Extended Materials and Methods

Data analysis

The TCGA project (1) has created single expression data sets for GBM that incorporate results from different platforms and participating centers. The UNC miR and mRNA expression data sets, representing 534 miRs and 17814 genes respectively, were used as the starting point for our analysis. Since miR-mRNA network inference presupposes the availability of both miR and gene level expression data, 290 samples that were represented in both the miR and mRNA expression sets were used for subsequent analysis.

The CLR method and application to GBM

For the analysis of the matched miR-mRNA expression profiling data we utilized a previously described method, namely the CLR (Context Likelihood of Relatedness) algorithm (2). The CLR algorithm builds upon the relevance network strategies, by applying a background correction step. After computing the mutual information between regulators and their potential target genes, CLR calculates the statistical likelihood of each mutual information value within its network context. The algorithm compares the mutual information between a miR/gene pair to the background distribution of mutual information scores for all possible miR/gene pairs that include either the miR or its target. After this background correction, the most probable interactions are those whose mutual information scores stand significantly above the background distribution of mutual information scores. This step removes many of the false correlations in the network by eliminating promiscuous cases in which one transcription factor weakly covaries with a large numbers of genes or one gene weakly covaries with many transcription factors. Such promiscuity

arises when the assayed conditions are inadequately or unevenly sampled, or when microarray normalization fails to remove false background correlations due to inter-lab variations in methodology. The mutual information for two discrete random variables X and Y is defined as:

$$MI(X,Y) = \sum_{i,j} P(x_i, y_j) \log\left(\frac{P(x_i, y_j)}{P(x_i)P(y_j)}\right)$$

where $P(x_i)$ is the probability that $X=x_i$. For genes, X and Y represent a transcription factor and its potential target gene, and x_i and y_i represent particular expression levels.

CLR uses the matrix of mutual information (MI) values between all probe sets on the array. The CLR algorithm estimates a likelihood of the mutual information score for a particular pair of genes, i and j by comparing the MI value for that pair of genes to a background distribution of MI values. The background distribution is constructed from two sets of MI values: MI_i, the set of all the mutual information values for gene i (in row or column i), and MI_j, the set of all the mutual information values for gene j (in row or column j). Because of the sparseness of biological regulatory networks, most MI scores in each row of the mutual matrix represent random background mutual information. We approximate this background MI as a joint normal distribution with MI_i and MI_jas independent variables and compute the z-score for each of those distributions

$$Zi = \frac{MI(i^{th}\ entry\ in\ i^{th}\ row) - mean(i^{th}\ row)}{\sigma(i^{th}\ row)} \text{ and}$$

$$Zj = \frac{MI(j^{th}\ entry\ in\ j^{th}\ column) - mean(j^{th}\ column)}{\sigma(j^{th}\ column)}.$$

The new metric for the relationship between gene i and gene j is then $Z_{ij} = \frac{Z_i + Z_j}{\sqrt{2}}$,

where Z_i and Z_j are the z-scores of the marginal distributions (3). The new metric is then used as a measure of the relationships among the genes. P-values corresponding to the entire set of z-

scores are then calculated by comparing to the normal distribution; a stringent FDR of 5% is finally applied to the computed p-values which defines the inferred network.

Because the CLR algorithm relies on variability in the dataset to estimate significant edges, we first filtered the 534 miRs and 17814 mRNAs based on the variability they showed in the 290 GBM samples. We chose to apply CLR on those miR and mRNA transcripts, which had variance across the 290 samples, higher than the median variance across all species. This yielded 267 miR and 8907 transcripts among which we looked to infer regulatory relationships. Similarly for the subtype specific analysis, we required that the miR and mRNA transcripts, have variance across their respective subtypes, higher than the median variance across all species in that subtype.

Copy number analysis

Segmented copy number data were downloaded from the TCGA data portal (http://tcga-data.nci.nih.gov/tcga) and were processed using the CN Tools (http://www.bioconductor.org/packages/2.6/bioc/html/CNTools.html) package of the Bio-conductor project to derive a matrix with genes as rows and samples as columns by assigning each gene the corresponding segment value for each sample. The SGOL (Segment Gain or Loss) score was then calculated for each gene by summarizing the positive or negative values for a given gene across samples using the cghMCR package (http://www.bioconductor.org/packages/2.6/bioc/html/cghMCR.html). Genes with large positive and small negative SGOL scores have high amplitude and/or recurrence of alterations across samples. The threshold used for a single sample was >0.3 for gain and < -0.3 for loss.

Thresholds of SGOL > 28 and SGOL < -32, corresponding to the 90th percentiles of gain and loss SGOL analyses respectively, were set to filter genes based on copy number data.

Correlating gene expression with copy number

Gene weight analysis can be used to evaluate the influence of copy number alterations on gene expression (4). In this work, gene weight analysis was applied to find genes or miRs with high correlation between their expression and copy number levels. For a particular probe gene/miR, expression values of all samples were divided into two groups according to copy number status. For example, expression in tumors with copy number alterations (CNA) compared to without CNA. The Gene weight score was calculated by the following formula:

$$GW = \frac{m_T - m_R}{\sigma_T + \sigma_R},$$

where m_T, m_R and, σ_T, σ_R denote the mean and

standard deviations for expression values for two groups T and R respectively. Statistical significance was determined by permuting sample labels for expression data. A low p-value indicates a strong correlation between the expression and copy number level of the gene or miR under consideration.

Survival analysis

To dichotomize GBM patients based on the expression levels of a targeted miR, an optimized cut-off value was determined that defined high expression and low expression groups with the most significant difference in overall survival. Statistical significance was estimated by log-rank test on the two corresponding Kaplan-Meier survival curves. P-values were adjusted by

Bonferroni correction for the number of cut-off values tested. Hazard ratios, P-values, and C-indexes were calculated in R statistical environment.

Permutation test

We used a permutation based approach to evaluate the strength of miR:mRNA edges in a particular subtype. We first apply the CLR algorithm to samples belonging to the 4 different subtypes separately. Significances of miR:mRNA associations that are identified exclusively in one subtype are assessed by three permutation tests as follows. Let us denote the 4 subtypes by A, B, C and D, with A being the subtype in which the uniqueness of a given correlation is being evaluated.

i. Compute the difference in miR:mRNA correlations between subtype A and one of the other subtypes, say subtype B.
ii. Permute the sample labels between subtypes A and B and compute the difference in correlations in the new set. Repeat this for a 1000 permutations to get the background distribution.
iii. The p-value is assigned to the original difference calculated in step (i) by comparing it against the background distribution.
iv. Repeat steps (i)-(iii) for subtypes C and D.
v. If the three p-values computed in steps (i)-(iv) are all less than 0.01, the correlation is considered unique to subtype A.

Identification of discriminatory miRs between PN and MS subtypes

To identify a core network of PN and MS differences, we looked for miR hubs that had significant number of connections to either PN and MS gene signatures, which were defined as genes that are highly expressed in the respective groups. We scored each miR hub with:

$$score\_mirna(i) = \frac{d1 - d2}{(Lmes + Lpro)/2 + Lhub}$$

where d1 is the number of MS signatures genes which are negatively correlated with the miR expression and the number of PN signature genes which are positively correlated with the miR expression; d2 is the number of MS signatures genes which are positively correlated with the miR and the number of PN signature genes which are negatively correlated with the miR. Lmes is the size of the MS signature, Lpro is the size of the PN signature and Lhub is the number of connections identified by CLR for the miR. This score would be high in two cases; miRs whose edges were to genes that showed negative correlation with MS genes and positive correlation with PN genes or vice versa. The score obtained for each miR was Z-transformed and a false positive rate of 0.001 was applied to identify 8 miR hubs with the greatest discriminatory potential between the PN and MS subtypes.

Transcription factors analysis

Transcription factor binding site motifs from the TRANSFAC database (5) (release 10.1) were identified in genomic regions comprising 8kb upstream and 2kb downstream of the transcription start site of genes based on coincidental prediction by the CisGenome (conservation cutoff = 50, likelihood ratio = 500, third order background Markov model) and MotifScanner (prior probability of 1 motif copy = 0.9) programs (6-9). Two sets of genes corresponding to mir-34a CLR targets that define either the PN or MS signatures were compared for differential over-representation of transcription factor binding sites using this approach. For each TF the

proportion of genes with at least one corresponding binding site were compared between the two groups. 34 vertebral transcription factors were thus identified (p.val< 0.05). This list was further screened for transcription factors that were overexpressed and had a statistically significant negative correlation with mir-34a (corr< -0.3, p-val< 0.05) in the PN group.

Vectors and constructs

The pHAGE EF1αL GFP UBC GFP, pHAGEEF1S GFP UBC GFP and pHAGE CMV RFP UBC GFP were kindly provided by Darrel Kotton, the pLB vector was originally made by Stephan Kissler and obtained through Addgene. The pWZL and pBabe vectors were originally made by Bob Weinberg and obtained through Addgene. The Gateway compatible expression vectors were obtained by removing the GFP and RFP cassette in the pHAGE EF1αL GFP UBC GFP and pHAGE CMV RFP UBC GFP and inserting a gateway cassette in the NotI/BamHI cloning sites. The Lenti Decoy system has been engineered by inserting a gateway cassette in the EcoRI site of the pLB vector and in the BamHI site of the pHAGE CMV RFP UBC GFP downstream to the sequence encoding for the fluorescent protein. All the constructs were sequenced.

The pre-mir-34a sequence was PCR amplified from a mir-34a expressing vector (MSCV PIG mir-34a, provided by Dr. Joshua Mendell) and made compatible for gateway cloning.The Decoy sequence was designed as previously described (10). The sequence was inserted in the Lenti Decoy vectors by Gateway recombination. The Lenti-Cre construct was provided by Dr. Ronald Depinho.

The Entry constructs and the pLKO.1 based shRNA constructs were obtained through the DFCI/Harvard Dna Core Facility. The lentiviral based pre-miR expression clones were obtained

from System Biosciences. The 3'UTR dual reporters were purchased from Genecopoeia. A control Firefly vector lacking the miR-34a binding site was used as negative control. The luciferase lentiviral construct for live bioimaging was purchased from System Biosciences.

Cell culture and virus production

Glioma cell lines, human and mouse astrocytes were grown in DMEM supplemented with 10% FBS and 1% pen/strep. Human and mouse neurospheres were maintained in Neurosphere's Medium (Stemcell Technologies) following manufacturer's instructions in presence of EGF/FGF (20ng/ml). Lentiviruses were produced by transfecting 293T cells by the calcium phosphate method with the expression vectors and a 2nd generation packaging system (pCMV dr8.74/pMD2.G).Viruses were harvested 72h after transfection and directly used for infecting target cells (8 hours in presence of 5ug/ml of polybrene). Human and mouse neurosphere were transduced by the spin-infection method (1000 RPM for 1 hour in an Eppendorf 5810 centrifuge).

Protein and miR collection and quantification

Cells were harvested 48-72 hours after infection, proteins were isolated in RIPA buffer (BBP) in presence of phosphatase and protease inhibitors (Roche), resolved in 4-12% Bis-Tris acrylamide gels and transferred on PVDF membranes; membranes were incubated overnight in presence of the following antibodies: PDGFRA (C-20) (Santa Cruz Biotechnologies sc-338), c-MYC (C-19) (Santa Cruz Biotechnologies), β-ACTIN (Sigma Aldrich, A1978), c-MET, CDK4, CDK6 (Cell Signaling), Smad4 (B-8) (Santa Cruz Biothecnology), Id1 (C-20), Id3 (C-20) (Santa Cruz Biothecnology). miRs were isolated using the MirVana kit (Ambion) following

manufacturer's instructions. miR levels were quantified by real-time RT-qPCR by using the QuantiMir Kit (SBI) following manufacturer's instructions and normalized to an internal control (U6 snRNA). qPCR arrays were run following manufacturer's instructions (SA Bioscience).

Proliferation, tumorigenicity and self-renewal assays

For *in vivo* sub-cutaneous tumorigenicity assays, $1.5x10^6$glioma cell lines and genetically engineered astrocytes were transplanted subcutaneously into the flanks of female Ncr/nude mice (Taconic) after being infected with the appropriate virus and followed for tumor development daily. Tumors were measured every 4 days, starting from day 10 post-injection. Tumor volumes were calculated according the following formula $V=L \times W^2/2$. Mice were sacrificed when sick or when tumors reached 1.5 cm in their largest diameter or become ulcerated. At termination of the experiment, tumors were harvested, and processed for pathological and molecular analyses. For tumor-free survival studies mice were censored upon appearance of a palpable nodule. Brain orthotopic injections were performed as previously described (11). All animal experiments were approved by Harvard's Institutional Animal Care and Use Committee (IACUC).

To measure self-renewal, spheroids were dissociated by mechanical disruption, then re-plated at clonal density in non-adherent cultures on ultra-low binding 12-well plates (Corning). Secondary spheroids were counted 3-5 days later.

Live bioimaging

For whole-body bioluminescent imaging, mice were intraperitoneally injected with $100mg/kg^{-1}$ D-Luciferin (Goldbio), and after 5min, analysed using an IVIS Spectrum system

(Caliper LifeSciences). Quantification was performed using Living Image software (Caliper LifeSciences) following the manufacturer's instructions.

Reporter assays

For 3'UTR reporter assays E6/E7T were seeded in 6-well plates at a density of $5x10^5$ cells per well and transfected with 2ug of total DNA with Fugene 6 following manufacturer's instructions. Luciferase activity was measured after 48 hours and normalized to the internal Renilla control. A control Firefly vector lacking the miR-34a binding site was used as negative control. All the reporter systems were obtained from Genecopoeia.

miRNA in situ hybridization

FFPE tissue microarrays are first deparaffinized in 3 exchanges of xylene, rehydrated with an ethanol gradient, then treated with 20 µg/mL Proteinase K (Roche Diagnostics) for 10 minutes at 37°C, fixed with 4% formaldehyde for 10 minutes, rinsed with 0.13 M 1-methylimidazole and refixed with 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC, Thermo Scientific) for 1 hour as described (12). Endogenous peroxidases are blocked in 1% H_2O_2 in PBS for 30 minutes and slides are prehybridized for 30 minutes at 50°C in hybridization oven (Advanced Cell Diagnostics) with hybridization buffer containing 50% formamide (American Bioanalytical), 5X SSC (American Bioanalytical), 50µg/mL Heparin (Sigma), 0.1% Tween 20 (Sigma), 500µg/mL yeast tRNA (Invitrogen) adjusted to pH 6. Slides were hybridized for 1 hour with 200 nM double digoxigenin (DIG) LNA modified probes for miR-34a (Exiqon). Alongside control slides were employed with 200nM double DIG Scrambled probe and 25 nM 5`DIG labeled U6 probe. Slides are then stringently washed in 2X SSC at the hybridization

temperature once and twice at room temperature followed by washes in PBS 0.1% Tween (PBS-T) and PBS, then blocked in 2% BSA (Sigma) in PBS before incubating with with Anti-Digoxigenin-POD, Fab fragments from sheep (Roche Diagnostics) diluted 1:100 and rabbit anti-GFAP (Dako) diluted 1:100 in block (2% BSA in PBS) for 1 hour at room temperature. The slides are washed twice with PBS-T then once with PBS. The miRNA signal is detected with the TSA Plus Cyanine 5 system (Perkin Elmer), the slides are washed again with PBS-T and PBS as above, and GFAP is detected with Alexa 546-conjugated goat anti-rabbit secondary antibody (Molecular Probes) diluted 1:100 in block for one hour. Finally, the slides are mounted with Prolong mounting medium containing 4', 6-Diamidino-2-phenylindole (DAPI, Molecular Probes). Serial sections of control index slides are ran alongside each run to assess reproducibility and negative control scrambled and postive control U6 are also used for each run.

Automated Quantitative Analysis (AQUA®) quantifies fluorescent intensity within specific subcellular compartments and has been described in detail previously (13). In brief, a series of high resolution monochromatic in- and out-of-focus images were obtained for each histospot using the signal from the DAPI, S100 (GP100)-Alexa 546 and the target-Cy5 channel by the PM-2000 microscope. The GFAP signal is used to distinguish stromal and non-stromal elements by creating a tumor "mask". The AQUA® scores of miR-34 in the tumor mask was calculated by dividing the pixel signal intensity by the area of the tumor mask. AQUA scores are then normalized by the exposure time and bit depth at which the images were captured. Histospots containing less than 0.17mm^2 of tumor were excluded from analysis.

1st (lower) quartile of the AQUA score distribution was used as cut-off to assign patients to the low or high expressors groups.

REFERENCES

1. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455:1061-8. PMCID: 2671642.

2. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, et al. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. 2007;5:e8.

3. Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. J Evol Biol. 2005;18:1368-73.

4. Carrasco DR, Tonon G, Huang Y, Zhang Y, Sinha R, Feng B, et al. High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients. Cancer Cell. 2006;9:313-25.

5. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 2006;34(Database issue):D108-10. PMCID: 1347505.

6. Aerts S, Thijs G, Coessens B, Staes M, Moreau Y, De Moor B. Toucan: deciphering the cis-regulatory logic of coregulated genes. Nucleic Acids Res. 2003;31:1753-64. PMCID: 152870.

7. Aerts S, Van Loo P, Thijs G, Mayer H, de Martin R, Moreau Y, et al. TOUCAN 2: the all-inclusive open source workbench for regulatory sequence analysis. Nucleic Acids Res. 2005;33(Web Server issue):W393-6. PMCID: 1160115.

8. Ji H, Vokes SA, Wong WH. A comparative analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors. Nucleic Acids Res. 2006;34:e146. PMCID: 1669715.

9. Ji H, Wong WH. TileMap: create chromosomal map of tiling array hybridizations. Bioinformatics. 2005;21:3629-36.

10. Gentner B, Schira G, Giustacchini A, Amendola M, Brown BD, Ponzoni M, et al. Stable knockdown of microRNA in vivo by lentiviral vectors. Nat Methods. 2009;6:63-6.

11. Zheng H, Ying H, Yan H, Kimmelman AC, Hiller DJ, Chen AJ, et al. p53 and Pten control neural and glioma stem/progenitor cell renewal and differentiation. Nature. 2008;455:1129-33.

12. Pena JT, Sohn-Lee C, Rouhanifard SH, Ludwig J, Hafner M, Mihailovic A, et al. miRNA in situ hybridization in formaldehyde and EDC-fixed tissues. Nat Methods. 2009;6:139-41. PMCID: 2838186.

13. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. Nat Med. 2002;8:1323-7.

Supplementary Figures Legend

Supplementary Figure 1. Characteristic miR-mRNA edges detected by CLR. miR-mRNA relationships detected by application of CLR to all GBM samples. In this example, PN samples (red dots) display under-expression of miR-34a and over-expression of PDGFRA while a reverse trend is apparent in MS samples (blue dots). These strong subtype differences account for a significant global negative correlation and identification of the miR-34a: PDGRA edge by CLR.

Supplementary Figure 2. Concomitant genetic loss of *p53* and *Pten* in murine neural stem cells and progenitor cells promotes the formation of proneural malignant gliomas. (A) malignant gliomas in *p53L/L;PtenL/+;Gfap-Cre* morphologically resemble human proneural Glioblastomas. Here H&E section showing oligondroglioma-like areas. **(B)** Gene set enrichment analysis of spheroid cultures derived from p53/Pten deficient mouse malignant gliomas *(p53L/L;PtenL/+;Gfap-Cre)*. A clear enrichment is observed for PN signature genes (n=14, p=2.2 x 10-16). Abbreviations: PN=Proneural, MS=Mesenchymal, CL=Classical, NL=Neural, TS=tumor spheroids.

Supplementary Figure 3. Expression levels of the specific miRNAs in *p53/Pten-/-* E13 embryonic neural stem cells infected with lentiviral miRNA precursors (supporting information to Figure 2D). Pre malignant embryonic spheroids where infected with a lentiviral vector expressing a microRNA precursor under a constitutive promoter (CMV). Total RNA was harvested 72h after the infection and miRNA levels were measured by RT-qPCR. Levels were normalized to the control cells. Error bars represent Standard Deviation of replicates.

Supplementary Figure 4. miR-34a expression in an independent cohort of GBM (supporting data to Figure 3B). Representative TMA spots with high and low expression of miR-34a.

Supplementary Figure 5. miR-34a levels are significantly lower in proneural GBM.

(A) miR-34a expression in the four subtypes of GBM. miR-34a is strongly down-regulated in the PN subtype in TCGA GBM dataset (n=290, p=1.92e-19). **(B)** Expression levels of miR-34a in murine wild-type E13 embryonic neural stem cells (nsc), *p53/Pten-/-* E13 embryonic neural stem cells and *p53/pten-/-* malignant PN-like spheroids by RT-qPCR. Error bars represent standard deviation of experimental replicates. **(C-D)** Acute inactivation of *p53* and *Pten* in primary astrocytes isolated from 3 weeks old *p53L/L; PtenL/L* mice. Passage 4 cells were infected with a lentiviral vector carrying the Cre recombinase or a GFP control. Cells were harvested at 96 hours and genomic DNA, total RNA and protein were isolated for analysis. miR-34a expression was significantly decreased after ablation of *p53* and *Pten*. Abbreviations: PN=Proneural, MS=Mesenchymal, CL=Classical, NL=Neural, NSC=neural stem cell, TS=tumor spheroids.

Supplementary Figure 6. miR-34a impairs tumor formation in orthotopic transplants (Supporting information to Figure 4). **(A)** H&E staining of mouse brains ortothopically transplanted with *p53/Pten-/-* PN-like glioma cells harboring miR-34a or the control vector. Arrows indicate the invasive edge of the tumors. **(B)** Expression levels of miR-34a in parental cells 72h after infection and in short term cultures established from control and miR-34a expressing tumors. Tumors arising under the selective pressure of miR-34a overexpression show lower levels of the miR compared to the parental cells.

Supplementary Figure 7. miR-34a impairs the self-renewal of human proneural malignant spheroids. (A) Re-introduction of miR-34a in human PN Glioblastoma spheroids (TS543) decreases their self-renewal potential. Error bars represent extreme values (upper graph) and standard deviation of experimental triplicates (lower graph). **(B)**
Expression levels of miR-34a measured by RT-qPCR 72h after infection with miR-34a or ct vector. Error bars represent Standard Deviation of replicates.

Supplementary Figure 8. miR-34a loss promotes tumorigenesis in vivo and induces the expression of its oncogenic targets (supporting data to Figure 5). (A) Expression levels of miR-34a in immortalized mouse astrocytes infected with the decoy or the control vector assessed by RT-qPCR 72h after the infection. Error bars represent the Standard deviation of replicates. **(B)** Western blots showing the expression levels of the oncogenic targets of miR-34a in E6/E7T cells transduced with the miR-34a decoy or the control vector. Abbreviations: IAP= *Ink4a/Arf-/-;Pten-/-* ; IAPVIII= *Ink4a/Arf-/-Pten-/-;EgfrVIII/+*.

Supplementary Figure 9. PDGFRA is a direct target of miR-34a. (A) 3'UTR luciferase reporter assays showing that PDGFRAis a direct target of miR-34a. Experiments were done in E6/E7T cells transiently transfected with a reporter construct containing the PDGFRA 3'UTR reporter or the control reporter lacking the miR-34a binding site along with the pre-miR-34a expressing vector, the decoy vector and the appropriate controls. Luciferase activity was measured after 48h and normalized to the internal control (Renilla). Error bars represent experimental triplicates. * $P< 0.05$, ** $P<0.01$, using the two-tailed Student's t-test. **(B)** Expression levels of PDGFRA in human and mouse Glioblastoma cell lines upon miR-34a expression. Cell lysates

were collected 72h after the infection with the miR-34a expressing vector or the control vector. Abbreviations: R.L.A.= Relative Luciferase Activity

Supplementary Figure 10. PDGFRA rescues the reduced self-renewal phenotype caused by miR-34a over-expression (supporting data to Figure 6B). (A) In-vitro assays showing that PDGFRA rescues the impaired self-renewal phenotype associated with miR-34a expression in mouse PN-like malignant glioma (TSG2). **(B)** In-vitro assays showing that PDGFRArescues the impaired self-renewal phenotype associated with miR-34a expression in human PN GBM spheroids (TS543). Error bars represent standard deviation of experimental triplicates. **(C)** Expression levels of PDGFRA measured by Western Blot. Cell lysates were collected 72-96h after the infection with the appropriate vectors.

Supplementary Figure 11. Inverse Correlation between the expression levels of miR-34a, PDGFRA and Smad4 in the GBM TCGA dataset. Heat-map showing the expression levels of miR-34a and its targets Pdgfra and Smad4 in the GBM TCGA dataset (n=290). Samples were clustered respective to miR-34a expression levels. A significant inverse correlation was found between the expression levels of miR-34a and both Pdgfra (p= 9.324e-13) and Smad4 (p= 1.068e-07).

Supplementary Figure 12. Smad4 is a direct target of miR-34a. 3'UTR luciferase reporter assays showing that Smad4 is a direct target of miR-34a. Experiments were done in E6/E7T cells transiently transfected with a reporter construct containing the Smad4 3'UTR reporter or the control reporter lacking the miR-34a binding site along with the pre-miR-34a expressing vector,

the decoy vector and the appropriate controls. Luciferase activity was measured after 48h and normalized to the internal control (Renilla). Error bars represent experimental triplicates. * P< 0.05, ** P<0.01, using the two-tailed Student's t-test. Abbreviations: R.L.A.= Relative Luciferase Activity

Supplementary Figure 13. Smad4 rescues the impaired self-renewal phenotype caused by miR-34a over-expression (supporting data to Figure 6C). In-vitro assays showing that Smad4 rescues the impaired self-renewal phenotype associated with miR-34a expression in mouse PN-like malignant glioma (TSG2). Error bars represent Standard Deviation of experimental triplicates.

Supplementary Tables Legend

Supplementary Table 1. List of 290 TCGA samples used in the study.

Supplementary Table 2. List of miR-mRNA edges identified by CLR. The correlations are computed using the 290 samples data set.

Supplementary Table 3. List of putative direct miR-mRNA edges in the network.

Supplementary Table 4. miR-mRNA edges identified as unique to each of the 4 molecular subtypes. (A) Classical (B) Neural (C) Proneural (D) Mesenchymal.

Supplementary Table 5. miR and mRNA nodes in regions of copy number aberration in the sub-network corresponding to the 8 discriminatory miRs.

Supplementary Table 6. Survival analysis results for the 8 discriminatory miRs. mir-34a expression is strongly associated with survival (univariate p-val = 0.002036, dichotomized logrank p-val = 2.2e-05).

Supplementary Table 7. Multivariate Cox regression analysis to correct the possible confounding effect of selected clinical variables in the entire TCGA dataset and in the 4 molecular subclasses.

Supplementary Table 8. Patients' characteristics (validation cohort, n=220).

Supplementary Table 9. AQUA values (raw data) for miR-34a expression in Glioblastoma TMAs (validation cohort).

Supplementary Table 10. List of transcription factors enriched in the PN subtype. Three transcription factors were identified that were (i) significantly overrepresented in PN vs MS samples (ii) significantly overexpressed in PN vs MS (iii) had statistically significant negative correlations (< -0.3) with mir-34a in PN samples.