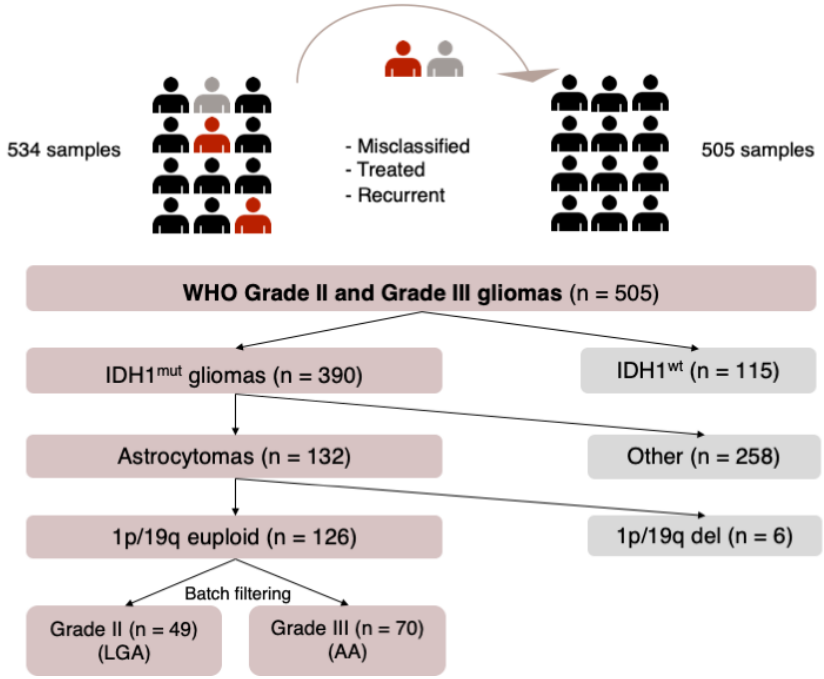
Supplementary Figures

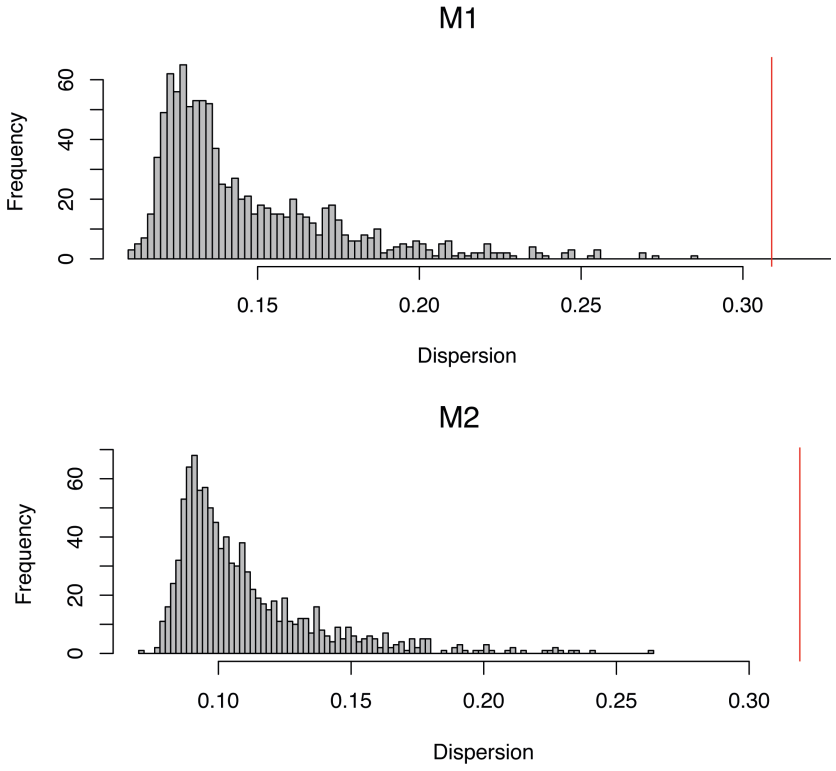
Supplementary Figure S1: Sample selection diagram.

RNA-sequencing, somatic mutation and SNP-array data for 534 grade II and grade III glioma samples were acquired from TCGA. Recurrent, previously treated samples, and samples for which the ICD-O-3 histology code did not match the tumour grade (misclassified) were omitted. The tumour samples were further selected based on the *IDH1* mutation status, histology and 1p/19q ploidy. Samples, that were analysed in independent batches or batches where only either grade II or grade III tumours were sequenced, were excluded from down-stream analysis. *AA – anaplastic astrocytoma, LGA – low-grade astrocytoma.*



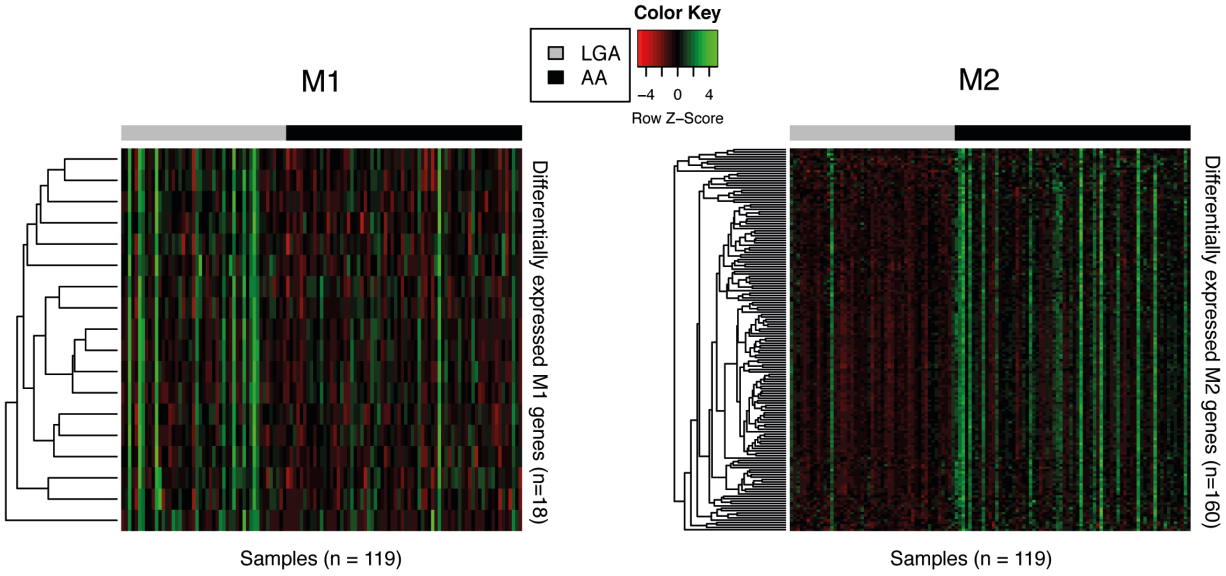
Supplementary Figure S2. Significance of differential expression.

Distribution of module-wise dispersion statistic quantifying the average change in correlation between pairs of genes drawn from module M1 (upper panel) and M2 (lower panel) in 1,000 bootstrap permutation samples of LGA and AA. The red vertical line indicates the value of the dispersion statistic observed in the data. The dispersion statistic observed in M1 and M2 modules was significantly higher than expected by chance (*P*-value < 0.001). *AA – anaplastic astrocytoma, LGA – low-grade astrocytoma.*



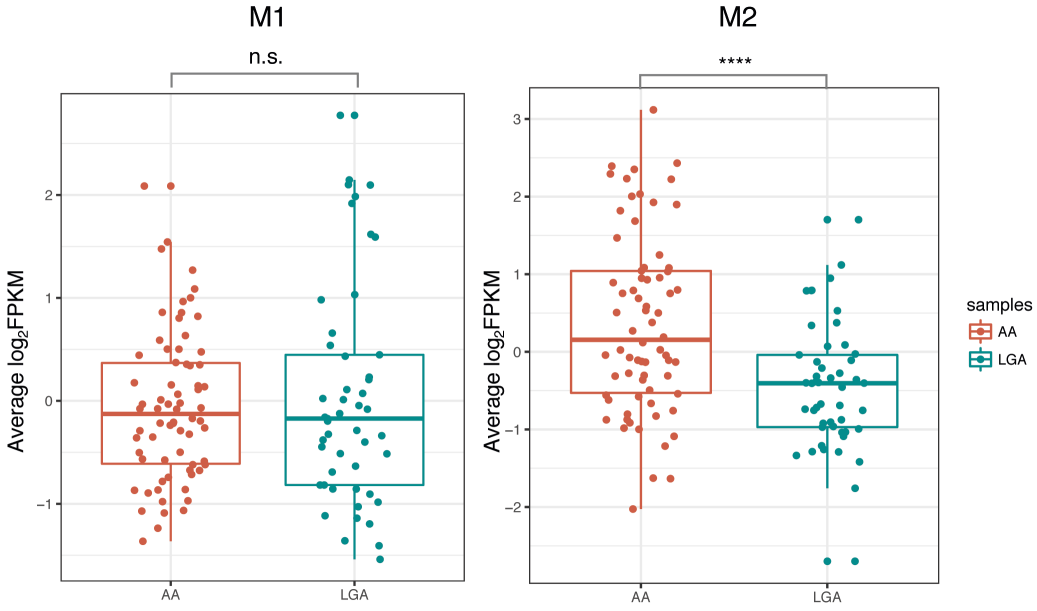
Supplementary Figure S3. Differential expression analysis between AA and LGA.

Heatmap of differentially expressed module M1 (left panel) and M2 (right panel) genes between LGA and AA gliomas. Of module M1 genes, 18 out of 272 were significantly differential (down-regulated in AA) and 160 out of 171 module M2 genes were differentially expressed (up-regulated in AA). *AA – anaplastic astrocytoma, LGA – low-grade astrocytoma.*



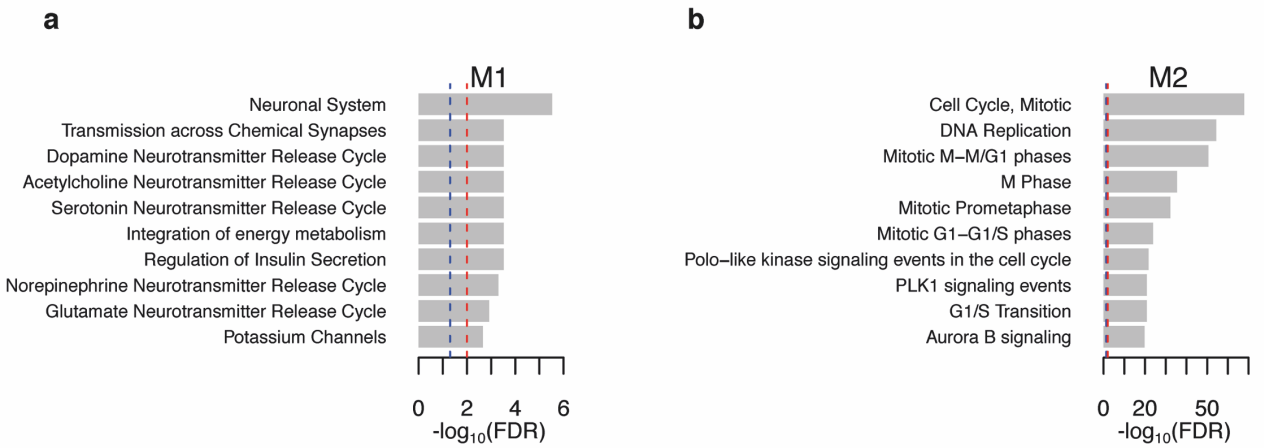
Supplementary Figure S4. Average expression of module M1 and M2 genes in LGA and AA gliomas.

The average normalised $\log_2$(FPKM) of module M1 did not differ significantly (*P*-value = 0.60) (right panel), while the average M2 expression was significantly higher in the AA samples compared to LGA (*P*-value = 1.32×10^{-5}) (left panel) (Mann-Whitney U-test). Each data point represents average expression across all module genes within each sample. *P*-value < 0.0001 ****. *AA – anaplastic astrocytoma, LGA – low-grade astrocytoma*



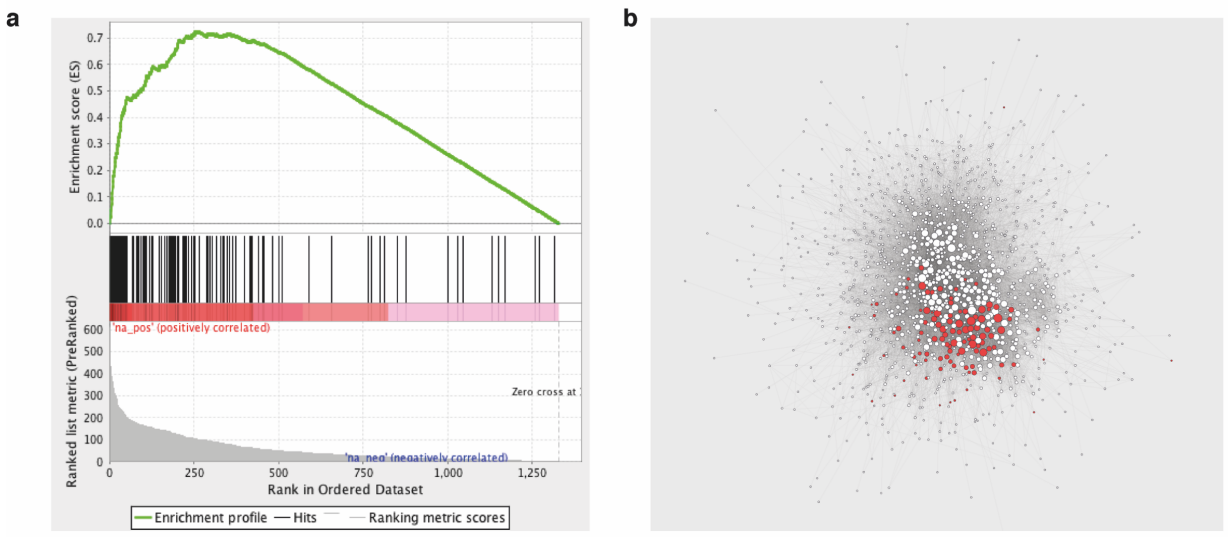
Supplementary Figure S5. Pathway enrichments results of module M1 and M2 from Web-Gestalt Pathway Commons analysis.

a, Bar-plot showing the top 10 most significant functional enrichments of module M1 genes. The top hit includes genes of the neuronal system. **b**, Bar-plot of top 10 most significant functional enrichments in module M2. The cell cycle pathway genes were most significantly enriched. The horizontal axis represents the significance of over-representation of biological pathway genes in modules calculated by hypergeometric test. Blue dashed line represents FDR = 0.05, and red dashed line FDR = 0.01.



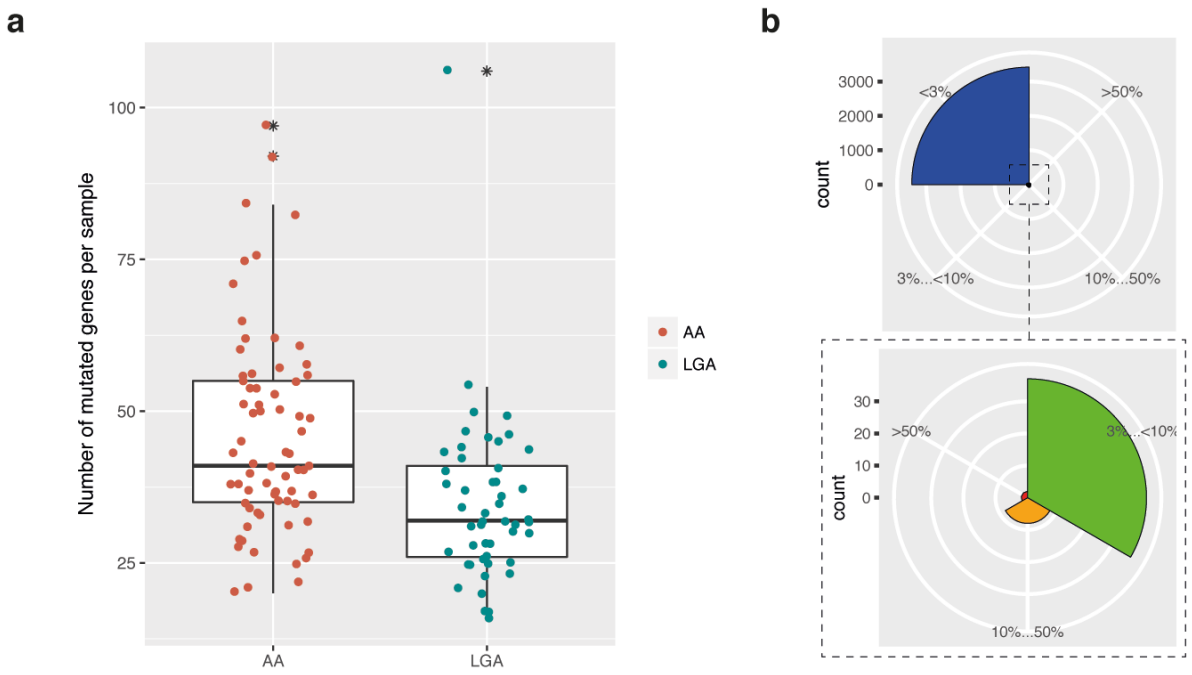
Supplementary Figure S6. Cell cycle genes with high connectivity are significantly enriched for the M2 genes.

a, Cell-cycle genes (GO:0007049 , n = 1768) with known protein-protein interaction (PPI) annotation (n=1328/1768) ranked by the degree of connection (number of interaction partners within the cell-cycle gene list). Module M2 genes with known PPI (133/177) were used as the set list for the GSEA. The M2 genes were significantly (*P*-value < 0.001) enriched among the most highly connected cell-cycle genes. **b**, Graphical representation of M2 genes in a perfuse force directed network layout. The size of the node corresponds to the number of protein-protein connections of each node. Module M2 genes are marked in red.



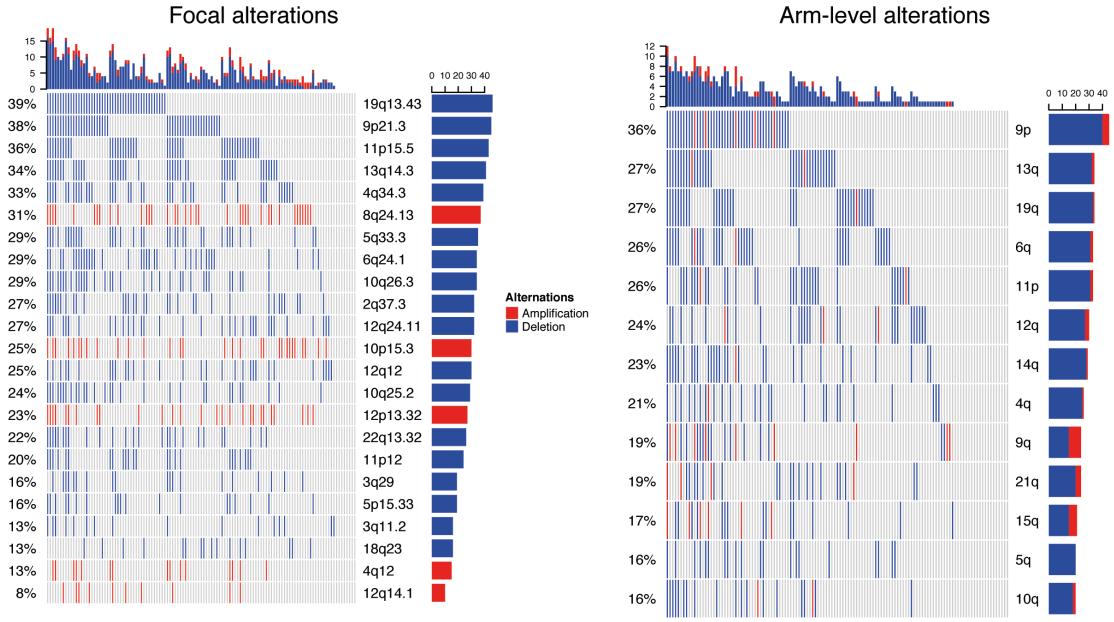
Supplementary Figure S7. Frequency of somatic mutations (base substitutions and small insertions/deletions) in *IDH1*-mutated 1p/19q euploid astrocytomas.

a, On average, the AA tumours have more somatically mutated genes per sample than the LGAs (P-value = 0.0002, independent two-sample t-test). b, Majority of genes are mutated in only less than 3% of the *IDH1*-mutated and 1p/19q euploid astrocytomas (n = 119). The most commonly shared mutations amongst the *IDH1*-mutated and 1p/19q euploid astrocytomas were the *TP53* mutation (106/119, 89.1%) and the *ATRX* mutation (85/119, 71.4%). *AA – anaplastic astrocytoma, LGA – low-grade astrocytoma.*



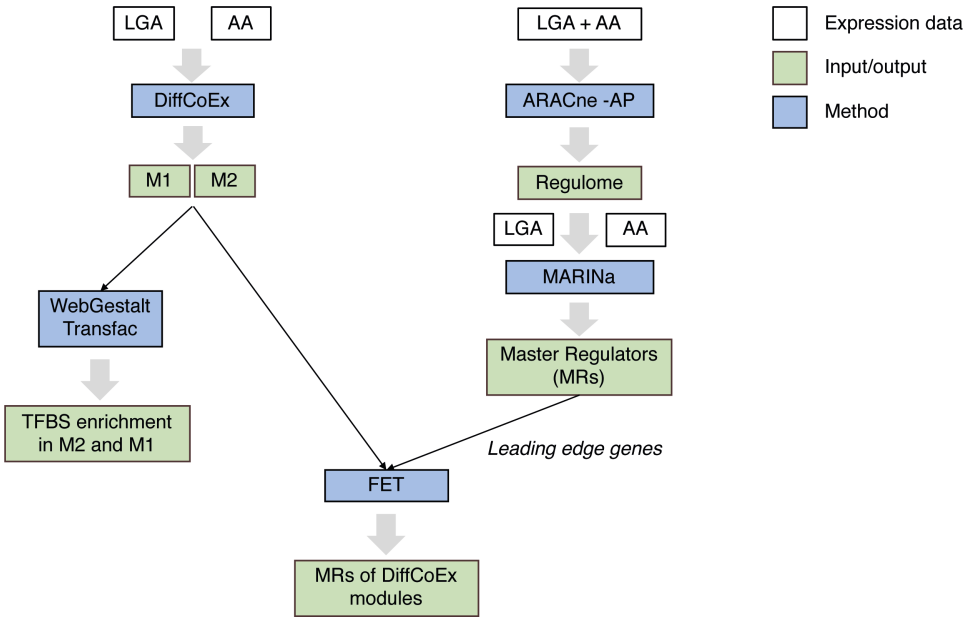
Supplementary Figure S8. Heatmap and frequencies of copy-number alterations across the *IDH1*-mutated and 1p/19 euploid astrocytoma cohort.

Significant (FDR < 0.1) focal copy-number alterations in LGA and AA detected by GISTIC2.0 algorithm with threshold 0.1 (left panel) and significant (FDR < 0.1) chromosome arm-level alterations in LGA and AA (right panel). The bar-plot on top of the heatmaps shows the number of samples with a given combination of significant alterations (*y*-axis). The bar-plot on the side of the heatmap shows the overall number of samples with a given alteration (*x*-axis). Each row in the heatmap specifies an alteration and the columns represent individual samples.



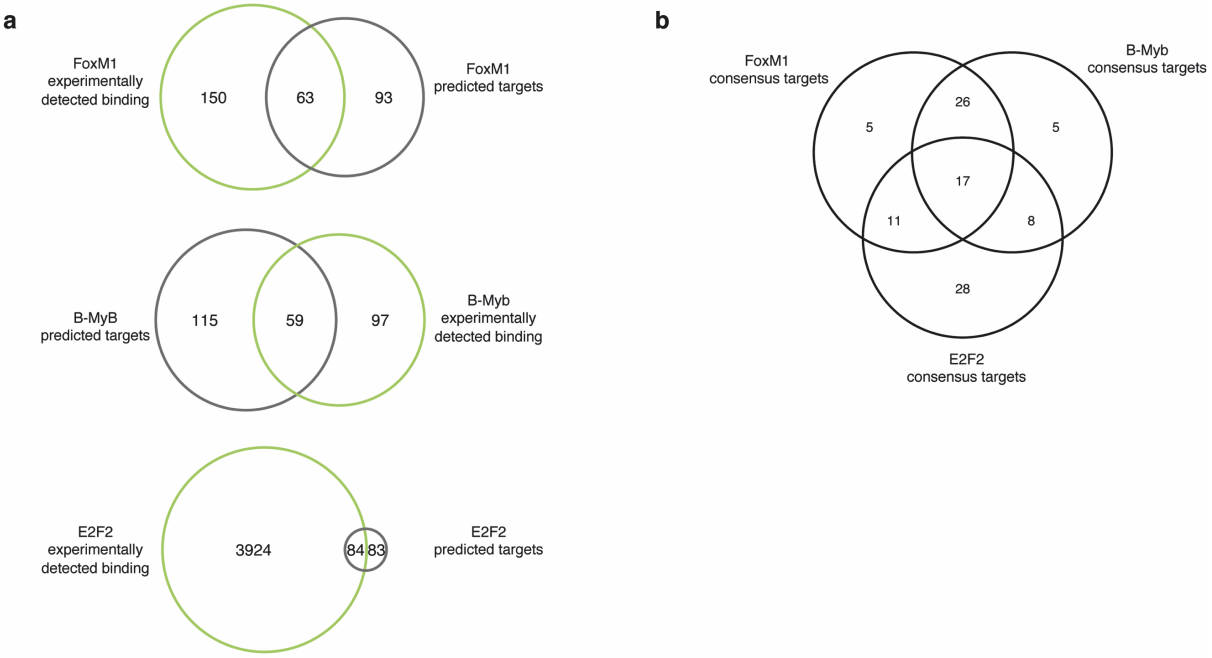
Supplementary Figure S9. Workflow to determine the source of regulatory influence on DiffCoEx modules M1 and M2.

Differentially co-expressed modules of genes (M1 and M2) were identified by DiffCoEx algorithm. *IDH1*-mutated astrocytoma specific gene-regulatory relationships between TF and potential targets (regulome) were inferred using ARACNe- AP. The master regulators (MRs) of phenotypic change from LGA to AA were derived using the MARINa algorithm, which makes use of the gene expression to prioritise regulons enriched for genes differential between the two conditions. The most highly enriched genes for each MR regulon at the positive and negative extremes (leading edge) were extracted. The overlap between the MR-targets and differentially co-expressed module genes was used to prioritize TFs that are predicted to regulate the DiffCoEx modules. In parallel, the significant over-representation of TFBS was sought using web-based tools - WebGestalt and Transfac database. Co-variate corrected $\log_2$(FPKM) values from AA and LGA were used to build the differential co-expression modules and the regulome. *AA – anaplastic astrocytoma, ARACNe-AP –Accurate Reconstruction of Cellular Networks with Adaptive Partitioning, FET – Fisher's exact test, DiffCoEx – differential co-expression analysis, LGA – low-grade astrocytoma*, *MARINa – Master Regulator Inference algorithm, MR – master regulator, TF – transcription factor, TFBS – transcription factor binding site, WebGestalt – Web-based Gene Set enrichment analysis toolkit*



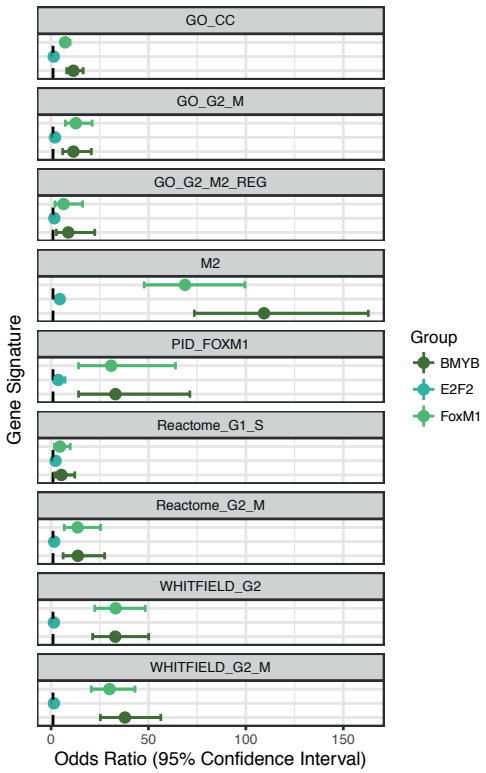
Supplementary Figure S10. Overall concordance between the predicted and validated TF of FoxM1, B-Myb and E2F2.

a, Gobal overlap between TF targets predicted by MARINa (grey) and experimentally detected in ChIP-seq analysis (green). **b**, Overlap of consensus targets (intersect of MARINa predicted and ChIP-detected TF targets) of FoxM1, B-Myb and E2F2 in module M2.



Supplementary Figure S11. Overlap of ChIP-seq targets of FoxM1, B-Myb and E2F2 with gene signatures.

Enrichment of ChIP-seq detected targets of FoxM1, B-Myb and E2F2 among gene sets of Gene Ontology Cell Cycle (GO_CC) (n = 1768), Gene Ontology Cell Cycle G2 to M progression (GO_G2_M2) genes (n = 138), Gene Ontology Cell Cycle G2 to M progression regulating (GO_G2_M2_REG) genes (n = 59), FOXM1 transcriptional network (PID_FOXM1) (n = 40), Reactome Cell Cycle G1 to S progression (Reactome_G1_S) (n = 115), Reactome Cell Cycle G2 to M progression (Reactome_G2_M) (n = 82), genes peaking in synchronized HeLa cells during G2 phase (WHITFIELD_G2) (n = 182), and genes peaking in synchronized HeLa cells during G2 to M transition (WHITFIELD_G2_M) (n = 216). The odds ratio was calculated by FET. Module M2 genes have the strongest enrichment of FoxM1, B-Myb and E2F2 ChIP-seq detected targets.



Supplementary Figure S12. M2 network depicted as consensus transcription factor (TF) - target interactions for master regulators (MRs) *FOXM1*, *MYBL2* and *E2F2*.

Differential expression in response to **a**, 10nM and **b**, 2nM resveratrol treatment is shown in yellow to red (absolute $\log_2$ fold change value of gene expression between resveratrol and DMSO treated primary astrocytoma-derived cells). Network edges represent an interaction between MR and its target as inferred by MARINa detected by ChIP-seq (intersect). Edge colours are assigned to individual MR-target interactions. All TFs in the network are noted as triangles. The size of the node represents the number of MR-target interactions for a given gene and the border colour symbolizes up- or down-regulation in response to resveratrol.