profiles in gliomas of various grades. We showed several novel findings based on this resource, but we are confident that this resource will pave the way to the potential future strategies in the treatment of gliomas.

Figure Legends

Figure 1. Experimental workflow

Figure 2 Construction of an atlas of regulatory sites in gliomas.

(A) Numbers of samples from gliomas of different grades for which ATAC-seq, ChIP-seq and RNA-seq data were collected.

(B) Representative peaks of ATAC-seq, ChIP-seq, RNA-seq at the *NRXN2* gene in pilocytic astocytomas (PA04) and glioblastoma (GB01) samples. Note high expression of the *NRXN2* in PA in comparison to GB, which correlates with the lack of the ATAC-seq signal in GB (shadowed blocks).

(C) Total numbers of peaks identified in all samples across performed experiments.

(D) Numbers of identified common and variable active regulatory elements, genes and transcripts.

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(E) Example of promoter-enhancer interactions identified using Hi-C data (Won et al., 2016). Brown dots represent "anchors", while brown horizontal lines depict predicted interaction sites.

Figure 3. Global characterization of chromatin structure and its relationship with gene expression.

(A) Commonality of histone marks, open chromatin regions and gene expression patterns between samples. Color scale represents the percentage of patients sharing a particular genomic element status. Bar height represents a number of genomic elements (scaled to 100%).
(B) Venn diagram shows intersection of the promoter regions marked by H3K4me3, H3K27ac and open chromatin sites detected with ATAC-seq.
(C) Evolutionary conservation of the identified promoters (green) and enhancers (purple) measured by PhastCons 100-Way scores. Mean scores for the identified regulatory regions are compared to mean scores obtained for random genomic intervals. P-values were calculated with one-sided Mann-Whitney U test.
(D) Correlation between H3K4me3 (green), H3K27ac (purple) and transcript expression. Normalized ChIP-seq coverage is plotted against an expression level of transcripts of protein coding genes, divided into quantiles.
(E) Hierarchical clustering of samples based on gene expression profiles. Shades of blue indicate tumor grade (from light = normal brain to dark = Grade IV and black PG). In the table on the right side the subtype of glioblastoma samples (GBM) is reported as well as assignment of glioma samples to one of Pan Glioma RNA Expression Clusters (LGr) as well as *IDH1* gene mutation status (IDH).

(F) Correlation between H3K27ac ChIP-seq coverage on enhancers and transcript expression. Numbers above bars indicate numbers of enhancers in each group.

Figure 4 Chromatin activity profiles indicate presence of normal brain signature in pilocytic astrocytomas

(A,B) Scatter-plots representing abundance of the H3K4me3 (A) and H3K27ac (B) marks obtained for PAs (X-axes) and DA/GBs (Y-axes). A single point on the scatter-plots represents an average abundance (ChIP-input) of a corresponding histone mark around transcription start site [TSS+/-2kb]. P-values shown above the plots were estimated with the Wilcoxon rank-sum test.
(C) Comparison of abundance values shown on Figure 4A (X-axis) and 4B (Y-axis). The dashed frame indicates the strongest difference in H3K27ac abundance (top 10%) values.
(D) Heatmaps showing H3K27ac abundance around the TSS for the top 10% regions (each row corresponds to a TSS, sorted by signal abundance)
(E) ATAC-seq profiles of 4 PA samples and 4 DA/GB samples (see Methods for details) around TSS overlap with the regions selected on Figure 4C. The blue and red lines

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

correspond to PA and DA/GB samples respectively.

(F) Average H3K27ac profiles around TSS. The blue and red lines correspond to the top 10% regions selected in Figure 4C and were computed for the PA and DA/GB samples respectively. The gray lines correspond to regions not selected on Figure 4C. The thick lines correspond to the average profiles, while colored areas give reference of the confidence interval for mean (CI).

(G) Differences of the gene expression levels computed either for genes corresponding to the top 10% selected regions (left boxplot) or to the rest of the regions (right boxplot). P-value shown above the plot and indicating differential regulation was estimated with the Wilcoxon rank-sum test.

(H) Results of Gene Ontology over-representation analysis performed for the genes corresponding to the top 10% regions. The barplot shows scaled Bonferroni corrected p-values and the vertical, dashed red line stands for the significance threshold (Pv = 0.05).

(I) Logos of top DNA-binding motifs enriched in the top 10% regions. The enrichment was computed with regard to the rest of the analyzed regions. P-values indicated above the logos stand for the significance of the enrichment and were computed with a tool from MEME-Suite (see Methods for details).

Figure 5. A group of promoters targeted by PRC2 exhibits H3K4me3 hypermethylation and DNA hypomethylation in PAs.

(A) Hierarchical clustering of samples based on the presence of H3K4me3 mark at the promoters. Color scale and numbers on the heatmap indicate similarity between H3K4me3 pattern at the promoters (dark blue, 1 - identical; white, 0 - maximally dissimilar). Red box – identified cluster of PA samples.

(B) Expression of transcripts associated with promoters exhibiting H3K4me3 hypermethylation in PAs. Enrichment of binding sites from the ENCODE transcription factor ChIP-seq data in PAs specific promoters compared to promoters active in all samples. The number of peaks for each DNA-binding protein is given in parentheses; enrichment values per megabase are shown next to each bar and indicated with proportional coloring (red – highest, blue-lowest).

(C) Enrichment of binding sites from the ENCODE transcription factor ChIP-seq data in PAs specific promoters compared to promoters active in all samples. The number of peaks for each DNA-binding protein is given in parentheses; enrichment values per megabase are shown next to each bar and indicated with proportional coloring (red – highest, blue-lowest).

(D) DNA methylation in promoters exhibiting H3K4me3 hypermethylation in PAs and having an EZH2 binding site. Methylation levels were determined with WGBS. Samples with *IDH* mutations were excluded from the plot.

(E) DNA methylation in promoters exhibiting H3K4me3 hypermethylation in PAs and having an EZH2 binding site in an independent cohort, determined by HumanMethylation450 BeadChip array.

(F) GO terms enriched in the set of the PA-specific genes with EZH2 binding sites in

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

promoters. The enrichment was calculated using the PANTHER Overrepresentation Test (Mi et al., 2019).

Figure 6. Intersection of the Atlas data with fetal brain Hi-C data reveals FOXM1-ANXA2R network operating in glioblastoma and impacting patients survival.

(A) A graphical representation of the chromatin contacts identified based on fetal brain Hi-C data.

(B) A heatmap shows normalized average expression levels of 17 differentially expressed genes (FDR < 0.01) that were selected in the contact analysis (Figure 6A, see Methods for details). Genes which expression levels were increased in high grade gliomas are depicted.

(C) A barplot shows significance of enhancer-promoter co-activation for the 5 genes shown in Figure 6B. The co-activation was computed with Pearson's correlation of H3K27ac enrichment in the putative enhancer and H3K4me3 enrichment in the corresponding promoter.

(D) Expression levels of *FOXM1* gene in glioma samples. The adjusted P-value (FDR) corresponds to differences in expression levels and was computed with edgeR Bioconductor package. Logo showed above the plot represents FOXM1 DNA-binding motif.

(E, F) Scatterplots show co-expression of *FOXM1* (X-axes) and *ANXA2R* (Y-axes) in our (E) and TCGA (F) data. Gene expression levels are represented as log-transformed FPKMs. Corresponding Pearson's correlation coefficients are shown above the plots. Tumor grades were color-coded.

(G) Kaplan-Meier curves plotted for the TCGA data. The patients were stratified by either by *ANXA2R* (top part) or by *FOXM1* (bottom part) expression levels into three equinumerous sub-groups. Differences in survival rates between sub-groups were assessed with the log-rank test.

(H) *ANXA2R* and *FOXM1* mRNA expression in normal human astrocytes (NHA), glioma patient derived cell cultures (WG4 and IPIN) and established glioma cell lines (LN229, LN18, U87) was evaluated by RT-qPCR. Data were normalized to the expression of *GAPDH* mRNA determined in the same sample. P values were considered significant when *P < 0.05 and **P < 0.01 (t-test).

(I) A coverage plot of WG4 Capture-C profile representing interactions around the viewpoint, ANXA2R promoter fragment, marked with a green arrow. Putative enhancer is denoted by purple arrow indicating region enriched in fragment reads. Fragment-based raw data is visualized as grey dots, whereas blue dots represent smoothed data. The trend line is blue whereas quantiles are shown as light-grey bands.

(J) ChIP-qPCR analysis of FOXM1 binding to the *CCNB1* gene promoter (left part) that serves as a positive control, and to the *ANXA2R* proximal and distal regulatory regions (middle part). Results are calculated as % of input, mean ± SD (n = 3) and compared to IgG. T-test was used for statistical analysis. Antibodies from two different companies showed around 2-fold enrichment in FOXM1 binding to the *ANXA2R* gene promoter over a neutral IgG

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgments

This work was supported by Polish National Science Centre grant DEC-2015/16/W/NZ2/00314] and by the Foundation for Polish Science TEAM-TECH Core Facility project "NGS platform for comprehensive diagnostics and personalized therapy in neuro-oncology" (BW, BG). We would like to thank Marcin Kruczyk for his contributions to the concept of the study, Luca Giorgetti and Mariya Kryzhanovska for their help in performing the Capture-C experiments and Jacek Koronacki for comments on the manuscript.

Author Contributions

Conceptualization: JK BK MJD KD BWi KS; Methodology: BK, KS, Investigation in the lab: KS, RG, SK, BG; Computational investigation: BWo, MM, AMac, JM, AD, IG, HK, KS, MD, MJD, MJ, BWi; Resources: KK, WG, MR, TC AMar, Writing - Original draft: KS, JM, MM, BK, BWi; Writing - Review & editing: BWi, BK, KS, MJD, MM, KD, JK. All authors have read and approved the final version of the manuscript.

Declaration of Interests

The authors declare no competing interests.

References

Azzarelli, Roberta, Benjamin D. Simons, and Anna Philpott. "The developmental origin of brain tumours: a cellular and molecular framework." *Development* 145.10 (2018).

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The Mammalian Epigenome. Cell *128*, 669–681.

Bogdanovic, O., Fernandez-Minan, A., Tena, J.J., de la Calle-Mustienes, E., Hidalgo, C., van Kruysbergen, I., van Heeringen, S.J., Veenstra, G.J.C., and Gomez-Skarmeta, J.L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res. *22*, 2043–2053.

Bonn, S., Zinzen, R.P., Girardot, C., Gustafson, E.H., Perez-Gonzalez, A., Delhomme, N., Ghavi-Helm, Y., Wilczyński, B., Riddell, A., and Furlong, E.E.M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat. Genet. *44*, 148–156.

Cadieux, B., Ching, T.T., VandenBerg, S.R., and Costello, J.F. (2006). Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation. Cancer Res. *66*, 8469–8476.

Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A.,

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., et al. (2016). Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. Cell *164*, 550–563.

Chang, W.-T., Chen, H., Chiou, R.-J., Chen, C.-Y., and Huang, A.-M. (2005). A novel function of transcription factor alpha-Pal/NRF-1: increasing neurite outgrowth. Biochem. Biophys. Res. Commun. *334*, 199—206.

Ciechomska, I.A., Przanowski, P., Jackl, J., Wojtas, B., and Kaminska, B. (2016). BIX01294, an inhibitor of histone methyltransferase, induces autophagy-dependent differentiation of glioma stem-like cells. Sci. Rep. *6*, 38723.

Crosby, M.E., and Almasan, A. (2004). Opposing roles of E2Fs in cell proliferation and death. Cancer Biol. Ther. *3*, 1208–1211.

D'Souza, S., Kurihara, N., Shiozawa, Y., Joseph, J., Taichman, R., Galson, D.L., and Roodman, G.D. (2012). Annexin II interactions with the annexin II receptor enhance multiple myeloma cell adhesion and growth in the bone marrow microenvironment. Blood *119*, 1888–1896.

Fontebasso, A.M., Liu, X.-Y., Sturm, D., and Jabado, N. (2013). Chromatin Remodeling Defects in Pediatric and Young Adult Glioblastoma: A Tale of a Variant Histone 3 Tail. Brain Pathol. *23*, 210–216.

Fullard, J.F., Giambartolomei, C., Hauberg, M.E., Xu, K., Voloudakis, G., Shao, Z., Bare, C., Dudley, J.T., Mattheisen, M., Robakis, N.K., et al. (2017). Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. Hum. Mol. Genet. *26*, 1942–1951.

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. Nature *489*, 91–100.

Grandori, C., Cowley, S.M., James, L.P., and Eisenman, R.N. (2000). The Myc/Max/Mad Network and the Transcriptional Control of Cell Behavior. Annu. Rev. Cell Dev. Biol. *16*, 653–699.

Harewood, Louise, Kamal Kishore, Matthew D. Eldridge, Steven Wingett, Danita Pearson, Stefan Schoenfelder, V. Peter Collins, and Peter Fraser. "Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours." *Genome biology* 18, no. 1 (2017): 125.

Jin, X., Kim, L.J.Y., Wu, Q., Wallace, L.C., Prager, B.C., Sanvoranart, T., Gimple, R.C., Wang, X., Mack, S.C., Miller, T.E., et al. (2017). Targeting glioma stem cells through combined BMI1 and EZH2 inhibition. Nat. Med. *23*, 1352.

Johar, K., Priya, A., Dhar, S., Liu, Q., and Wong-Riley, M.T.T. (2013). Neuron-specific specificity protein 4 bigenomically regulates the transcription of all mitochondria- and nucleus-encoded cytochrome c oxidase subunit genes in neurons. J. Neurochem. *127*, 496–508.

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Jones, D.T.W., Kocialkowski, S., Liu, L., Pearson, D.M., Bäcklund, L.M., Ichimura, K., and Collins, V.P. (2008). Tandem Duplication Producing a Novel Oncogenic BRAF Fusion Gene Defines the Majority of Pilocytic Astrocytomas. Cancer Res. *68*, 8673–8677.

Jones, D.T.W., Hutter, B., Jäger, N., Korshunov, A., Kool, M., Warnatz, H.-J., Zichner, T., Lambert, S.R., Ryzhova, M., Quang, D.A.K., et al. (2013). Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. Nat. Genet. *45*, 927.

Karlic, R., Chung, H.-R., Lasserre, J., Vlahovicek, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. Proc. Natl. Acad. Sci. *107*, 2926–2931.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693–705.
Lahoz, E.G., Xu, L., Schreiber-Agus, N., and DePinho, R.A. (1994). Suppression of Myc, but not E1a, transformation activity by Max-associated proteins, Mad and Mxi1. Proc. Natl. Acad. Sci. *91*, 5503 LP-5507.

Lambert, S.R., Witt, H., Hovestadt, V., Zucknick, M., Kool, M., Pearson, D.M., Korshunov, A., Ryzhova, M., Ichimura, K., Jabado, N., et al. (2013). Differential expression and methylation of brain developmental genes define location-specific subsets of pilocytic astrocytoma. Acta Neuropathol. *126*, 291–301.

Larjavaara, S., Mäntylä, R., Salminen, T., Haapasalo, H., Raitanen, J., Jääskeläinen, J., and Auvinen, A. (2007). Incidence of gliomas by anatomic location. Neuro. Oncol. *9*, 319–325.
Liu, C., Kanazawa, T., Tian, Y., Mohamed Saini, S., Mancuso, S., Mostaid, M.S., Takahashi, A., Zhang, D., Zhang, F., Yu, H., et al. (2019). The schizophrenia genetics knowledgebase: a comprehensive update of findings from candidate gene studies. Transl. Psychiatry *9*, 205.

Louis, D.N., Perry, A., Reifenberger, G., Deimling, A. Von, Figarella, D., Webster, B., Hiroko, K.C., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System : a summary. Acta Neuropathol. 131(6):803-20

Mack, S.C., Singh, I., Wang, X., Hirsch, R., Wu, Q., Villagomez, R., Bernatchez, J.A., Zhu, Z., Gimple, R.C., Kim, L.J.Y., et al. (2019). Chromatin landscapes reveal developmentally encoded transcriptional states that define human glioblastoma. J. Exp. Med. *216*, 1071 LP-1090.

Maji, S., Chaudhary, P., Akopova, I., Nguyen, P.M., Hare, R.J., Gryczynski, I., and Vishwanatha, J.K. (2017). Exosomal Annexin II Promotes Angiogenesis and Breast Cancer Metastasis. Mol. Cancer Res. *15*, 93–105.

Maleszewska, M., and Kaminska, B. (2013). Is glioblastoma an epigenetic malignancy? Cancers (Basel). *5*, 1120–1139.

Manni, I., Tunici, P., Cirenei, N., Albarosa, R., Colombo, B.M., Roz, L., Sacchi, A., Piaggio, G., and Finocchiaro, G. (2002). Mxi1 inhibits the proliferation of U87 glioma cells through down-regulation of cyclin B1 gene expression. Br. J. Cancer *86*, 477—484.

Network, T.C.G.A.R., McLendon, R., Friedman, A., Bigner, D., Van Meir, E.G., Brat, D.J., M. Mastrogianakis, G., Olson, J.J., Mikkelsen, T., Lehman, N., et al. (2008). Comprehensive

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genomic characterization defines human glioblastoma genes and core pathways. Nature *455*, 1061.

Piao, Y., Kim, H.G., Oh, M.S., and Pak, Y.K. (2012). Overexpression of TFAM, NRF-1 and myr-AKT protects the MPP+-induced mitochondrial dysfunctions in neuronal cells. Biochim. Biophys. Acta - Gen. Subj. *1820*, 577–585.

Pidsley, R., Viana, J., Hannon, E., Spiers, H., Troakes, C., Al-saraj, S., Mechawar, N., Turecki, G., Schalkwyk, L.C., Bray, N.J., et al. (2014). Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. Genome Biol. 15(10):483

Racko, D., Benedetti, F., Dorier, J., and Stasiak, A. (2018). Are TADs supercoiled? Nucleic Acids Res. *47*, 521–532.

Reifenberger, G., Wirsching, H.-G., Knobbe-Thomsen, C.B., and Weller, M. (2016). Advances in the molecular genetics of gliomas — implications for classification and therapy. Nat. Rev. Clin. Oncol. *14*, 434.

Reitman, Z.J., Paolella, B.R., Bergthold, G., Pelton, K., Becker, S., Jones, R., Sinai, C.E., Malkin, H., Huang, Y., Grimmet, L., et al. (2019). Mitogenic and progenitor gene programmes in single pilocytic astrocytoma cells. Nat. Commun. *10*, 3731.

Sharma, V., Malgulwar, P.B., Purkait, S., Patil, V., Pathak, P., Agrawal, R., Kulshreshtha, R., Mallick, S., Julka, P.K., Suri, A., et al. (2017). Genome-wide ChIP-seq analysis of EZH2-mediated H3K27me3 target gene profile highlights differences between low- and high-grade astrocytic tumors. Carcinogenesis *38*, 152–161.

Shiozawa, Y., Havens, A.M., Jung, Y., Ziegler, A.M., Pedersen, E.A., Wang, J., Wang, J., Lu, G., Roodman, G.D., Loberg, R.D., et al. (2008). Annexin II/Annexin II receptor axis regulates adhesion, migration, homing, and growth of prostate cancer. J. Cell. Biochem. *105*, 370–380.

Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. Nat. Rev. Mol. Cell Biol. *15*, 703.

Tome-Garcia, J., Erfani, P., Nudelman, G., Tsankov, A.M., Katsyv, I., Tejero, R., Zhang, B., Walsh, M., Friedel, R.H., Zaslavsky, E., et al. (2018). Analysis of chromatin accessibility uncovers TEAD1 as a regulator of migration in human glioblastoma. Nat. Commun. *9*, 4020.

Won, H., de la Torre-Ubieta, L., Stein, J.L., Parikshak, N.N., Huang, J., Opland, C.K., Gandal, M., Sutton, G.J., Hormozdiari, F., Lu, D., et al. (2016). Chromosome conformation elucidates regulatory relationships in developing human brain. Nature *538*, 523–527.

Collected tumor specimens

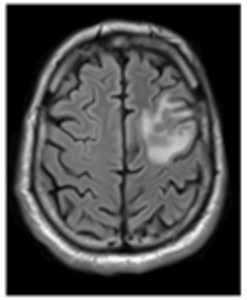
Pilocytic Astrocytoma samples (**PA**)

Glioblastoma samples (**GB**)

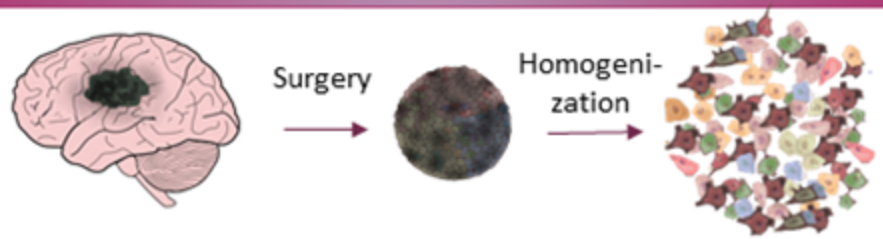
WHO II/III samples (**DA**)

Clinical data

- Age
- Gender
- Pathology
- Tumor location
- Molecular testing

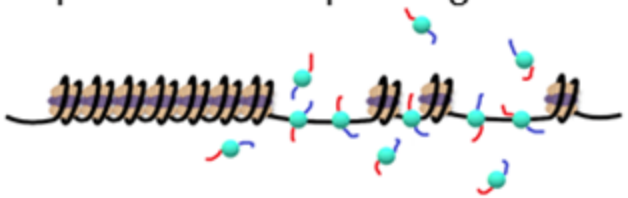


Experimental workflow

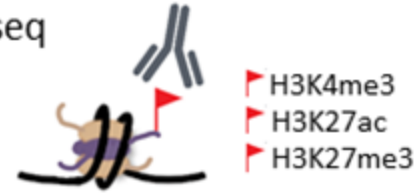


- Gene expression analysis with RNA-seq

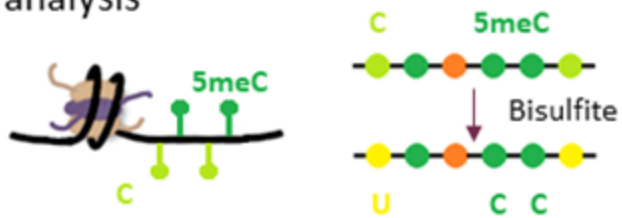
- Open chromatin profiling with ATAC-seq



- Chromatin modifications profiling with ChIP-seq

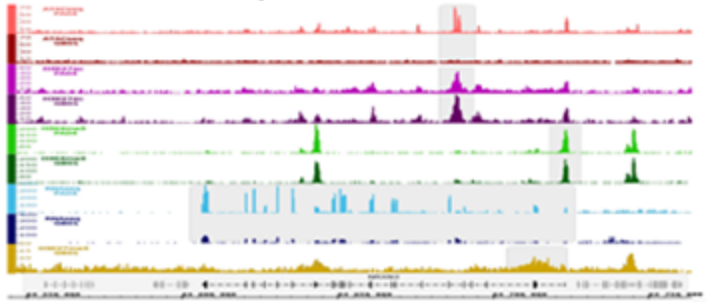


- Genome-wide DNA methylation analysis

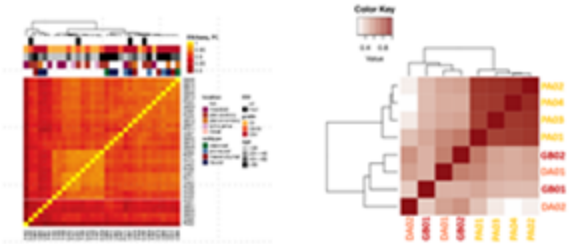


Data analysis and visualization

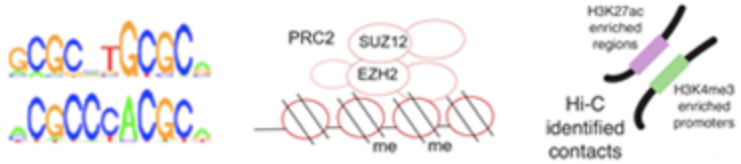
- Genome browser with coverage for each experiment



- Transcriptional and epigenomic differential analysis across glioma grades



- Identification of regulatory mechanisms behind observed data

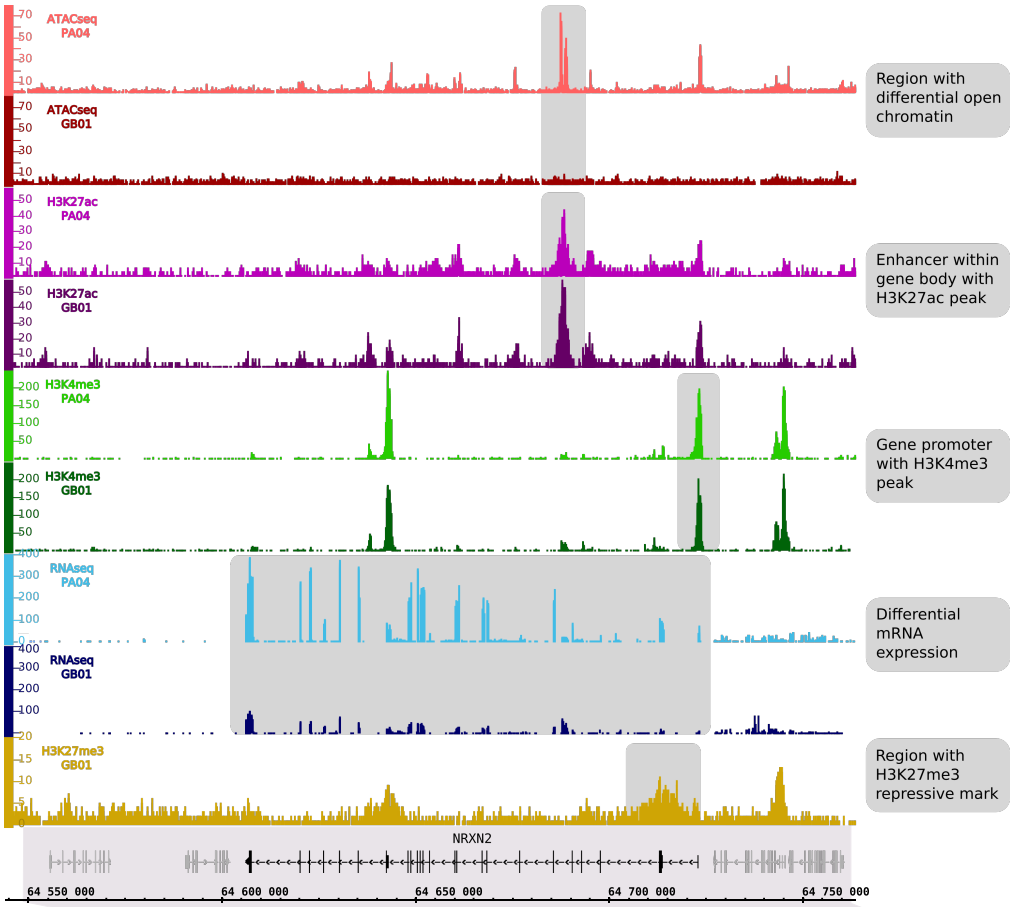


bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A

	ATAC-seq	H3K27ac	H3K4me3	RNA-seq	H3K27me3	sum
PA	4	3	7	11	0	25
DA	2	5	6	7	5	25
GD	2	8	11	15	7	43
sum	8	16	24	33	12	**93**

B



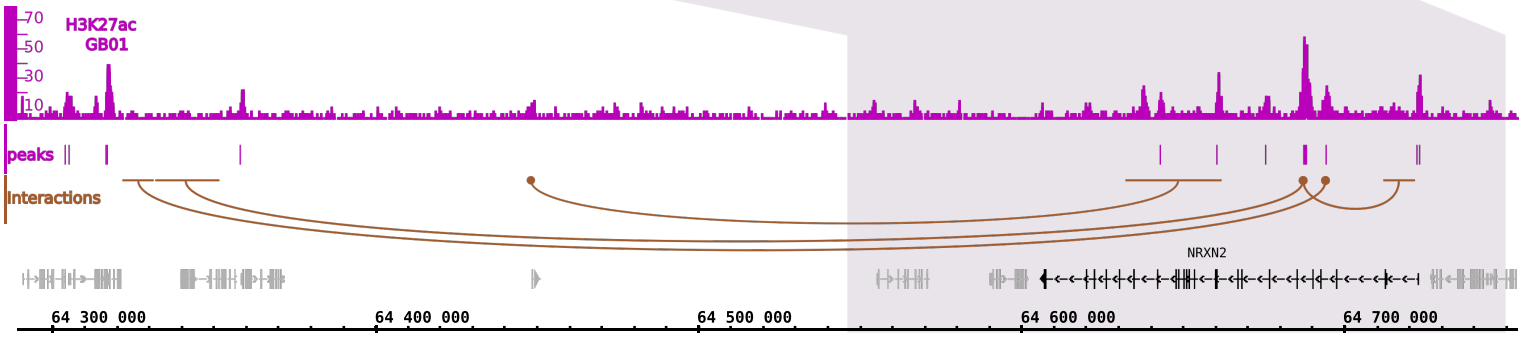
C

	ATAC-seq	H3K4me3	H3K27ac	H3K27me3
total number of peaks	3,050,914	523,359	351,018	585,162

D

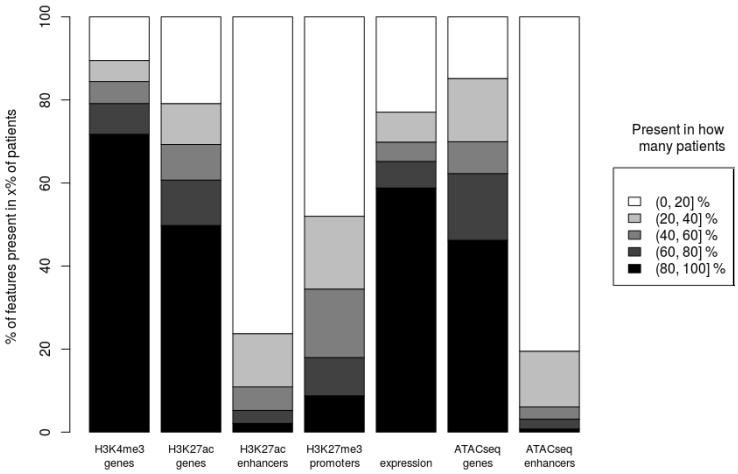
	H3K27ac		H3K4me3	
	common peaks	variable peaks	common peaks	variable peaks
genes	12,728	8,241	17,294	6,807
transcripts	57,212	35,599	78,565	25,191
putative enhancers	10,673	34,875		

E

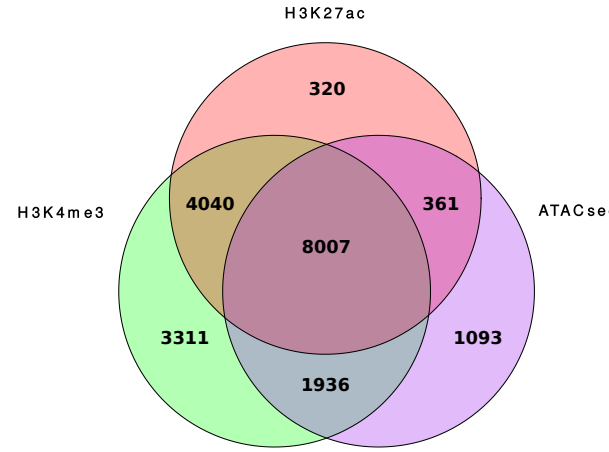


bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

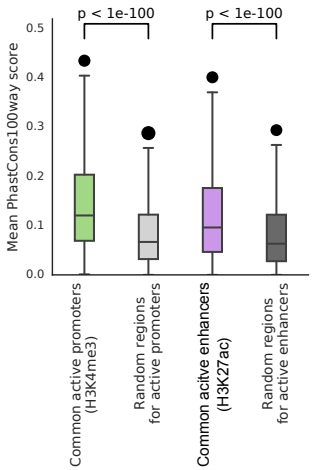
A



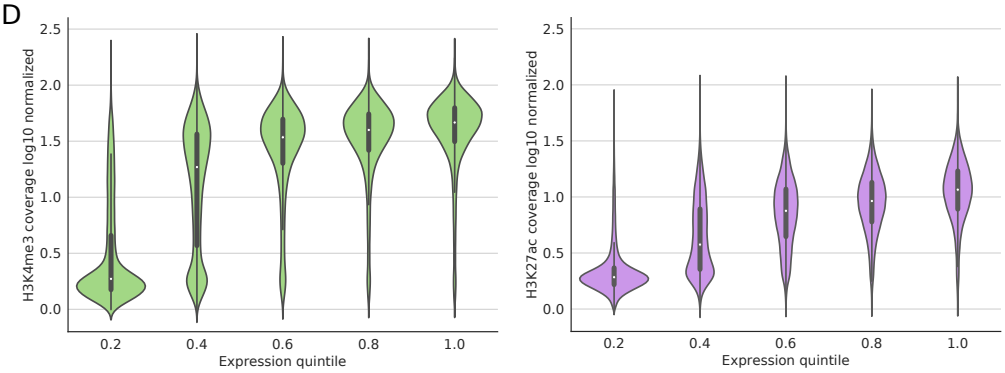
B



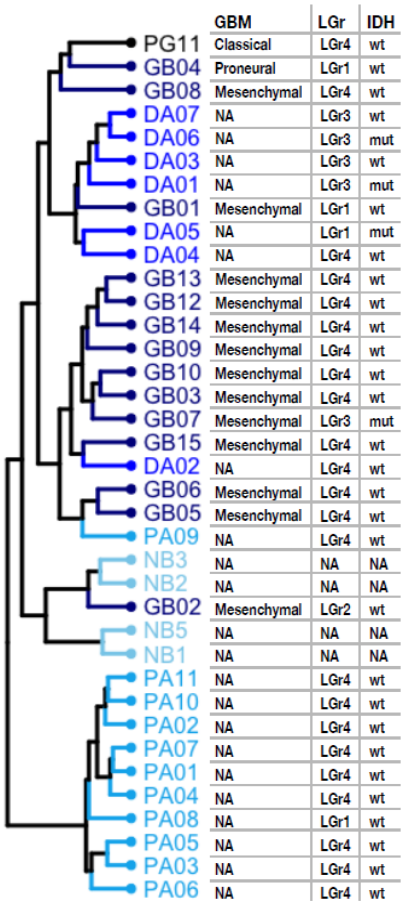
C



D

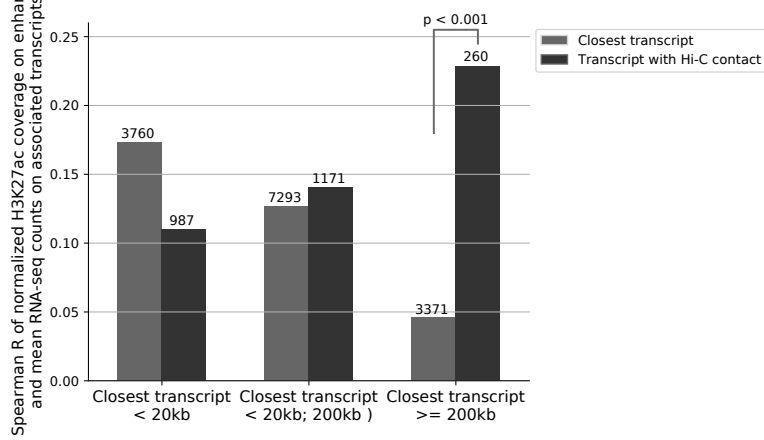


E



	GBM	LGr	IDH
PG11	Classical	LGr4	wt
GB04	Proneural	LGr1	wt
GB08	Mesenchymal	LGr4	wt
DA07	NA	LGr3	wt
DA06	NA	LGr3	mut
DA03	NA	LGr3	wt
DA01	NA	LGr3	mut
GB01	Mesenchymal	LGr1	wt
DA05	NA	LGr1	mut
DA04	NA	LGr4	wt
GB13	Mesenchymal	LGr4	wt
GB12	Mesenchymal	LGr4	wt
GB14	Mesenchymal	LGr4	wt
GB09	Mesenchymal	LGr4	wt
GB10	Mesenchymal	LGr4	wt
GB03	Mesenchymal	LGr4	wt
GB07	Mesenchymal	LGr3	mut
GB15	Mesenchymal	LGr4	wt
DA02	NA	LGr4	wt
GB06	Mesenchymal	LGr4	wt
GB05	Mesenchymal	LGr4	wt
PA09	NA	LGr4	wt
NB3	NA	NA	NA
NB2	NA	NA	NA
GB02	Mesenchymal	LGr2	wt
NB5	NA	NA	NA
NB1	NA	NA	NA
PA11	NA	LGr4	wt
PA10	NA	LGr4	wt
PA02	NA	LGr4	wt
PA07	NA	LGr4	wt
PA01	NA	LGr4	wt
PA04	NA	LGr4	wt
PA08	NA	LGr1	wt
PA05	NA	LGr4	wt
PA03	NA	LGr4	wt
PA06	NA	LGr4	wt

F



bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

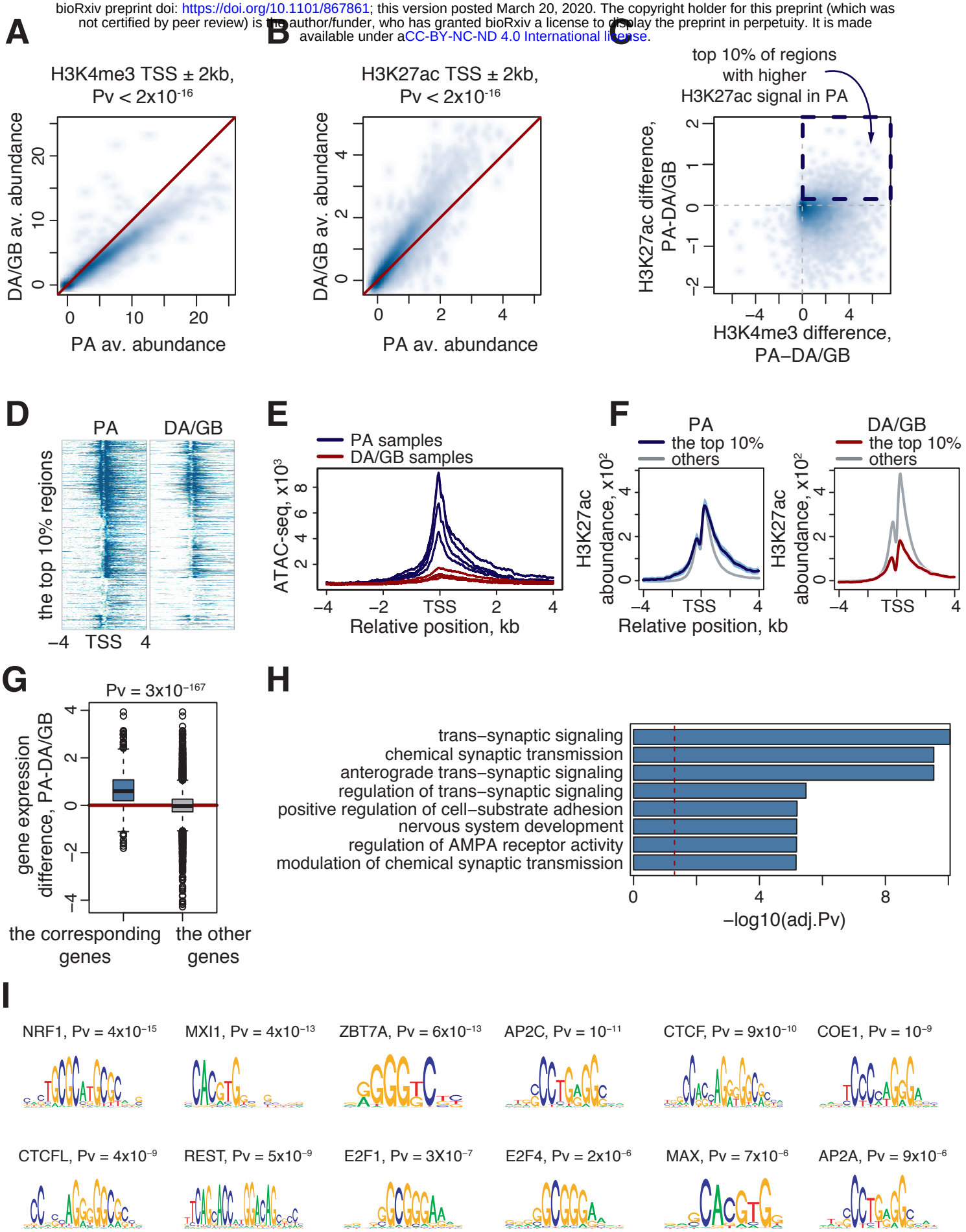
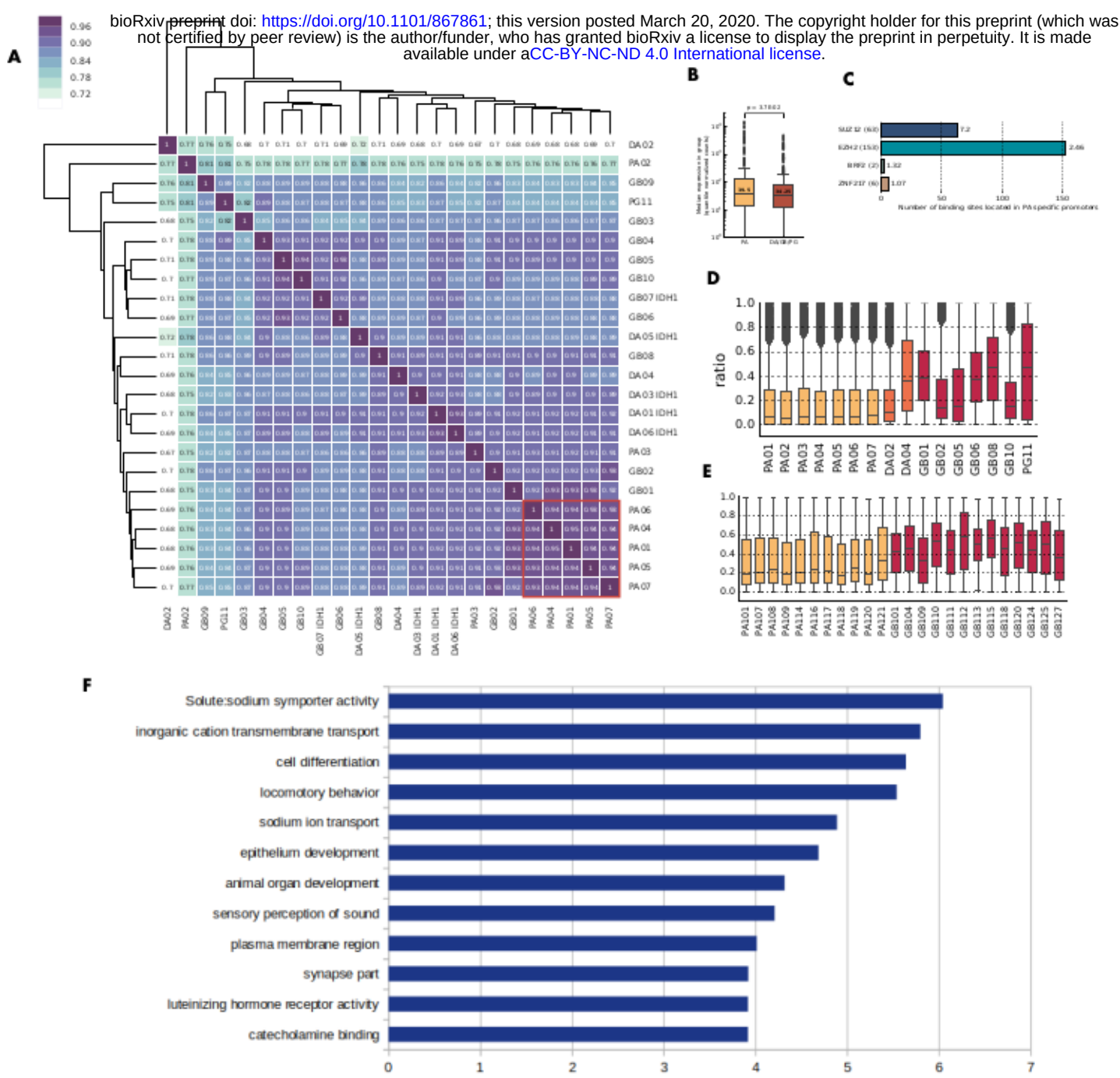


Figure 4

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

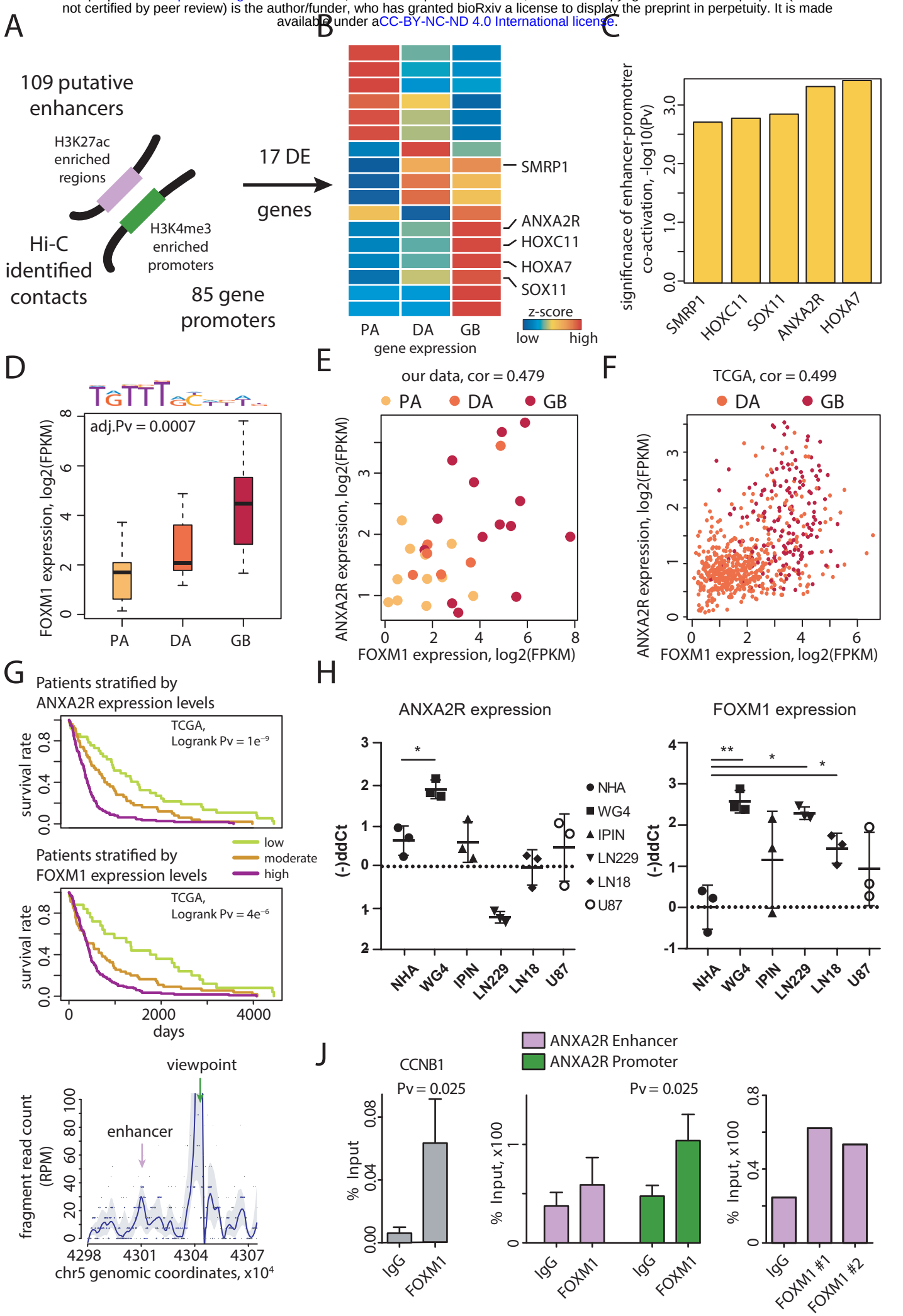


Figure 6

STAR+METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-H3K4me3	Merck Millipore	Cat. number 07-473, RRID:AB_1977252
Rabbit polyclonal anti-H3K27ac	Active Motif	Cat. number 39133, RRID:AB_2561016
Rabbit monoclonal anti-H3K27me3	Abcam	Cat. number ab192985, RRID:AB_2650559
Rabbit polyclonal anti-FOXM1	Diagenode	Cat. number C15410232-100
Rabbit Normal IgG Control Antibody, Unconjugated	Cell Signalling	Cat. number 2729, RRID:AB_1031062
Biological Samples		
See Table S1	This paper	NA
Chemicals, Peptides, and Recombinant Proteins		
37% Formaldehyde	Sigma-Aldrich	CAS number 50-00-0
A- and G-Sepharose beads	Merck Millipore	Cat. number 16-156 & 16-266
Adapters and USER enzyme	New England Biolabs	Cat. number E7335L
Ampure Beads	Beckman Coulter	Cat. number A63880
DMEM	ThermoFisher Scientific	Ref 31885-023
DMEM/F-12, GlutaMAX™	ThermoFisher Scientific	Ref 31966-021
DNA polymerase	EURx	Cat. number E2500-01
Dynabeads Protein A	ThermoFisher Scientific	Cat. number 10002D
FBS	ThermoFisher Scientific	Cat. number 1056000
MboI	NEB	Cat. number R0147M
Proteinaise K	Applied Biosystems	Cat. number 4333793
RNase A	Invitrogen	Cat. number 12091-021
SimplySafe™	EURx	Cat. number E4600-01
SYBR Green chemistry	Applied Biosystem by Thermo Fisher Scientific	Cat. Number 4385612
T4 DNA ligase	NEB	Cat. number M0202T
TRI Reagent	Sigma-Aldrich	Product No.T9424
Critical Commercial Assays		

Agilent DNA High Sensitivity Kit	Agilent Technologies, Ltd.	Cat. number 5067-4626
Agilent RNA 6000 Nano Kit	Agilent Technologies, Ltd.	Cat. number 5067-1511
ChIP-IT Express Immunoprecipitation Kit	Active Motif	Cat. number 53008
EZ-96 DNA Methylation-Gold kit	Zymo Research	Cat. number D5007
Infinium HumanMethylation 450 BeadChip	Illumina	Cat. IDs:WG-314-1003
KAPA Stranded mRNA Sample Preparation Kit	Kapa Biosystems	Kit code KK8420 Roche cat no. 07962193001
NEBNext Ultra Library Prep Kit for Illumina	Ilumina	Cat. number E7370S/L
Nextera DNA Library Preparation Kit	Illumina	Cat. number FC-121-1030
Syngen GEL/ PCR Mini Kit	Syngen Biotech	Cat. number SY201010
QIAsymphony DNA Midi kit	Qiagen	Cat. number 931255
QuantiFluor double stranded DNA System	Promega	Cat. number E2670
RNeasy Mini kit	(Qiagen, Hilden, Germany)	Cat. Number 74104
xGen Lockdown Probes and Reagents kit	IDT	Cat. number 1072280
Zymo DNA Clean and Concentrator 5	Zymo Research	Cat. number D4003T
Deposited Data		
DNA methylation data from human brain cerebellum and cortex	(Pidsley et al., 2014)	GSE61431
DNA methylation data from pilocytic astrocytoma samples	(Lambert et al., 2013)	GSE44684
Gene expression data from pilocytic astrocytoma samples	(Lambert et al., 2013)	GSE44971
Experimental Models: Cell Lines		
LN229	ATCC	CRL-2611
LN18	ATCC	CRL-2610
U87	ATCC	HTB-14
WG4	Bozena Kaminska Lab	NA
IPIN	Bozena Kaminska Lab	NA
Oligonucleotides		
IDH_mut_F GGATGCTGCAGAAGCTATAA	Oligo.pl	NA
IDH_mut_R CATGCAAAATCACATTATTGCC	Oligo.pl	NA
GAPDH_ChIP_F TACTAGCGGTTTTACGGGCGCAC	Oligo.pl	NA
GAPDH_ChIP_R TCGAACAGGAGGAGCAGAGAGCGA	Oligo.pl	NA
HOXA7_ChIP_F AGATGCGGAAATTGGCCTCAG	Oligo.pl	NA
HOXA7_ChIP_R TCCTACGACCAAAACATCCCC	Oligo.pl	NA

CCNB1_prom_ChIP_F CGCGATCGCCCTG-GAAACGCA	Oligo.pl	NA
CCNB1_prom_ChIP_R CCCAGCAGAAACCAACAGC-CGT	Oligo.pl	NA
ANXA2R_enh_ChIP_F ATGGACAAACAAACCAA-CAAACA	Oligo.pl	NA
ANXA2R_enh_ChIP_R AGTGCAGTCCATGCAGGTTA	Oligo.pl	NA
ANXA2R_prom_ChIP_F TGTGACCACCGATTCCACTG	Oligo.pl	NA
ANXA2R_prom_ChIP_R ACAGACATTTGCTACGGGCA	Oligo.pl	NA
ANXA2R_qPCR_F CAAGTACAGCGAAGCCCACT	Oligo.pl	NA
ANXA2R_qPCR R CTGAGTCTGTCGGGTTCCTC	Oligo.pl	NA
FOXM1_qPCR_F GCAGCGACAGGTTAAGGTTG	Oligo.pl	NA
FOXM1_qPCR_R GTCATGCGCTTCCTCTCAGT	Oligo.pl	NA
Software and Algorithms		
Basic4CSeq Biocoductor package	walterC. 2019	
bedtools	(Quinlan and Hall, 2010)	
bigWigAverageOverBed		https://github.com/ENCODE-DCC/kentUtils
Bowtie 2.2.6.2	(Langmead, 2013)	
BSMAP	(Xi and Li, 2009)	
CytoMeth	Draminski et al. in prep.	https://github.com/mdraminski/CytoMeth
Dendextend R package	(Galili, 2015)	
EDASeq	(Risso, 2013)	
FastQC		https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
FASTQ Trimmer	(Blankenberg et al., 2010)	
FASTX		http://hannonlab.cshl.edu/fastx_toolkit/
Filter FASTQ	(Blankenberg et al., 2010)	
F-seq1.85	(Boyle et al., 2008)	
HiCEnterprise		https://github.com/hansiu/HiCEnterprise
HiTC Bioconductor package	Servant N, 2012	
MACS2.1	(Zhang et al., 2008)	
methylKit	(Akalin et al., 2012)	
PANTHER Overrepresentation Test (Released 20171205)	(Mi et al., 2017)	
Picard		http://broadinstitute.github.io/picard/

PWMEnrich R package	(Stojnic and Diez, 2018)	
QuasR Bioconductor package	(Gaidatzis D, 2015)	
samtools	(Li et al., 2009a)	
Tophat2	(Kim et al., 2013)	
Trimmomatic	(Bolger et al., 2014)	
Other		
ENCODE Regulation 'Txn Factor' track data V3		http://hgdownload.soe.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeRegTfbsClustered/
GO Ontology database Released 2018-06-01	(Ashburner et al., 2000; Carbon et al., 2017)	
HOCOMOCO v11 database	(Kulakovskiy et al., 2018)	
PhastCons 100-Way scores		http://hgdownload.cse.ucsc.edu/goldenPath/hg38/phastCons100way/hg38.phastCons100way.bw

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human glioma samples

Freshly resected glioma specimens were acquired from several neurosurgery clinics: The Children's Memorial Health Institute, Public Central Clinical Hospital, Institute of Psychiatry and Neurology and Mazovian Brodno Hospital. The tissue collection protocol was approved by the Committees of Bioethics of the institutions listed above (protocol number #14/KBE/2012, #KBE/54/2016, #3/2016). Each patient provided a written consent for use of tumor tissues; afterwards samples were anonymized . Details concerning age, gender, clinical diagnosis, tumor location and molecular test results are provided in the Table S1. Tumor samples were transported in DMEM/F-12 medium on ice and processed directly surgical resection. Procedures were halted at a first safe stop within 1-3 h after acquisition. Tumor specimens were transferred cold PBS, minced with sterile scissors or scalpel on a Petri dish kept on ice and subsequently homogenized by using chilled manual glass douncer. Then homogenized material was aliquoted to perform different methods.

Human glioma cell lines

Human malignant U-87 MG, LN18 and LN229 glioblastoma cells were purchased from American Type Culture Collection (ATCC). WG4 and IPIN are primary glioma cell lines developed from GBM patient surgical samples as described before (Ciechomska et al., 2016).

Method Details

Established glioma cell lines were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (ThermoFisher Scientific) and 100 units/mL of penicillin and 100 µg/mL of streptomycin. Primary glioma cell lines were cultured in DMEM/Nutrient Mixture F-

12, GlutaMAX™ medium (DMEM/F-12, GlutaMAX™) supplemented with 10% fetal bovine serum (ThermoFisher Scientific) and antibiotics (100 U/mL penicillin, 100 µg/mL streptomycin).

Nucleic acids extraction from glioma samples

Total RNA and DNA were isolated using Tri-Reagent extraction (Sigma-Aldrich, Munich, Germany) starting from 50-100 mg of tissues(depending on the initial specimen size). RNA quality and yield were verified by Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) using a RNA 6000 Nano Kit (Agilent Technologies, Ltd.) and NanoDrop 2000 (Thermo Scientific, NanoDrop products, Wilmington, USA). DNA purity was estimated using NanoDrop 2000 (Thermo Scientific, NanoDrop products, Wilmington, USA).

Analysis of *IDH1/2* mutations

To check *IDH1/2* status, 20 ng of genomic DNA was amplified by polymerase chain reaction (PCR) with specific primers amplifying the exon 6,in which common *IDH* mutations occur. PCR reaction consisted standard buffer, forward primer 5'-GGATGCTGCAGAAGCTATAA-3', reverse primer 5'-CATGCAAAATCACATTATTGCC-3'. s and DNA polymerase (EURx) in a total volume of 25 mL, and followed denaturation at 95°C for 30 s, annealing at 54°C for 30 s and extension at 72°C for 50 s for 35 cycles. PCR products were separated on 1% agarose gel as 230 bp fragments, and visualized by SimplySafe™ (EURx) stainingSubsequently, PCR products were purified using Syngen GEL/ PCR Mini Kit (Syngen Biotech) and a mutational status of *IDH* gene in surgical glioma specimens was determined by Sanger sequencing.

RNA sequencing

Strand-specific polyA enriched RNA libraries were prepared using the KAPA Stranded mRNA Sample Preparation Kit according to the manufacturer's protocol (Kapa Biosystems, MA, USA). Briefly, mRNA molecules were enriched from 500ng of total RNA using poly-T oligo-attached magnetic beads (Kapa Biosystems, MA, USA). Obtained mRNA was fragmented and the first-strand cDNA was synthesized using a reverse transcriptase. Second cDNA synthesis was performed to generate double-stranded cDNA (dsDNA). Adenosines were added to the 3′ ends of dsDNA and adapters were ligated (adapters from NEB, Ipswich, MA, USA). Following the adapter ligation, uracil in a loop structure of adapter was digested by USER enzyme from NEB (Ipswich, MA, USA). Adapters containing DNA fragments were amplified by PCR using NEB starters (Ipswich MA, USA). Library evaluation was done with Agilent 2100 Bioanalyzer using the Agilent DNA High Sensitivity chip (Agilent Technologies, Ltd.) Mean library size was 300bp. Libraries were quantified using a Quantus fluorometer and QuantiFluor double stranded DNA System (Promega). Libraries were run in the rapid run flow cell and were paired-end sequenced (2x76bp) on HiSeq 1500 (Illumina, San Diego, CA 92122 USA).

ATAC-sequencing

In order to obtain single cell suspension, preliminarily disrupted tumor sample aliquots corresponding to 50-100 mg of tissue were passed through a syringe needle around 50 times. Mechanical homogenization was followed by centrifugation for 5 min at 4°C at 2400 g. Each pellet was resuspended in 10 ml of cold lysis buffer L1 (50 mM HEPES KOH, pH7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 10% glycerol, 5% NP-40, 0.25% Triton X-100, containing proteinase inhibitor cocktail) and incubated at 4°C for 20 min on a rocking shaker. Then mechanical forcing of tissue disruption was repeated, residual de-

bris were precleared by filtration through an 80-µm nylon mesh filter and eventually lysis buffer was replaced with PBS. Cell suspension was visually controlled under the microscope. Cells were counted automatically with NucleoCounter NC-100 and 50,000 cells were subsequently lysed as previously described (Buenrostro et al., 2013). Then transposition reaction was performed using Nextera DNA Library Preparation kit by Illumina, accordingly to the Buenrostro protocol. Reactions were cleaned up with Zymo DNA Clean and Concentrator 5 columns. The remainder of the ATAC-seq library preparation was performed as described previously (Buenrostro et al., 2013). Finally, ATAC-seq libraries were visualized on Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) and generated chromatograms were used to estimate DNA concentration. Libraries were run in the rapid run flow cell and were paired-end sequenced (2x76bp) on HiSeq 1500 (Illumina, San Diego, CA 92122 USA).

Chromatin immunoprecipitation (ChIP) on tissue samples

Cell suspensions corresponding to 100-400 mg of tumor tissue were aliquoted and spuned down at 1200 rpm for 10 min at 4°C in a swing bucket centrifuge (Eppendorf Centrifuge 5810R). Pellets were crosslinked in a crosslinking buffer containing 100 mM NaCl, 50mM HEPES, 1 mM EDTA, 0.5 mM EGTA and supplemented with 1% formaldehyde (Sigma-Aldrich) for 15 minutes at room temperature on a rocking shaker. Fixation was stopped by incubation with 0.125 M glycine for 5 min at RT. In order to wash away the excess of formaldehyde, fixed material was twice centrifuged at 1400 g for 10 min at 4°C and then pellets were resuspended in cold PBS supplemented with a protease inhibitor cocktail (Roche). Eventually, pellets were stored at -80°C before starting further procedure. Thawed on ice, pellets were resuspended in 5 ml of cold PBS supplemented with proteinase inhibitor cocktail and further homogenized with insulin syringes prior to obtain single cell suspension (if needed fixed tissue was additionally processed with mechanical homogenizer [PRO Scientific, PRO200] before using syringes). Homogenization step was followed by centrifugation for 5 min at 4°C at 2400 g. Each pellet was resuspended in 10 ml of cold lysis buffer L1 (50 mM HEPES KOH, pH7.5, 140 mM NaCl, 1 mM EDTA pH 8.0, 10% glycerol, 5% NP-40, 0.25% Triton X-100, containing a proteinase inhibitor cocktail) and incubated at 4°C for 20 min on a rocking shaker. Subsequently, ice-cold manual glass douncer was used to release the nuclei and the remaining material was collected by centrifugation for 10 min at 4°C at 1700g. L1 buffer was changed for 10 ml of warmed lysis buffer L2 (200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0, containing proteinase inhibitor cocktail) and incubated at room temperature for 20 min on a rocking shaker, then manual glass homogenizer was used once again. This step was followed by centrifugation for 5 min at 4°C at 1700 g. Collected material was resuspended in 0.5 ml of L3 buffer (1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 10 mM Tris pH 8.0, 100 mM NaCl, 0.1% Na-deoxycholate, 0.17 mM N-lauroyl sarcosine, containing protease inhibitors) and sonicated using a Bioruptor Plus Sonicator (Diagenode) for 3 x 15 cycles (30 ON: 30 OFF) set on HIGH conditions. Lengths of chromatin fragments (200-500 bp) were evaluated on agarose gels. Prior to do gel electrophoresis, batches of 10 ul of chromatin were collected by centrifugation at 14, 000 rpm (Eppendorf Centrifuge 5430) and after adding 1xTE buffer up to 300 ul, samples were incubated overnight at 65°C with occasional shaking. After reversal of crosslinking, samples were treated with RNase I at 0.1 mg/ml for 30 min at 37°C, and then treated with proteinase K (final concentration 0.4 mg/ml) for 1.5 h at 55°C. The DNA was purified by phenol–chloroform extraction followed by ethanol precipitation, and recovered in 20 µl of water. Additionally, the efficiency of chromatin isolation was evaluated with the measurement on Quantus Fluorometer (Promega).

DNA-protein complexes were immunoprecipitated with 5 µg of antibody against H3K4me3 (cat. Number 07-473, Merck Millipore), H3K27ac (cat. Number 39133, Active Motif) and H3K27me3 (cat. number ab192985, Abcam). In order to estimate specificity and purity of ChIP reaction, immunoprecipitation with normal IgG was performed simultaneously (cat. Number 2729S, Cell Signalling). Generally, 30 µg of fragmented chromatin was added per ChIP reaction. Depending on antibody type, different ChIP protocols were applied. Anti-H3K4me3 immunoprecipitation and input sample processing were carried out accordingly to the protocol provided with ChIP-IT Express Immunoprecipitation Kit (Cat. Number 53008, Active Motif) with some modifications: i) the volume of IP reaction was equal 0.5 or 1 ml, depending on chromatin concentration in L3 buffer, ii) each washing step of magnetic beads was repeated twice, iii) elution lasted 30 min. Decrosslink and DNA purification was performed as for sonication test. Anti-H3K27ac and anti-H3K27me3 immunoprecipitation was performed accordingly to the RoadmapEpigenomics protocol for ChIP-seq on tissues (**http://www.roadmapepigenomics.org/protocols/type/experimental/**).Briefly, 30 µg of pre-cleared chromatin was incubated for 4 h at 4°C with A- and G-Sepharose beads Cat. number 16-156 & 16-266, Millipore) mix with pre-bound appropriate antibodies. The beads were washed once with RIPA-150 buffer, twice with RIPA-500, twice with RIPA-LiCl and twice with TE buffer, and then bound chromatin was eluted with freshly made elution buffer. After decrosslink and DNA purification, acquired DNA was measured with Quantus fluorometer using **Agilent 2100 Bioanalyzer and High Sensitivity DNA kit (Agilent Technologies, Ltd)**. Before DNA library preparation, an immunoprecipitated material was evaluated by qPCR with specific primers designed to amplify the *GAPDH* promoter region (active chromatin marks) or the *HOXA7* gene body region (repressive chromatin mark) using SYBR Green chemistry (Cat. Number 4385612, Applied Biosystem by Thermo Fisher Scientific) and QuantStudio 12K Flex Real-Time PCR System (data not shown).

ChIP sequencing

DNA libraries for chromatin immunoprecipitation with respective antibodies were prepared using QIAseq Ultra Low Input Library Kit (QIAGEN, Hilden, Germany). Briefly, DNA was end-repaired, adenosines were added to the 3′ ends of dsDNA and adapters were ligated (adapters from NEB, Ipswich, MA, USA). Following the adapter ligation, uracil was digested by USER enzyme from NEB (Ipswich, MA, USA) in a loop structure of adapter. Adapters containing DNA fragments were amplified by PCR using NEB starters (Ipswich MA, USA). Library quality evaluation was done with Agilent 2100 Bioanalyzer using the Agilent DNA High Sensitivity chip (Agilent Technologies, Ltd.) Quantification and quality evaluation of obtained samples were done using Nanodrop spectrophotometer (Thermo Scientific, NanoDrop products, Wilmington, USA), Quantus fluorometer (Promega Corporation, Madison, USA) and 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). Mean library size was 300 bp. Libraries were run in the rapid run flow cell and were single-end sequenced (65bp) on HiSeq 1500 (Illumina, San Diego, CA 92122 USA).

ChIP-qPCR on cultured glioma cells

Primary glioma WG4 cell cultures were harvested at 90% confluency using standard trypsinization protocol and $1x10^7$ cells were collected by centrifugation (1200 rpm, 10 min, 4°C, Eppendorf Centrifuge 5810R), and fixed in 1% formaldehyde (ThermoFisher Scientific) diluted in PBS for 15 minutes at room temperature on a rocking shaker. Formaldehyde was quenched by adding 0.125 M glycine and incubating samples for 5 min at RT. Fixed cells were centrifuged at 1400 g for 10 min at 4 °C and

washed twice with PBS as described. Cell pellets were resuspended in 10 ml of cold L1 buffer and incubated at 4°C for 20 min on a rocking shaker to force cell swelling. Subsequently, ice-cold manual glass douncer was used to aid nuclei release which were collected by centrifugation (10 min, 4°C, 1200 rpm, Eppendorf Centrifuge 5810R) in order to reduce sample volume. After centrifugation, cell pellets were resuspended in 1 ml of buffer, transferred to Eppendorf tubes and pushed through insulin syringe needle several times. Nuclei were centrifugated (10 min at 4°C at 1200 rpm , Eppendorf Centrifuge 5430), and reconstituted in 50 µl SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH=8.0) and incubated on ice for 30 min. Nuclei release was controlled under the microscope and 300 µl of 1x TE were added to dilute SDS. Chromatin fragmentation was performed in Eppendorf tubes using Bioruptor Plus Sonicator (Diagenode) with the following program: 3 x 13 cycles (30 ON: 30 OFF), HIGH conditions. Sonication efficacy was examined using agarose gel electrophoresis as described above in *ChIP on tissue samples* section. Chromatin isolated from $1x10^7$ cells was used per one immunoprecipitation reaction either with 5 µg FOXM1 (cat. no. C15410232-100, Diagenode) or normal rabbit IgG antibody (cat. no. PP64B, Merck Millipore). Immunoprecipitated complexes were captured on Dynabeads Protein A (cat. no. 10002D, ThermoFisher Scientific) accordingly to the RoadmapEpigenomics protocol mentioned before (Dynabeads part) with modified elution step performed as follows: 125 µl elution buffer was added to washed beads and resuspended beads were kept in 65°C for 10 min, this elution step was repeated and eventually eluate was pooled. Decrosslink, RNase, proteinase K treatments and DNA purification was performed as described in the *ChIP on tissue samples* section. Real-time PCR amplifications of the studied *ANXA2R* regulatory regions were performed applying SYBR Green chemistry (Cat. Number 4385612, Applied Biosystem by Thermo Fisher Scientific) on QuantStudio 12K Flex Real-Time PCR System device using primers indicated in Key Sources Table. The *CCNB1* gene promoter served as a positive control for FOXM1 binding. Results were calculated as % of input, mean ± SD (n = 3) and compared to IgG values. T-test was used for evaluation fo sgnificance.

DNA methylation sequencing

DNA samples) were bisulfite-converted using EZ DNA Methylation-Lightning Kit (Zymo Research, Irvine, CA, USA). Probes from SeqCap Epi CpGiant Enrichment Kit (Hoffmann-La Roche, Basel, Switzerland) were used to enrich each Bisulfite-Converted Sample Library in the predetermined various genome regions of >80.4 Mb capture size comprising >5.6 million of CpG sites on both DNA strands. The libraries were prepared according to the Hoffmann-La Roche's "NimbleGen SeqCap Epi Library Workshop Protocol, v1.0" and "SeqCap Epi Enrichment System User's Guide, v1.2". Briefly, the concentration of genomic DNA was measured using a Quantus fluorometer with QuantiFluor dsDNA System (Promega, Madison, WI, USA) and 1 µg of input DNA together with 165 pg of Bisulfite-Conversion Control (viral unmethylated gDNA; SeqCap Epi Accessory Kit; Hoffmann-La Roche) were fragmented using Focused-ultrasonicator Covaris M220 (Covaris, Inc., Woburn, MA, USA) to an average size of 200 ± 20 bp. DNA fragments were checked using the High Sensitivity DNA Kit on a 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA). Next, DNA fragments were "End-Repaired", "A-Tailing" was performed and Index Adapters ligated using KAPA LTP Library Preparation Kit (KAPA Biosystems, Wilmington, USA), SeqCap Adapter Kit A and B (Hoffmann-La Roche) and DNA purification beads (Agencourt AMPure XP Beads; SeqCap EZ Pure Capture Bead Kit; Hoffmann-La Roche). Then, DNA fragments, enlarged by adapters, were size selected Agencourt AMPure XP Beads (SeqCap EZ Pure Capture Bead Kit) using Solid Phase Reversible Immobilization technology to discard DNA fragments larger than ~450 and smaller than

~250 bp. Next, the libraries were bisulfite-converted using EZ DNA Methylation-Lightning Kit (Zymo Research) and amplified by Pre-Capture Ligation Mediated PCR (LM-PCR). After cleaning on Agencourt AMPure XP Beads (SeqCap EZ Pure Capture Bead Kit), quality and concentrations of Amplified Bisulfite-Converted Sample Libraries were determined using NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and Quantus with QuantiFluor dsDNA System (Promega), respectively. Also a size of DNA fragments was analyzed using the High Sensitivity DNA Kit on a 2100 Bioanalyzer (Agilent Technologies, Inc.). Next, 1 µg of each Amplified Bisulfite-Converted Sample Library was hybridized (47 °C, 67 ± 2 h) with probes from SeqCap Epi CpGiant Enrichment Kit (Hoffmann-La Roche), bound to the Capture Beads (SeqCap EZ Pure Capture Bead Kit; Hoffmann-La Roche) and sequentially washed out of contamination and unspecific DNA in buffers of SeqCap Hybridization and Wash Kit (Hoffmann-La Roche). Finally, the Captured Bisulfite-Converted Sample Libraries were amplified in Post-Capture LM-PCR, cleaned up using Agencourt AMPure XP Beads (SeqCap EZ Pure Capture Bead Kit) and the Amplified Captured Bisulfite-Converted Sample Libraries were submitted to the last quality check where the quality and the concentrations of the final libraries were determined using NanoDrop (Thermo Fisher Scientific) and Quantus with QuantiFluor dsDNA System (Promega), respectively. A size of the obtained DNA fragments was also analyzed using the High Sensitivity DNA Kit on a 2100 Bioanalyzer (Agilent Technologies, Inc.). Libraries were run in the rapid run flow cell and were paired-end sequenced (2x76bp) on HiSeq 1500 (Illumina, San Diego, CA 92122 USA).

DNA methylation bead chip analysis

Genomic DNA was extracted using the QIAsymphony DNA Midi kit (Qiagen, Crawley, UK). Bisulphite conversion of 500 ng of each sample was performed using the EZ-96 DNA Methylation-Gold kit (Zymo Research, Orange, CA, USA). Bisulphite-converted DNA was used for hybridization on the Infinium HumanMethylation 450 BeadChip (Bibikova et al., 2011)[according to manufacturer's instructions. Illumina GenomeStudio software was used to extract the raw signal intensities of each probe (without background correction or normalization).

Quantitative PCR (qRT-PCR)

Total RNA was extracted from glioma cells using the RNeasy Mini kit (Qiagen, Hilden, Germany) and purified using RNeasy columns accordingly to the manufacturer's instructions. cDNA was synthetized by extension of oligo(dT) primers with SuperScript III Reverse Transcriptase (Invitrogen, USA). Real-time PCR was performed applying SYBR Green chemistry (Cat. Number 4385612, Applied Biosystem by Thermo Fisher Scientific) on QuantStudio 12K Flex Real-Time PCR System device using primers indicated in Key Sources Table. Amplified product was normalized to the endogenous expression of *GAPDH* and represented as minus delta delta Ct values. P values were considered significant when *P < 0.05 and **P < 0.01 (t-test).

Capture-C assay

WG4 cells were cultured, harvested and crosslinked following the same protocol as described in the ChIP-qPCR section. Cell pellet was resuspended in 500 µL of freshly prepared ice-cold lysis buffer (10 mM Tris-HCl, pH 8, 10 mM NaCl, 0.2% Igepal NP-40 (Sigma), 1x protease inhibitor cocktail (Roche)), and incubated on ice for 30 min and followed by centrifugation to pellet nuclei (5 min, 1200 rpm, 4 °C, Eppendorf Centrifuge 5430). Cells were washed once with 300 ul 1x NEB2 buffer, and nuclei were extracted by 40 min incubation at 37°C in 190 µl 0.5%SDS 1xNEB2 buffer. In order to

quench SDS, samples were transferred on ice and 400 µL of 1X NEBuffer2 and 120 µL of 10% Triton X-100 were added (in the mentioned order), and kept at 37°C for 15 min. After centrifugation and washing with 1X NEBuffer2, the digestion reaction with 400U of MboI was performed overnight in 37°C (300 µl final volume) in termomixer with shake 800 RPM. The next day, nuclei were pelleted again, reconstituted in 200 µl 1xNEB2 buffer and additional 200 U of MboI were added for two more hours before heat inactivation (65°C, 15 min). After centrifugation, supernatant was discarded and obtained pellets were resuspended in 1.2 ml of ligation mix consisting of 1x T4 DNA Ligase buffer, 1% Triton X-100 and 0.1 mg/ml BSA. Eventually 5 µL of 2000 U/µL T4 DNA Ligase (NEB) was added and incubated at 16ºC for 5h to perform slowligation. Then samples wers centrifuged and nuclei were resuspended in 200 µl 1X NEBuffer2. Decrosslink was performed by overnight incubation at 65 ºC. The next day, we performed RNAse A (10 µL of 10 mg/mL, 30 min, 37°C) and Proteinase K treatment (20 µL of 20 mg/mL,1.5 h, 55ºC) followed by standard fenol:chloroform extraction to purify DNA. At the same time non-digested and non-ligated controls were processed, as 10 µl of reaction at the crucial steps were taken to check digestion and ligation efficiency by running DNA on 0.8% agarose gel. Properly digested and re-ligated samples were subsequently used for library preparation. First, DNA was fragmented using Diagenode Bioraptor Pico device (cat. No. B01060010). Aliquots of 100 µl DNA (10 µg/µl) were distributed to 6 Eppendorf 1.5 ml tubes, , and the following conditions: 20"ON/60"OFF, 10 cycles were applied program. Sonication efficiency was tested by loading 50 ng DNA on the 1.2 % agarose gel. After AMPure beads (Beckman Coulter) purification 1 µg of fragmented DNAs were used for sequencing library generation using NEBNext Ultra Library Prep Kit for Illumina (cat. No. E7370S/L) following a manufacturer's protocol. The library was hybridized to probes capturing the *ANXA2R* promoter fragment. Probe design excluded repeating elements present in the *ANXA2R* promoter region. In order to capture both strands of a selected restriction fragment, biotinylated olignucleotides complementary to 5' end of each strand were used (IDT). Lyophilized DNA oligonucleotides were reconstituted to a concentration of 2.9 µM, mixed and diluted in equimolar quantities to final 2.9 nM concentration. Subsequently, Hybridization Capture of DNA libraries using xGen Lockdown Probes and Reagents kit (IDT) was used following a manufacturer's instruction with an exception of using NEBNext Ultra Library Prep Kit for Illumina (cat. No. E7370S/L) instead of recommended libarary preparation kit. Final library was sequenced using MiSeq MICRO 2x151 bps 300 cyles PE Cycle Paired-end sequencing (4M reads).

Quality of fastq files was assessed using FastQC tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed with FASTQ Trimmer from 3' end to reach a read length of 81 base pairs. Reads were aligned to human hg38 genome using qAlign function from Bioconductor QuasR package with maxHits parameter set to 1 to obtain uniquely mapped reads. At this point fastq files were treated as files obtained out of single-end sequencing experiments. Generated BAM files were sorted with Picard, and subsequently same tool (MarkDuplicates option) was used to remove PCR duplicates. We calculated the number of reads mapped to MboI restriction fragments qAlign function (QuasR package). Briefly, new *qProject* object was created out of sorted and deduplicated BAM files, a list of restriction fragments for chromosome 5 generated with getRestrictionFragmentsPerChromosome function (HiTC package) were transformed into *query* object and mask parameter was set to exclude captured restriction fragment applying vmatchPattern function from Biostrings package. At this point read counts for each file were merged and normalized to „per million reads". Results were visualized with visualizeViewpoint function from Basic4Cseq package.

Data analysis

Fastq preprocessing. The quality of raw fastq files was assessed using FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and based on the results the following parameters of preprocessing were chosen: i) for RNA-seq reads were trimmed with FASTQ Trimmer (Blankenberg et al., 2010) (offset from 5' end: 11, offset from 3 ' end: 3), overrepresented sequences were removed using clip tool from FASTX package (http://hannonlab.cshl.edu/fastx_toolkit/) and reads which pair mate was filtered out were removed; ii) for ChIP-seq reads were trimmed with FASTQ Trimmer (offset from 5' end: 5 for samples GB07, GB10, 3 for the rest; offset from 3' end: 3), Filter FASTQ (Blankenberg et al., 2010) was used to keep only reads with quality above specified threshold (min quality 10, max number of bases below this quality: 6), for patients DA05 and GB03, additionally we run Trimmomatic (Bolger et al., 2014) with option “sliding window 4”, finally the reads were filtered by length using Filter FASTQ (min. length: 20); iii) for ATAC-seq reads were trimmed with FASTQ Trimmer (offset from 5' end: 14; offset from 3' end: 3), Filter FASTQ was used to keep only reads with quality above specified threshold (min quality 10, max number of bases below this quality: 5), the reads were filtered by length using Filter FASTQ (min. length: 20) and reads which pair mate was filtered out were removed. All fastq preprocessing steps were performed using Galaxy wrappers (Afgan et al., 2016).

Mapping, peak calling, read counts. ChIP-seq and ATAC-seq reads were mapped to hg38 using Bowtie 2.2.6.2 (Langmead and Steven L Salzberg, 2013) with default parameters. RNA-seq reads were mapped to hg38 transcriptome using Tophat2 (Kim et al., 2013), with default parameters, except for the following: library type=fr-firststrand, mate-std-dev=100, mate--inner-dist=300. For ChIP-seq and ATAC-seq duplicates were removed using samtools rmdup command (Li et al., 2009). ChIP-seq peaks were called using MACS2.1 (Zhang et al., 2008), with default parameters. For ATAC-seq, F-seq1.85 (Boyle et al., 2008) with default parameters was used, except for f=50. RNA-seq counts on genes and transcripts were calculated using samtools. Mapping and peak-calling with MACS were performed using Galaxy wrappers.

Defining active promoters and putative enhancers. Promoters were defined as region +/-1000 bp around every TSS. We presumed that a promoter has a H3K4me3 mark in a patient A if it overlaps with any H3K4me3 peak called in a patient A; the samea gene has a H3K4me3 mark if any of its promoters has a H3K4me3 mark. Similarly presumptions were made for H3K27ac marks and ATAC-seq signal. Overlaps were found using bedtools intersect tool (Quinlan and Hall, 2010). For every patient non-promoter H3K27ac peaks were defined as peaks that have no overlap with regions +/-1000 bp from any TSS. Putative enhancers were defined by merging all non-promoter peaks identified for each patient (individual peak files were merged into a single file and overlapping peaks were merged into one peak using bedtools merge tool). Each putative enhancer was denoted as active in a patient A if it overlapped with any non-promoter H3K27ac peak present in a patient A.

Common and variable marks. Active genes, promoters and enhancers were divided into two groups: common, which are present in at least threshold patients, and variable, which are present in less than threshold patients. Threshold value was chosen separately for every mark/type of genomic element. For promoters and genes with H3K4me3 mark, threshold = 20. For promoters and genes with H3K27ac mark, threshold = 10. For enhancers with H3K27ac mark, threshold = 4.

Coverage. Coverages over whole genome were calculated using bedtools genomecov tool. Coverages over specific intervals were calculated using bedtools intersect tool and custom python script calculate_coverage_over_features.py. Briefly, for every interval sum of coverages over all

positions were calculated; then the sum was divided by length of the interval to obtain mean coverage value. Normalized coverages were obtained by dividing every coverage value by mean coverage defined as summaric coverage (the sum of values on every position of the genome, including ones with no coverage) divided by the length of the genome.

Evolutionary conservation analysis. A collection of random genomic intervals was generated for both common active promoters and common active enhancers using bedtools shuffle. PhastCons 100-Way scores were obtained from http://hgdownload.cse.ucsc.edu/goldenPath/hg38/phastCons100way/hg38.phastCons100way.bw Mean PhastCons scores for all analyzed intervals were calculated with bigWigAverageOverBed from the kentUtils package (https://github.com/ENCODE-DCC/kentUtils). One-sided Mann-Whitney U test was calculated to assess significance of difference in conservation scores between identified regulatory elements and random genomic intervals.

Identification of enhancer-promoter contacts based on Hi-C data. Hi-C data for fetal brain samples from (Won et al., 2016) were used to infer long-range chromatin interactions for putative enhancers defined based on the presence of non-promoter H3K27ac peaks. First, coordinates of putative enhancers were converted from hg38 to hg19 using UCSC liftover tool, to match the Hi-C data. Then the HiCEnterprise method (https://github.com/hansiu/HiCEnterprise) was used to identify contacts within 2 Mb distance with 10 kb resolution (Kranas et al., 2019). Coordinates of identified contacts were converted back to hg38.

Correlation between histone marks on promoters and transcript expression. H3K4me3 and H3K27ac ChIP-seq coverages on promoter regions and RNA-seq counts on transcripts were normalized with quantile normalization. Only transcripts of protein-coding genes were considered. For genes with multiple alternative transcripts, the transcript with a highest ChIP-seq coverage on promoter was chosen. Spearman correlation between ChIP-seq coverage and transcript expression was calculated for all transcript-promoter pairs in all patients. For violin plot visualization, transcripts were divided into quantiles based on expression level (quantile normalized counts).

Correlation between H3K27ac level on enhancers and transcript expression. For each enhancer identified by the presence of H3K27ac mark, the closest transcript was found using bedtools closest. The enhancers were subsequently divided into three groups: i) having the closest transcript within less than 20 kilobases, ii) having closest transcript within 20 to 200 kb, iii) having closest transcript located more than 200 kb away. Spearman correlation between H3K27ac coverage and transcript expression was calculated for each of these groups of enhancers. Then, enhancers assigned to each group were checked for contacts with promoters, predicted from Hi-C data, and Spearman correlation was again calculated for enhancer-transcript pairs defined with Hi-C. In cases where one enhancer had predicted contacts with more than one promoter, mean of counts for transcripts associated with all contacting promoters was used in the correlation calculation. P-value for the difference between correlation obtained for closest enhancer-transcript pairs and pairs defined with Hi-C data was estimated by random sampling 260 enhancer-transcript pairs from the set of 3371 enhancer-transcript pairs with closest transcript located further than 200 kb away for 1000 times.

From all enhancer-promoter pairs predicted based on Hi-C data pairs with significant correlation between H3K27ac coverage and transcript expression were chosen based on the following criteria:

FDR < 0.1, Spearman R > 0.7, minimal read counts on transcript in a sample with highest expression = 20. Amongst analyzed pairs, 109 enhancer-promoter pairs fulfilled these criteria.

Clustering based on gene expression. RNA-seq row counts on genes of 37 cases (4 NB, 11 PA, 7DA, 14GB, 1 PG) were plotted for each tumor grade and hospital of a sample origin. No batch effect related to grade or hospital was detected(Risso et al., 2011). Two normalization procedures were considered following recommendations (Risso et al., 2011): (i) within-lane to adjust for GC-content and gene-length; (ii) between-lane, both implemented in EDASeq 3.8 R package (Risso et al., 2015). Genes with mean row counts across all samples below 10 were filtered out, and then the EDASeq 3.8 protocol was applied with full-quantile normalization method for within- and between-lane normalization. The filtering step resulted in 19,328 genes that were used for clustering. Dendextend R package (Galili, 2015) was used to perform hierarchical clustering with ward.D2 method with distance set to Euclidean. ssGSEA tool (gene-centric single sample Gene Set Enrichment Analysis of gene expression data) was used to define GBM subtypes (downloaded from https://github.com/broadinstitute/ssGSEA2.0) Input data for the analysis consisted 37 samples and 56,819 RPKM normalized gene expression values, converted to GCT file format. Additionally, four reference files (for each subtype separately) have been downloaded from Molecular Signatures Database v7.0 (http://software.broadinstitute.org/gsea/msigdb/index.jsp) and used in the analysis. To predict the subtype of Pan Glioma RNA Expression Cluster (LGr1, LGr2, LGr3, LGr4), a validation cohort from published studies, including 665 adult and pediatric gliomas (Ceccarelli et al., 2016) and 21,022 of their gene expressions levels were compiled. Next on that data we trained three classifiers (Naïve Bayes, SVM and Random Forest) and applied them as a committee of classifiers with voting mechanism on our 26 samples to predict the target variable (Pan Glioma Cluster).
Gene expression differential analyses were computed with edgeR Bioconductor package. For visualization and co-activation analyses FPKM (Fragments per Kilobase Million) values were calculated.

Comparison of ChIP-seq and ATAC-seq profiles in TSS-proximal regions. Genomic positions with the numbers of mapped reads above the significance threshold of z-score = 7 were identified as anomalous, and the tags mapped to such positions were discarded. Read frequencies were computed in 500-bp non-overlapping bins for each sample independently and normalized by the corresponding library sizes to represent values per one million of mapped reads. For the histone marks abundance was computed per bin as a difference between normalized frequencies obtained from chip and input experiments. For each TSS proximal region, the abundance values were computed as an average over abundance in bins overlapping the corresponding 4 kb region (TSS+/-2kb). The transcription start site proximal regions were defined as +/-2kb from the TSS. TSS proximal regions overlapping with other genes were excluded from this analysis. The averaged and detailed profiles around TSS were computed using 10 bp non-overlapping bins and smoothed using runmean function from caTools package with k=15.

Clustering based on H3K4me3 mark. For each sample a set of promoters with H3K4me3 mark was defined based on a presence of H3K4me3 ChIP-seq peak as described above. Pairwise similarities between samples were defined as Jaccard indices calculated for sets of promoters with H3K4me3 mark. These pairwise similarities were subsequently used to perform hierarchical clustering of

samples with scipy.cluster.hierarchy.linkage function and the UPGMA algorithm (method = “average”).

Identification of promoters specifically active/inactive in the PAs cluster. To identify promoters whose activity differentiate PA from higher grade glioma, for each annotated promoter we compared the number of samples from the PA cluster in which it was active (marked with H3K4me3) to the number of samples from a group of 13 higher grade samples (DA01, DA03, DA04, DA05, DA06, GB03, GB04, GB05, GB06, GB07, GB08, GB09, GB10) in which it was active. As specifically active in PA, we chose those active in at least four out of five samples from the PA cluster and in at most two higher grade samples and additionally having p-value from Fisher’s exact test below 0.003. Similarly, promoters specifically active in higher grades were defined as those active in at most one PA sample and at least 11 higher grades samples, and having the p-value below 0.003.

Enrichment analysis for DNA-binding proteins binding sites. The ENCODE Regulation 'Txn Factor' track data was downloaded from http://hgdownload.soe.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeRegTfbsClustered/ promoters specifically active in PAs, active in all samples or inactive in any sample were intersected with bed files defining binding sites of 161 proteins using bedtools intersect. Enrichments in binding sites were calculated as a fold change of the number of intersections between each protein binding sites and promoters specifically active in PAs and number of intersections for the same protein binding sites and active or inactive promoters (per megabase).

Protein binding motif analysis. To identify transcription factors that can potentially bind to the identified promoters with H3K27ac hyperacetylation or H3K4me3 hypermethylation in PA, we used HOCOMOCO-v11 motif database (Kulakovskiy et al., 2018). For hyperacetylated promoters the relative motif enrichment was computed with AME tool from MEME-Suite software (v5.0.4) [10.1186/1471-2105-11-165] and logos were obtained from the HOCOMOCO website. For hypermethylated promoters, motif enrichment was computed with PWMEnrich R package (Stojnic and Diez, 2013) using sequences of promoters inactive in all samples as background. The returned protein binding motifs were accepted if their p-value was below 0.001. To establish the similarity of motifs enriched for grade-specific groups (PAs vs. higher grade gliomas - HGG) on each of the backgrounds, the Jaccard index (Jaccard, 1901) implemented in ‘jaccard’ CRAN R package (Christopher Chung et al.) was used.

Analysis of enriched Gene Ontology terms. Enrichment of gene ontology terms for promoters with H3K4me3 hypermethylation in PAs and EZH2 binding sites was performed using the PANTHER Overrepresentation Test (Released 20171205) (Ashburner et al., 2000; Carbon et al., 2017) and Fisher’s Exact with FDR multiple test correction. For promoters with H3K27ac hyperacetylation in PAs, gene ontology analysis was performed with GO.db Bioconductor package and 0.05 significance threshold on Bonferroni corrected P-value (Fisher’s exact test) was used.

DNA methylation data processing. For DNA methylation, assessment of methylation sites was restricted to regions covered by SeqCap Epi CpGiant Methylation panel which captures 80.5 Mb (Roche). The methylation analysis workflow consisted of: read quality assessment, mapping to hg38 genome, removal of PCR duplicates, coverage statistics assessment and methylation level assignment. The software required to perform the aforementioned steps: FastQC, BSMAP (Xi and Li, 2009), Picard (http://broadinstitute.github.io/picard/) were selected based on Roche recommended

pipeline (NimbleGen, 2014) and compiled into a single CytoMeth tool (https://github.com/mdraminski/CytoMeth) enabling fast and transparent rough data processing . The following quality requirements were set: minimal bisulphite conversion 99.2% and at least 3, 000, 000 cytosines in a CpG context covered with at least ten reads. Out of 26 samples, 20 passed these requirements and were used to assess DNA methylation variability within promoter regions (+/-2000 bp from TSS) with a use of methylKit (Akalin et al., 2012).

For Infinium HumanMethylation 450 BeadChip data signal intensities were imported into R using the methylumi package [10.1093/bib/bbt054]. Initial quality control checks were performed using functions in the methylumi package. The methylation score for each CpG was represented as a β-value according to the fluorescent intensity ratio representing any value between 0 (unmethylated) and 1 (completely methylated). Mapping of CpGs was performed based on UCSC Genome Browser annotation data (Haeussler et al., 2019). CpGs beta values in the range from 0 to 0.2 were defined as hypomethylated and those from 0.8 to 1 as hypermethylated.

DNA methylation analysis for PAs specific promoters with EZH2 binding sites. Signal intensities for probes from HumanMethylation450 BeadChip array, which were located within PA specific promoters with EZH2 binding sites, were compared for 11 independent PAs and 13 GB samples. DNA methylation data from previously pubslihed studies (Lambert et al., 2013), and gene expression data from the altter were downloaded from Gene Expression Omnibus (GEO: http://www.ncbi.nlm.nih.gov/geo/) using the following accession numbers: GSE61431, GSE44684 and GSE44971, respectively. Quantile normalization was performed on signal intensities. For the expression data median intensities were calculated for each probe in each analyzed group of samples.

TCGA data analyses. The expression values for the TCGA samples were downloaded from TCGA website (RNASeqV2 set). Survival analyses were performed with TCGA data. The patients were stratified into three equinumerous sub-groups according to *ANXA2R* or *FOXM1* expression levels. Correlation between the expression level of the selected genes and patient survival times was analyzed using Kaplan-Meier model and log-rank test.

REFERENCES

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., et al. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. *44*, W3–W10.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. *13*.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: Tool for the unification of biology. Nat. Genet. *25*, 25–29.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., and Team, G. (2010). Manipulation of FASTQ data with galaxy. Bioinformatics *26*, 1783–1785.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Boyle, A.P., Guinney, J., Crawford, G.E., and Furey, T.S. (2008). F-Seq: A feature density estimator for high-throughput sequence tags. Bioinformatics *24*, 2537–2538.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Carbon, S., Dietze, H., Lewis, S.E., Mungall, C.J., Munoz-Torres, M.C., Basu, S., Chisholm, R.L., Dodson, R.J., Fey, P., Thomas, P.D., et al. (2017). Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res. *45*, D331–D338.

Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., et al. (2016). Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. Cell *164*, 550–563.

Christopher Chung, N., Miasojedow, B., and Gambin, A. Pattern Recognition of Chemical Fingerprints from Large-Scale Databases of Molecules and Reactions.

Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics *31*, 3718–3720.

Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., et al. (2019). The UCSC Genome Browser database: 2019 update. Nucleic Acids Res. *47*, D853–D858.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et du Jura. Bull. La Société Vaudoise Des Sci. Nat. *37*, 547–579.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*.

Kranas H, Tuszyńska I, Wilczynski B. 2019. HiCEnterprise: Identifying long range chromosomal contacts in HiC data. PeerJ Preprints 7:e27753v1 https://doi.org/10.7287/peerj.preprints.27753v1

Kulakovskiy, I. V., Vorontsov, I.E., Yevshin, I.S., Sharipov, R.N., Fedorova, A.D., Rumynskiy, E.I., Medvedeva, Y.A., Magana-Mora, A., Bajic, V.B., Papatsenko, D.A., et al. (2018). HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res. *46*, D252–D259.

Lambert, S.R., Witt, H., Hovestadt, V., Zucknick, M., Kool, M., Pearson, D.M., Korshunov, A., Ryzhova, M., Ichimura, K., Jabado, N., et al. (2013). Differential expression and methylation of brain developmental genes define location-specific subsets of pilocytic astrocytoma. Acta Neuropathol. *126*, 291–301.

Langmead and Steven L Salzberg (2013). Bowtie2. Nat. Methods *9*, 357–359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. *45*, D183–D189.

NimbleGen, R. (2014). How To... Evaluate NimbleGen SeqCap Epi Target Enrichment Data. 1–22.

Pidsley, R., Viana, J., Hannon, E., Spiers, H., Troakes, C., Al-Saraj, S., Mechawar, N., Turecki, G., Schalkwyk, L.C., Bray, N.J., et al. (2014). Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. Genome Biol. *15*, 483.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). GC-Content Normalization for RNA-Seq Data. BMC Bioinformatics *12*.

Risso, D., Eda, R., and Eda, G. (2015). EDASeq : Exploratory Data Analysis and Normalization for. 1–16.

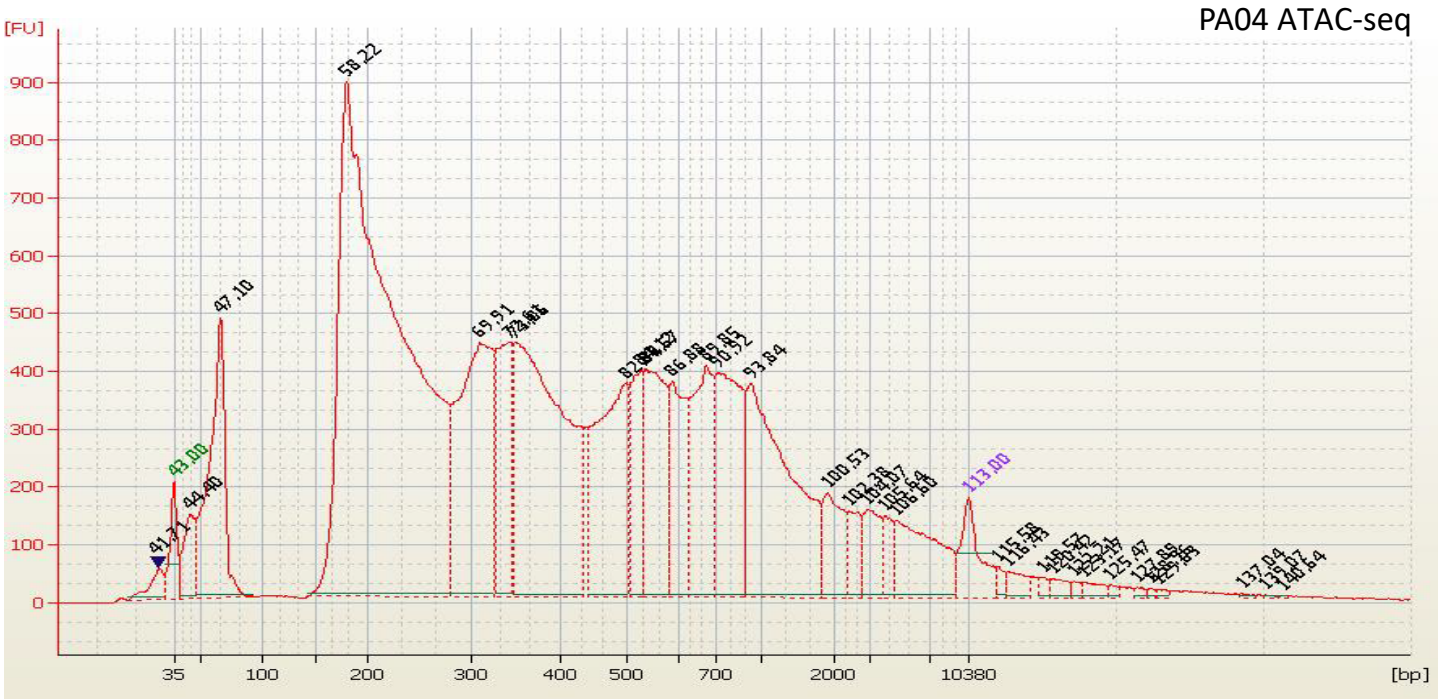
Stojnic, R., and Diez, D. (2013). PWMEnrich: PWM enrichment analysis. R Packag. Version.

Won, H., de la Torre-Ubieta, L., Stein, J.L., Parikshak, N.N., Huang, J., Opland, C.K., Gandal, M., Sutton, G.J., Hormozdiari, F., Lu, D., et al. (2016). Chromosome conformation elucidates regulatory relationships in developing human brain. Nature *538*, 523–527.

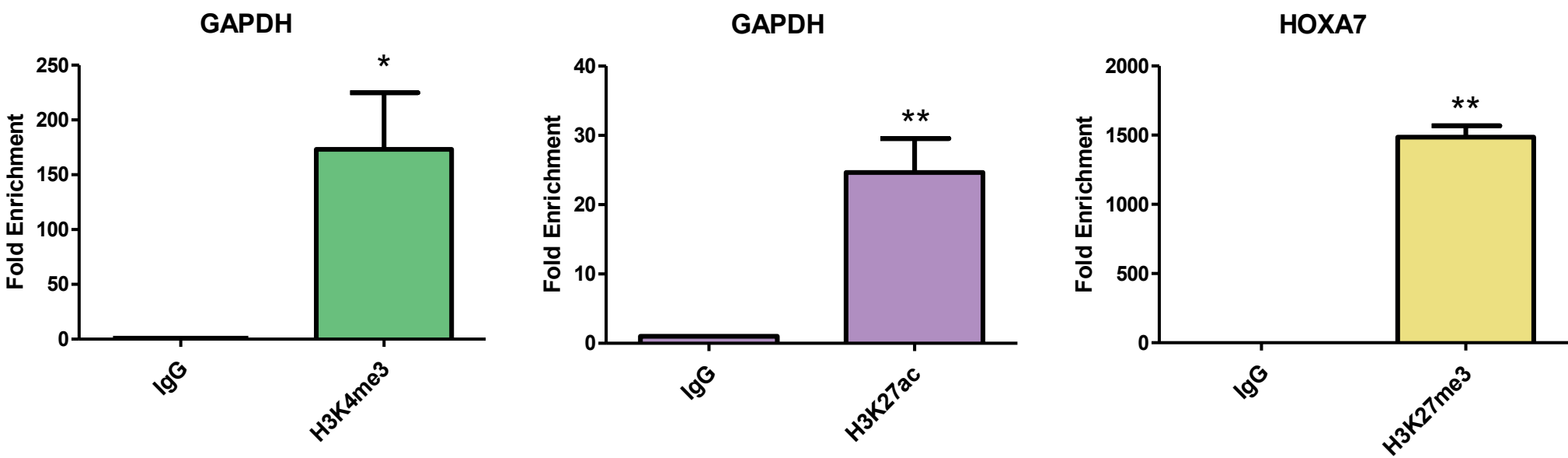
Xi, Y., and Li, W. (2009). BSMAP: Whole genome bisulfite sequence MAPping program. BMC Bioinformatics *10*.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*.

A


PA04 ATAC-seq
[FU]
900
800
700
600
500
400
300
200
100
0
35
100
200
300
400
500
700
2000
10380
[bp]

B


GAPDH
Fold Enrichment
250
200
150
100
50
0
*
IgG
H3K4me3
GAPDH
Fold Enrichment
40
30
20
10
0
**
IgG
H3K27ac
HOXA7
Fold Enrichment
2000
1500
1000
500
0
**
IgG
H3K27me3

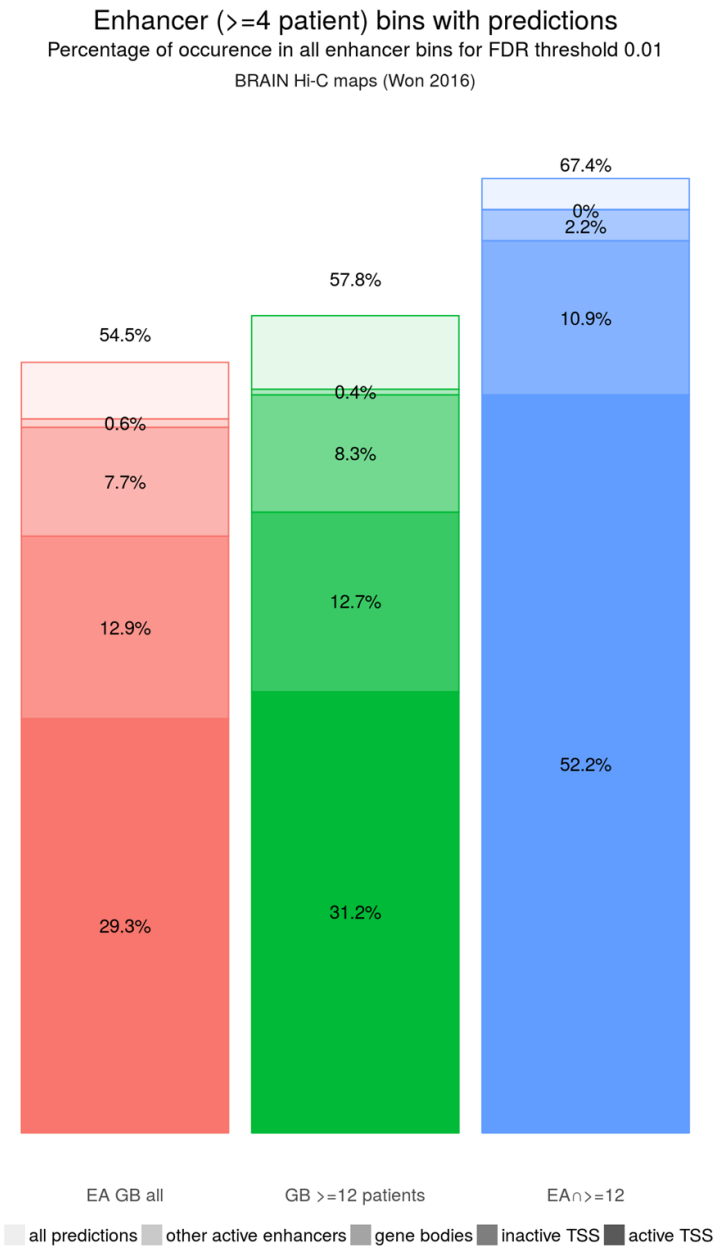
bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure 1

A. A characteristic appearance of a single nucleosome and their multiplications in the material isolated for ATAC-seq separated on DNA Agilent chips (PA04 sample).

B. Validation of immunoprecipitation efficiency with antibodies used in the experiments (for details see Methods). Chromatin immunoprecipitation (ChIP)-quantitative polymerase chain reaction (qPCR) analysis of histone modification enrichment in control regions: GAPDH promoter (active chromatin marks, H3K4me3, H3K27ac) and HOXA7 gene body (repressive mark, H3K27me3). Results are calculated as fold enrichment over negative control (immunoprecipitation with normal IgG) and represented as mean ± SD (H3K4me3 ChIP-qPCR n=6, H3K27ac ChIP-qPCR n=5, H3K27me3 ChIP-qPCR n=3n = 4). Statistical significance **P < 0.05 i **P < 0.01 (t-test).

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

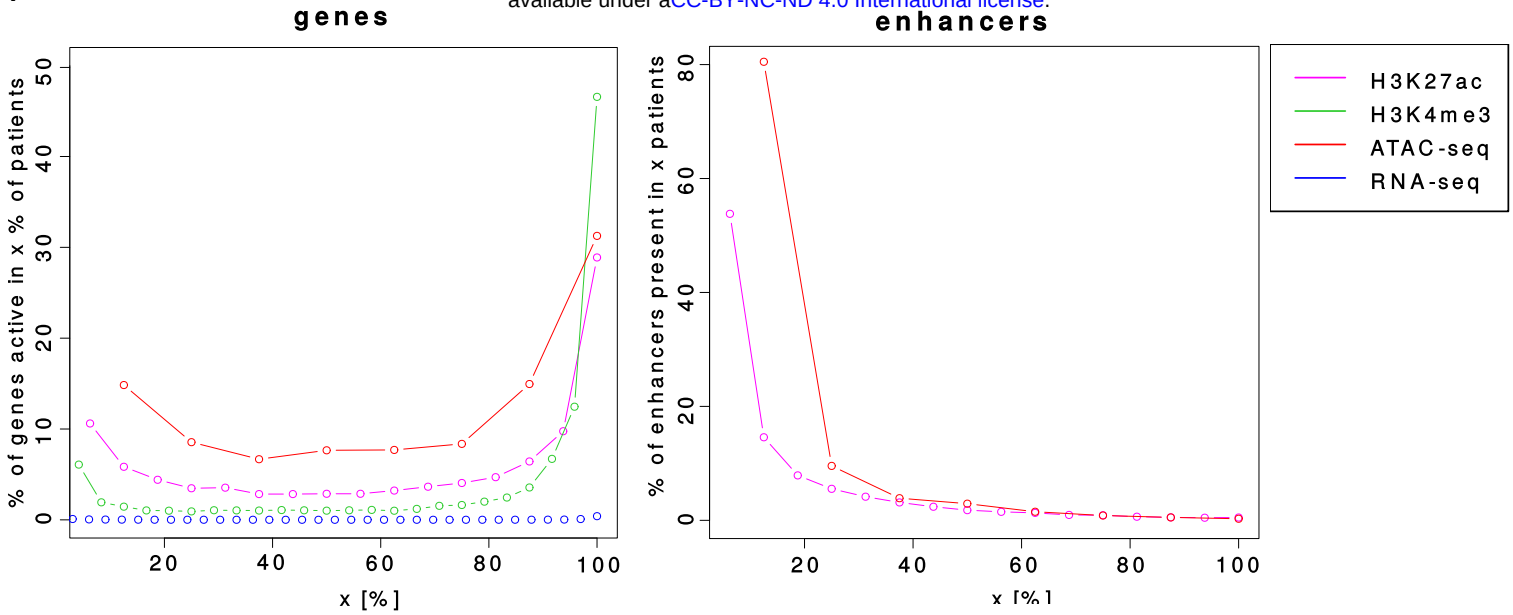


Supplementary Figure 2

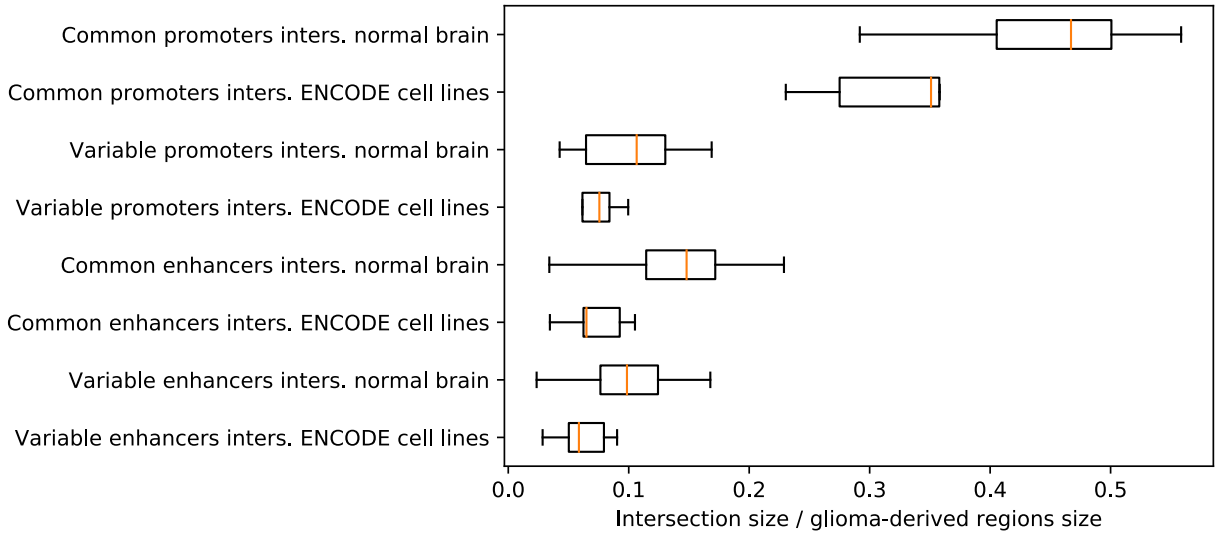
Percentage of enhancer bins predicted to be in contact with different genomic elements. EA GB all - all enhancers reported in EnhancerAtlas (Gao et al. 2016) and in this work in at least 4 samples; GB >= 12 patients - enhancers identified in this work in at least 12 samples; EA and >=12 - enhancers present both in EnhancerAtlas and at least 12 samples in this work.

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

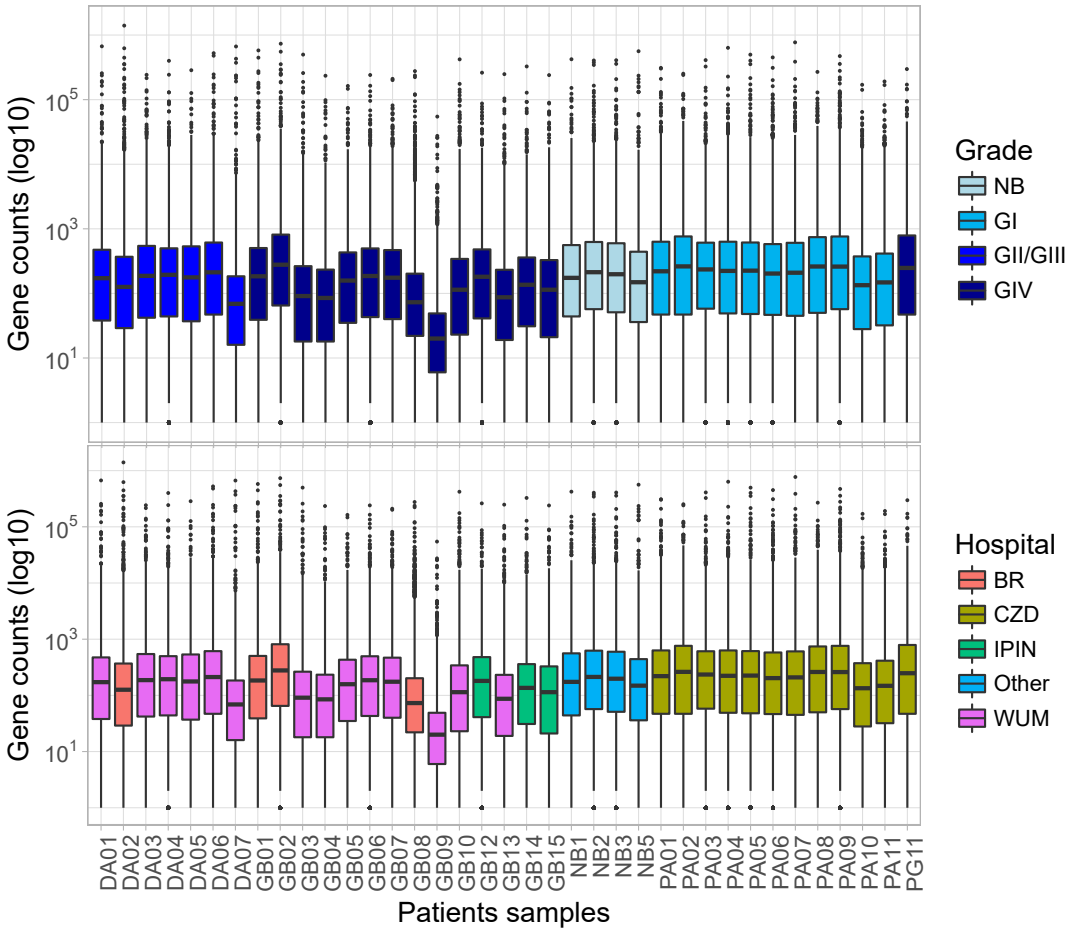
A



B



C



bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

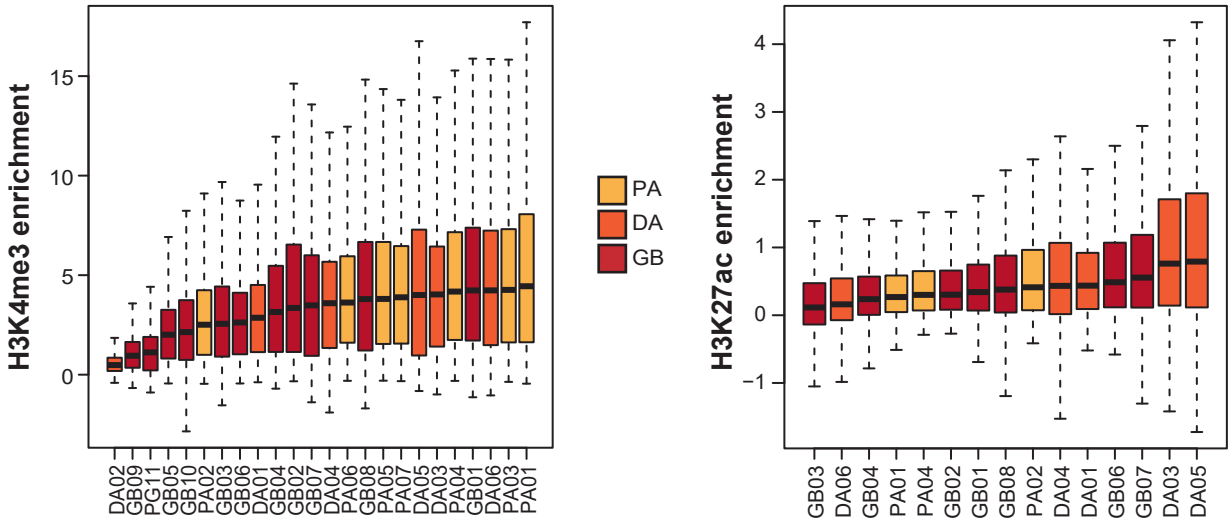
Supplementary Figure 3

A. Overlap of active genes and enhancers sets between patients. Colors indicate experimental method used to detect the activity.

B. Size of intersections between H3K27ac ChIP-seq peaks in active promoters and enhancers identified in this work (glioma-derived) and H3K27ac peaks from normal brain (control samples from Sun et al. 2016) or non-brain cell lines from the ENCODE project (Gerstein et al. 2012) vs total size of promoter/enhancer peaks detected in this work.

C. Comparison of RNA-seq results between different glioma Grades and hospitals where brain samples have been collected.

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

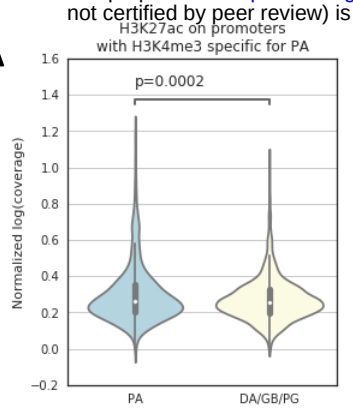


Supplementary Figure 4

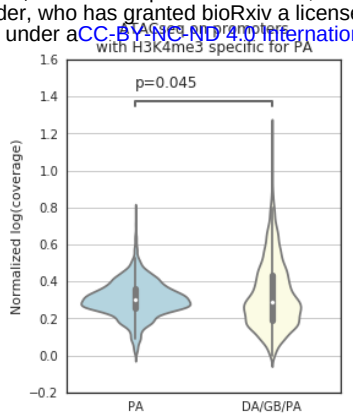
Boxplots representing global enrichment in H3K4me3 (left) and H3K27ac (right) in the regions around TSS in every single glioma sample form our cohort order from lowest to highest median value. The colour coding has been used to discriminate between major malignancy groups.

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A



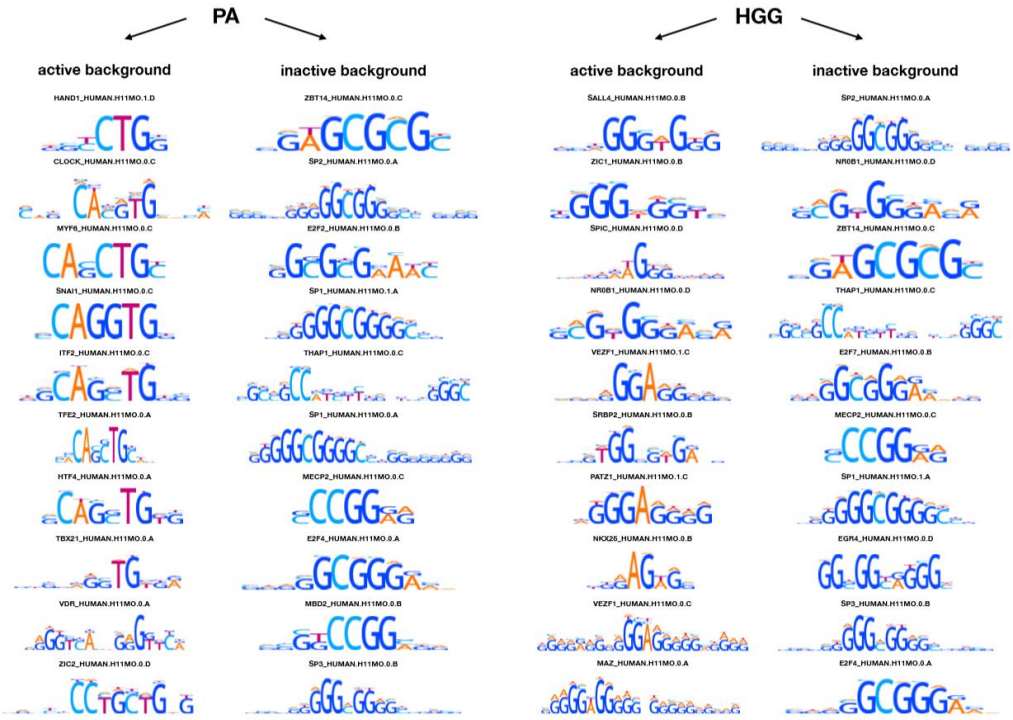
B



C

Rank	Target	PWM	Motif ID	Raw score	P-value	In top motifs
1	MA0506.1 NRF1		MA0506.1 NRF1	17	6.82e-141	40 %
2	MA1122.1 TFDP1		MA1122.1 TFDP1	3.89	6.32e-134	39 %
3	MA0632.1 Tcfl5		MA0632.1 Tcfl5	1.44	2.24e-129	40 %
4	MA0470.1 E2F4		MA0470.1 E2F4	6.23	4.86e-109	39 %
5	MA0131.2 HINFP		MA0131.2 HINFP	2.51	2.46e-104	33 %
6	MA1099.1 Hes1		MA1099.1 Hes1	1.37	4.43e-102	32 %
7	MA0516.1 SP2		MA0516.1 SP2	95	8.94e-101	31 %
8	MA0162.3 EGR1		MA0162.3 EGR1	1.74	2e-98	28 %
9	MA0732.1 EGR3		MA0732.1 EGR3	4.99	4.81e-93	27 %
10	MA0527.1 ZBTB33		MA0527.1 ZBTB33	1.14	1.88e-80	26 %

D



bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

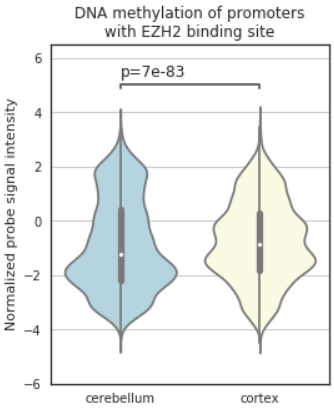
E

GO biological process complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
locomotory behavior	192	8	.85	9.43	+	2.86E-06	4.06E-03
↳behavior	555	12	2.45	4.89	+	7.14E-06	7.43E-03
sodium ion transport	169	7	.75	9.37	+	1.28E-05	1.17E-02
↳monovalent inorganic cation transport	415	10	1.83	5.45	+	1.78E-05	1.54E-02
↳cation transport	819	16	3.62	4.42	+	6.74E-07	1.76E-03
↳ion transport	1284	18	5.67	3.17	+	1.28E-05	1.25E-02
↳metal ion transport	630	16	2.78	5.75	+	2.07E-08	1.08E-04
sensory perception of sound	148	6	.65	9.17	+	6.14E-05	4.36E-02
↳nervous system process	1335	17	5.90	2.88	+	7.59E-05	4.94E-02
↳system process	1893	21	8.37	2.51	+	6.96E-05	4.73E-02
↳multicellular organismal process	6695	58	29.59	1.96	+	1.90E-09	1.48E-05
inorganic cation transmembrane transport	569	13	2.51	5.17	+	1.58E-06	3.09E-03
↳cation transmembrane transport	613	13	2.71	4.80	+	3.52E-06	4.23E-03
↳ion transmembrane transport	886	16	3.92	4.09	+	1.85E-06	3.21E-03
↳transmembrane transport	1208	17	5.34	3.18	+	2.20E-05	1.72E-02
↳inorganic ion transmembrane transport	650	14	2.87	4.87	+	1.20E-06	2.67E-03
epithelium development	1075	16	4.75	3.37	+	2.04E-05	1.67E-02
↳anatomical structure development	5189	46	22.93	2.01	+	3.50E-07	1.09E-03
↳developmental process	5527	52	24.43	2.13	+	1.80E-09	2.81E-05
cell differentiation	3574	35	15.80	2.22	+	2.27E-06	3.54E-03
↳cellular developmental process	3665	35	16.20	2.16	+	3.46E-06	4.50E-03
animal organ development	3077	29	13.60	2.13	+	4.78E-05	3.55E-02
↳system development	4244	38	18.76	2.03	+	6.15E-06	6.86E-03
↳multicellular organism development	4820	46	21.30	2.16	+	3.27E-08	1.28E-04

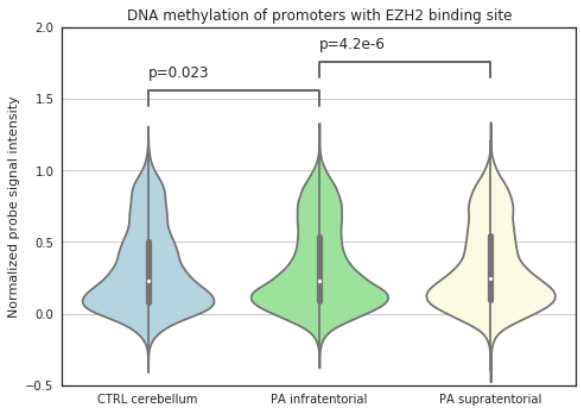
	Homo sapiens (REF)	upload_1 (▽ Hierarchy NEW! ⓘ)					
GO molecular function complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
luteinizing hormone receptor activity	2	2	.01	> 100	+	1.14E-04	3.80E-02
catecholamine binding	19	3	.08	35.72	+	1.20E-04	3.71E-02
solute:sodium symporter activity	68	6	.30	19.96	+	8.98E-07	1.39E-03
↳sodium ion transmembrane transporter activity	147	7	.65	10.77	+	5.33E-06	4.14E-03
↳monovalent inorganic cation transmembrane transporter activity	380	10	1.68	5.95	+	8.45E-06	4.37E-03
↳inorganic cation transmembrane transporter activity	585	14	2.59	5.41	+	3.51E-07	8.16E-04
↳cation transmembrane transporter activity	641	14	2.83	4.94	+	1.02E-06	1.18E-03
↳ion transmembrane transporter activity	896	15	3.96	3.79	+	9.89E-06	4.60E-03
↳transmembrane transporter activity	1066	16	4.71	3.40	+	1.84E-05	7.13E-03
↳transporter activity	1261	18	5.57	3.23	+	1.00E-05	4.24E-03
↳inorganic molecular entity transmembrane transporter activity	839	15	3.71	4.05	+	4.56E-06	4.24E-03
↳metal ion transmembrane transporter activity	432	14	1.91	7.33	+	9.21E-09	4.29E-05
↳solute:cation symporter activity	99	6	.44	13.71	+	6.99E-06	4.65E-03
↳symporter activity	134	6	.59	10.13	+	3.60E-05	1.29E-02
↳secondary active transmembrane transporter activity	219	8	.97	8.27	+	7.28E-06	4.24E-03

	Homo sapiens (REF)	upload_1 (▽ Hierarchy NEW! ⓘ)					
GO cellular component complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
synapse part	745	12	3.29	3.64	+	1.19E-04	4.74E-02
plasma membrane region	1095	15	4.84	3.10	+	9.62E-05	4.78E-02
↳plasma membrane part	2836	28	12.53	2.23	+	2.69E-05	5.36E-02
↳plasma membrane	5592	43	24.72	1.74	+	5.03E-05	3.34E-02
↳cell periphery	5712	44	25.25	1.74	+	3.44E-05	3.42E-02

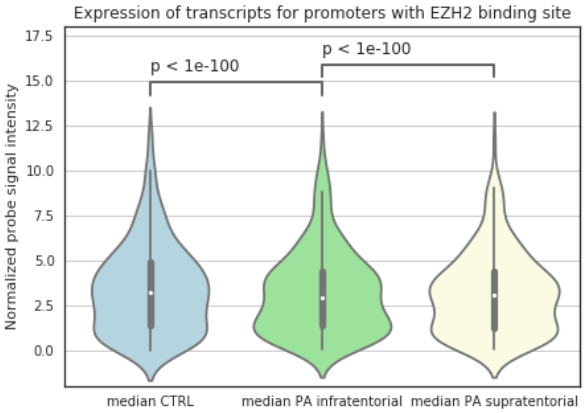
F



G



H



bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure 5

A. H3K27ac ChIP-seq signal on promoters for which H3K4me3 presence is specific for PA compared for PA and higher grade samples.

B. ATAC-seq signal on promoters for which H3K4me3 presence is specific for PA compared for PA and higher grade samples.

C. Transcription factor motifs enrichment in the promoters specifically active in PA compared to promoters inactive in all samples.

D. Transcription factor motifs enrichment in the promoters specifically active in PA or higher Grade gliomas (HGG) compared to promoters inactive in all samples (inactive background) or active in all samples (active background).

E. Enrichment of GO terms calculated for genes specifically active in PA and having the EZH2 protein binding site in promoters.

F. DNA methylation levels on promoters specifically active in PA and having the EZH2 protein binding site. Comparison between control cerebellum and cortex (based on Pidsley et al. 2014)

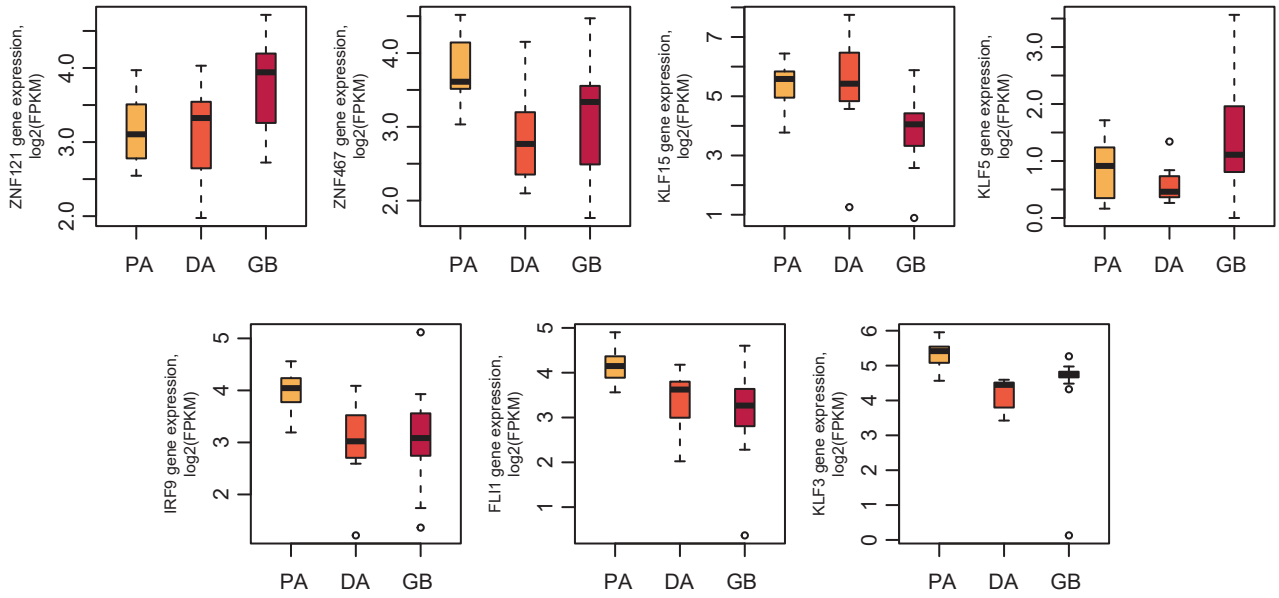
G. DNA methylation levels on promoters specifically active in PA and having the EZH2 protein binding site. Comparison between control cerebellum and pilocytic astrocytomas located in infra- or supratentorial brain regions (based on Lambert et al. 2013)

H. Expression of transcripts related with promoters specifically active in PA and having the EZH2 protein binding site. Comparison between control cerebellum and pilocytic astrocytomas located in infra- or supratentorial brain regions (based on Lambert et al. 2013).

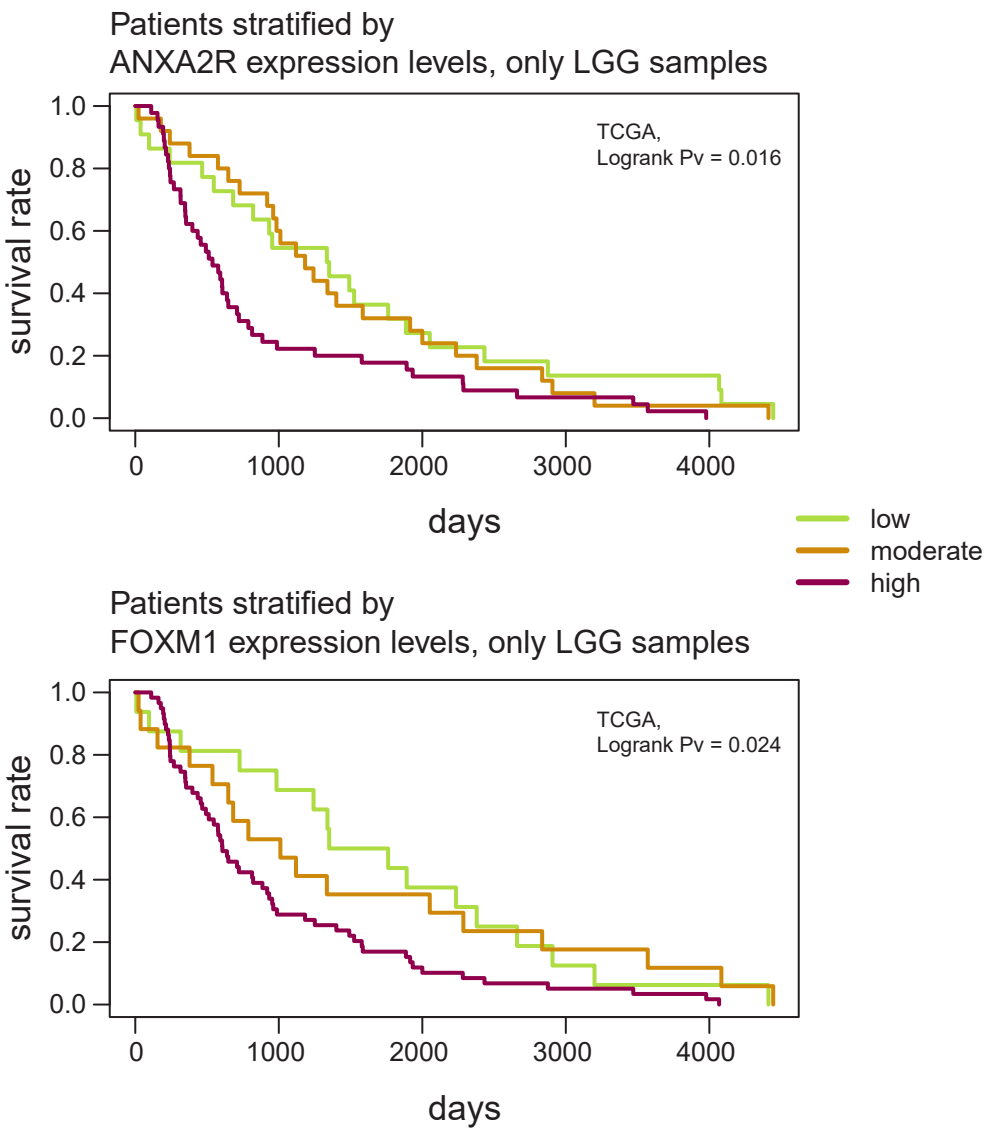
All p-values were calculated with the one-sided Mann-Whitney U test.

bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A



B



bioRxiv preprint doi: https://doi.org/10.1101/867861; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure 6

A. Expression levels of genes coding for transcription factors identified to have a potential binding site in the putative ANXA2R enhancer region. Gene expression levels are represented as log-transformed FPKMs. The colour coding has been used to discriminate between major malignancy groups.

B. Kaplan-Meier curves plotted for the low grade glioma (LGG) data from TCGA database. The patients were stratified by either by ANXA2R (top part) or by FOXM1 (bottom part) expression levels into three equinumerous sub-groups. Differences in survival rates between sub-groups were assessed with the log-rank test.