Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2014 May 06.

Published in final edited form as:

Nature. 2010 January 21; 463(7279): 318–325. doi:10.1038/nature08712.

The transcriptional network for mesenchymal transformation of brain tumors

Maria Stella Carro[1,†,∏], **Wei Keat Lim**[2,3,*,†], **Mariano Javier Alvarez**[3,4,†], **Robert J. Bollo**[5], **Xudong Zhao**[1], **Evan Y. Snyder**[6], **Erik P. Sulman**[7], **Sandrine L. Anne**[1,†], **Fiona Doetsch**[10], **Howard Colman**[8], **Anna Lasorella**[1,10,11], **Ken Aldape**[9], **Andrea Califano**[1,2,3,4,¶], and **Antonio Iavarone**

[1]Institute for Cancer Genetics, Columbia University Medical Center, New York, NY, 10032

[2]Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, 10032

[3]Center for Computational Biology and Bioinformatics, Columbia University Medical Center, New York, NY, 10032

[4]Joint Centers for Systems Biology, Columbia University Medical Center, New York, NY, 10032

[5]Department of Neurosurgery, New York University School of Medicine & NYU Langone Medical Center, New York, NY, 10016

[6]Burnham Institute for Medical Research, La Jolla, CA 92037

[7]Division of Radiation Oncology, M.D. Anderson Cancer Center, Houston, Texas 77030

[8]Department of Neuro-Oncology, M.D. Anderson Cancer Center, Houston, Texas 77030

[9]Department of Pathology, M.D. Anderson Cancer Center, Houston, Texas 77030

[10]Department of Pathology, Columbia University Medical Center, New York, NY, 10032

[11]Department of Pediatrics, Columbia University Medical Center, New York, NY, 10032

[12]Department of Neurology, Columbia University Medical Center, New York, NY, 10032

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

[¶]To whom correspondence should be addressed. califano@c2b2.columbia.edu (Andrea Califano), ai2102@columbia.edu (Antonio Iavarone)..

[†]These authors contributed equally.

[∏]Present address: Department of General Neurosurgery, Neurocenter and Comprehensive Cancer Research Center, University of Freiburg, Breisacher Str. 64, D-79106 Freiburg, Germany

[*]Present address: Therasis Inc., 462 First Avenue, Suite 908, New York, NY 10016.

[‡]Present address: Rockefeller University, RRB 750, 1230 York Avenue, New York, NY 10065.

Author Contributions A.C. and A.I. conceived the ideas for this study. A.C. designed the computational systems biology approach and A.I. the experimental platform. M.S.C. prepared constructs, performed the biochemical experiments and the microarrays, conducted biological experiments and analyses, assisted in mouse intracranial injections and performed tumor xenograft immunohistochemistry and tumor analysis. W.K.L. performed reverse engineering, master regulator, and statistical analyses. M.J.A. conducted gene expression, bioinformatics, and statistical analyses. R.J.B. and E.Y.S. provided experimental material. X.Z. and F.D. assisted in mouse intracranial injections. E.P.S., H.C. and K.A. provided reagents, performed the arrayCGH/expression analysis and primary human tumor immunohistochemistry. S.L.A. performed cell culture immunofluorescence microscopy and analysis. A.L. assisted in primary NSC experiments, performed intracranial injection and assisted in the analysis of mouse xenografts. A.I. and A.C. wrote the manuscript with contributions from all other authors. M.S.C., W.K.L. and M.J.A. contributed equally to this work.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Abstract

Inference of transcriptional networks that regulate transitions into physiologic or pathologic cellular states remains a central challenge in systems biology. A mesenchymal phenotype is the hallmark of tumor aggressiveness in human malignant glioma but the regulatory programs responsible for implementing the associated molecular signature are largely unknown. Here, we show that reverse-engineering and unbiased interrogation of a glioma-specific regulatory network reveal the transcriptional module that activates expression of mesenchymal genes in malignant glioma. Two transcription factors (C/EBPβ and Stat3) emerge as synergistic initiators and master regulators of mesenchymal transformation. Ectopic co-expression of C/EBPβ and Stat3 reprograms neural stem cells along the aberrant mesenchymal lineage whereas elimination of the two factors in glioma cells leads to collapse of the mesenchymal signature and reduces tumor aggressiveness. In human glioma, expression of C/EBPβ and Stat3 correlates with mesenchymal differentiation and predicts poor clinical outcome. These results reveal that activation of a small regulatory module is necessary and sufficient to initiate and maintain an aberrant phenotypic state in cancer cells.

High-grade gliomas (HGGs) are the most common brain tumors in humans and are essentially incurable[1]. The defining hallmarks of aggressiveness of glioblastoma multiforme (GBM) are local invasion and neo-angiogenesis[2, 3]. A recently established notion postulates that neoplastic transformation in the central nervous system (CNS) converts neural cells into cell types manifesting a mesenchymal phenotype, a state associated with uncontrolled ability to invade and stimulate angiogenesis[4, 5]. Gene expression studies have established that over-expression of a “mesenchymal” gene expression signature (MGES) and loss of a proneural signature (PNGES) co-segregate with the poor prognosis group of glioma patients[4]. Yet, differentiation along the mesenchymal lineage is virtually undetectable in normal neural tissue during development. Thus, it is unclear whether drift toward the mesenchymal lineage is an aberrant event that occurs during brain tumor progression or whether glioma cells recapitulate the rare mesenchymal plasticity of neural stem cells (NSCs)[4–7]. The molecular events that activate the MGES while suppressing the PNGES signature, thus imparting a highly aggressive phenotype to glioma cells, remain unknown.

Efforts to identify TFs that are Master Regulators (MRs) of specific cancer signatures, based on cellular-network models, have yet to produce experimentally validated discoveries, likely because these networks are still poorly mapped, especially within specific mammalian cellular contexts[8]. Notwithstanding, recent developments in genome-wide reverse engineering were successful in identifying causal, rather than associative interactions[9–12], and showed promise in the identification of dysregulated genes within developmental and tumor-related pathways[13–17]. Thus, we reasoned that context-specific regulatory networks, inferred by unbiased reverse engineering algorithms may provide sufficient accuracy to allow estimating (a) the activity of TFs from that of their transcriptional targets or *regulons* and (b) the identity of TFs that are MRs of specific eukaryotic signatures[18, 19] from the overlap between their regulons and the signatures. We applied the above paradigms to unravel the MRs causally linked to activation of the MGES in malignant glioma (Supplementary Fig. 1).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A transcriptional module is linked to the MGES of HGGs

We first asked whether copy number variation may account for the aberrant expression of MGES genes in HGGs. Integrated analysis of gene expression profiles and array comparative genomic hybridization (aCGH) of 76 HGGs showed no correlation between mean expression and DNA copy number of MGES genes in proneural, mesenchymal, and proliferative tumors (Supplementary Fig. 2).

We thus used the ARACNe algorithm[9] to assemble a genome-wide repertoire of HGGs-specific transcriptional interactions (the HGG-interactome) from 176 gene expression profiles of grade III (anaplastic astrocytoma) and grade IV (GBM) samples previously classified into three molecular signature groups – proneural, proliferative, and mesenchymal (Supplementary Table 1a–c)[4, 20, 21]. ARACNe is an information theoretic approach for the inference of TF-target interactions from large sets of gene expression profiles[9, 16], further refined to infer directed (i.e. causal) interactions[12, 22] (see Methods). ARACNe predicted 92,660 transcriptional interactions, 1,217 of which were between TFs and 102 of 149 MGES genes[4], represented across all the gene expression profile data.

Next, we applied a novel Master Regulator Analysis (MRA) algorithm to the HGG-interactome. The algorithm computes the statistical significance of the overlap between the regulon of each TF (i.e., its ARACNe-inferred targets) and the MGES genes (*p*-values computed by Fisher Exact Test, FET). From a list of 928 TFs (Supplementary Table 2), MRA inferred 53 MGES-specific TFs, at a *False Discovery Rate* (FDR) < 5% (Supplementary Table 3a). These were ranked based on the total number of MGES targets they regulated. The top six TFs (Stat3, C/EBP, bHLH-B2, Runx1, FosL2, and ZNF238) collectively controlled >74% of the MGES genes (Fig. 1a). C/EBPβ and C/EBPδ were grouped as they form stoichiometric homo and heterodimers with identical DNA binding specificity and redundant transcriptional activity[23]. We thus use the term C/EBP to indicate the TF complex with the union of their targets as the corresponding regulon. Consistent with their previously reported activity[24, 25], Spearman correlation analysis revealed that five of these TFs are likely activators (Stat3, C/EBP, bHLH-B2, Runx1, and FosL2) and one is likely a repressor (ZNF238). Overlap between the regulons of these TFs was highly significant (Supplementary Table 4). MRA analysis of the PNGES and Proliferative (PROGES) signatures of HGGs detected virtually no overlap among candidate MRs of the three signatures, with the notable exception of few TFs inversely associated with MGES and PNGES activation (Supplementary Table 5).

Next, we used stepwise linear regression (SLR) to infer simple, quantitative regulation models for each MGES gene. Specifically, the log-expression of each MGES gene was fitted by the linear combination of the log-expression of a small number of genes (1 – 5) (see Methods), selected among 53 ARACNe-inferred and 52 additional TFs, whose DNA-binding signatures were enriched in promoters of MGES genes. Six TFs were in both lists, for a total of 99 TFs (Supplementary Table 3b). TFs were then ranked based on the fraction of MGES genes they regulated. Surprisingly, the top six MRA-inferred TFs were also among the eight controlling the largest number of MGES targets, based on SLR analysis (Supplementary Table 6). Indeed, the three with the highest linear-regression coefficient

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

values were C/EBP (α = 0.40), bHLH-B2 (α = 0.41), and Stat3 (α = 0.40), further establishing them as likely MGES-MR candidates. The next strongest TF, ZNF238, had a negative coefficient (α = −0.34) confirming its role as a candidate MGES repressor.

Validation of the mesenchymal regulatory module

To determine whether these TFs bind the promoter region of their predicted MGES targets, we performed chromatin immunoprecipitation (ChIP) in a human glioma cell line. On average, TF-specific antibodies but not control antibodies immunoprecipitated 80% of the tested genomic regions (Fig. 1b–f). Lentivirus-mediated shRNA silencing of C/EBPβ, Stat3, bHLH-B2, FosL2, and Runx1 in glioma cells, followed by gene expression profiling and Gene Set Enrichment Analysis (GSEA) revealed that, following silencing of each TF, differentially expressed genes were highly enriched in their ARACNe-inferred targets but not in those of control TFs with equivalent regulon size (Supplementary Table 7a). Furthermore, differentially expressed genes were also enriched in MGES genes (Supplementary Table 7b).

Promoter occupancy analysis revealed a hierarchical and highly modular topology, with 8 of 10 possible intra-module interactions implemented (modularity p-value = 1.0×10^{-8} by FET, Fig. 2c). Specifically, C/EBPβ and Stat3 occupy their own promoter (Fig. 2a, b); C/EBPβ occupies the Stat3, FosL2, bHLH-B2, C/EBPβ, and C/EBPδ promoters (Fig. 2a); Stat3 occupies those of FosL2 and Runx1 (Fig. 2b); FosL2 occupies those of Runx1 and bHLH-B2 (Supplementary Fig. 3a) and bHLH-B2 occupies only the promoter of Runx1 (Supplementary Fig. 3b). C/EBP and Stat3 are at the top of this hierarchical regulatory module. They have autoregulatory loops and form feed-forward loops with a larger fraction of MGES genes (43%) than any of the other TF-pairs. ShRNA-mediated co-silencing of C/EBPβ and Stat3 in glioma cells produced >2-fold reduction of the mRNAs coding for the second layer TFs in the FF loops (bHLH-B2, FosL2, and Runx1, Fig. 2d), supporting their role as MRs. C/EBPβ and Stat3 bound the promoters of their MGES targets also in primary human GBM (Supplementary Fig. 3c, d).

To functionally validate the role of C/EBPβ and Stat3 as MRs of the MGES, we conducted gain and loss-of-function experiments. We transduced *v-myc* immortalized mouse NSCs known as C17.2[26–28] as well as primary murine NSCs derived from the mouse telencephalon at embryonic day 13.5 with retroviruses expressing C/EBPβ and a constitutively active form of Stat3 (Stat3C)[29]. ShRNA-mediated silencing targeted C/EBPβ and Stat3 in the human glioma cell line SNB19 and in serum-free cultures of tumor cells derived from primary GBM that propel the formation of GBM-like tumors after intracranial transplantation in immunodeficient mice[30] (GBM-derived brain tumor initiating cells, GBM-BTICs, see Fig. 5c–g). We generated a global dataset of 89 individual samples, including 55 knockdown experiments in human glioma cells and 34 ectopic expression experiments in mouse NSCs. Of the 149 genes in the MGES, 118 could be mapped to murine genes represented on the array (Supplementary Table 8). Quantitative RT-PCR (qRT-PCR) analysis showed that, following C/EBPβ or Stat3 shRNA silencing in GBM-BTICs and SNB19, the corresponding mRNA levels were significantly reduced compared to non-target control transduced cells (*C/EBPβ* fold ratio = 0.26, p ≤ 0.00108, *Stat3* fold ratio =

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

0.205, $p \leq 0.00109$ by U-test). Reciprocal changes followed ectopic expression of the two TFs in C17.2 and NSCs (Supplementary Table 9). QRT-PCR values and microarray-based measurements were highly correlated for Stat3 but not for C/EBPβ mRNA (Supplementary Fig. 4). Thus, we used the qRT-PCR values for C/EBPβ and Stat3 as more accurate read-outs for their mRNA expression. GSEA analysis confirmed that genes co-expressed with the two TFs were enriched in their respective ARACNe-inferred regulon genes but not in those of control TFs (Supplementary Table 10). Perturbation of C/EBPβ (Supplementary Fig. 5a, c) or Stat3 (Supplementary Fig. 5b, d) specifically affected the MGES signature ($p = 2.69\times10^{-2}$ and $p = 2.0\times10^{-4}$, respectively by GSEA). Common targets of C/EBP and Stat3 were 8-fold more enriched in MGES genes than targets controlled individually by each TF (Fig. 2e, $p = 2.25\times10^{-5}$). To test whether the two TFs may be involved in synergistic MGES control, we computed a metagene (C/EBPβ×Stat3) with expression proportional to the product of their mRNAs, such that the metagene should be highly correlated with the expression profile of any target synergistically regulated by the two TFs, under a multiplicative model (Fig. 2f). GSEA analysis confirmed that genes ranked by Spearman correlation to the C/EBPβ×Stat3 metagene were significantly enriched in MGES genes (Fig. 2g), suggesting that at least a subset of the MGES is synergistically regulated by the C/EBPβ×Stat3 pair.

We sought to establish whether (a) MRs inferred by our procedure would also be inferred when using an independent glioma sample datasets and (b) MRs identified on the basis of clinical outcome would overlap significantly with those inferred from MGES analysis. The Atlas-TCGA dataset[31] includes 77 and 21 samples associated with worst- and best-prognosis, respectively (92 samples with intermediate prognosis were not considered). Differential expression analysis identified a *TCGA Worst-Prognosis Signature* (TWPS), comprising 884 genes differentially expressed in the worst-prognosis vs. best-prognosis samples (at $p \leq 0.05$ by Student's t-test, Supplementary Table 11). GSEA analysis confirmed that MGES genes were markedly enriched in the TWPS signature ($p \leq 1.0\times10^{-4}$, Supplementary Fig. 6) indicating that the poor-prognosis group in the Atlas-TCGA dataset displays marked mesenchymal features. Despite partial overlap between MGES and TWPS genes (22.8%), 5 of the 6 MRs identified by MRA analysis from the original dataset were also found among the 10 most significant TFs identified by MRA analysis of the Atlas-TCGA dataset using the TWPS signature. C/EBP was the most significant TF, while Stat3 was in 7th position. C/EBPβ and C/EBPδ had the first and second best linear-regression coefficient by SLR analysis, respectively (Supplementary Table 12). These results indicate significant robustness of the approach both to dataset and signature selection.

Concurrent expression of C/EBPβ and Stat3 reprograms NSCs toward the mesenchymal lineage

We tested whether combined and/or individual expression of Stat3C and C/EBPβ in NSCs is sufficient to trigger the mesenchymal phenotype that characterizes HGGs. Introduction of C/EBPβ and Stat3C in C17.2 NSCs caused loss of neuronal differentiation and manifestation of a fibroblast-like morphology (Supplementary Fig. 7a, b). The morphological changes were associated with gain of the expression of the mesenchymal marker proteins SMA and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

fibronectin (Fig. 3a, Supplementary Fig. 7c, d) and induced expression of the mesenchymal genes *Chi3l1/YKL40, Acta2/SMA, CTGF* and *OSMR* (Fig. 3b). The individual expression of Stat3C or C/EBPβ was generally insufficient to induce either mesenchymal marker proteins or expression of mesenchymal genes (Fig. 3a, b). Removal of mitogens to Stat3C/C/EBPβ-expressing C17.2 cells resulted in further increase of the expression of mesenchymal genes and complete acquisition of mesenchymal properties such as positive alcian blue staining, a specific assay for chondrocyte differentiation (Fig. 3c, Supplementary Fig. 7e, f). The expression of Stat3C and C/EBPβ promoted migration in a wound assay and triggered invasion through the extracellular matrix in a Matrigel invasion assay in the absence or presence of PDGF (Fig. 3d, e, Supplementary Fig. 7g). The combined but not individual expression of Stat3C and C/EBPβ efficiently induced mesenchymal marker proteins and mesenchymal gene expression also in primary NSCs (Fig. 3f–h). Conversely, Stat3C and C/EBPβ abolished differentiation along the neuronal and glial lineages (Fig. 3i, Supplementary Fig. 7h). The C/EBPβ/Stat3C–induced mesenchymal transformation of primary NSCs was associated with withdrawal from cell cycle (data not shown). Thus, the combined introduction of C/EBPβ and Stat3C in NSCs prevents neural differentiation and triggers reprogramming toward an aberrant mesenchymal lineage.

C/EBPβ and Stat3 are essential for mesenchymal transformation and glioma aggressiveness

Transduction of GBM-BTICs cultures derived from two GBM patients (BTSC-20 and BTSC-3408) with specific shRNA-carrying lentiviruses silenced endogenous C/EBPβ and Stat3, eliminated expression of mesenchymal genes and depleted the tumor cells of the mesenchymal marker proteins fibronectin, collagen-5A1 and YKL40. Individual silencing of C/EBPβ or Stat3 produced variable inhibitory effects with the silencing of C/EBPβ typically carrying the most severe consequences (Fig. 4a–e, Supplementary Fig. 8a). Combined or individual silencing of C/EBPβ and Stat3 in the human glioma cell line SNB19 produced similar effects (Supplementary Fig. 8b–e). Silencing of the two TFs in SNB19 and GBM-BTICs reduced by >70% their ability to invade through Matrigel (Fig. 4f–i). Next, we determined the impact of C/EBPβ and Stat3 knockdown on brain tumorigenesis after intracranial injection of SNB19 in immunocompromised mice. We observed efficient tumor formation in all mice injected with shcontrol and shStat3 cells. However, only one of four mice from the shC/EBPβ and one of five mice from the shC/EBPβ+shStat3 groups developed tumors after 120 days from the injection (Fig. 5b). The histologic analysis demonstrated high-grade tumors, which displayed peripheral invasion of the surrounding brain as single cells and cell clusters in the shRNA control group as shown by anti-human vimentin staining (Fig. 5a). Staining for the endothelial marker CD31 revealed marked vascularization in the shRNA control group of tumors. Conversely, the single tumor in the shC/EBPβ+shStat3 group grew well circumscribed and was less angiogenic. Tumors in the shStat3 group and the single tumor in the shC/EBPβ group had an intermediate growth pattern and limited angiogenesis. Staining for fibronectin, collagen-5A1 and YKL40 were readily detected in the tumors from the control group but absent or barely detectable in the single tumors from the shC/EBPβ and shC/EBPβ+shStat3 groups (Fig. 5a). Tumors derived from shStat3 cells displayed an intermediate phenotype with reduced expression of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mesenchymal markers compared with tumors in the shcontrol group but higher than that observed in the tumors in the shC/EBPβ and shC/EBPβ+shStat3 groups (shcontrol > shStat3 > shC/EBPβ> shC/EBPβ+shStat3). Intracranial transplantation of GBM-BTICs transduced with shRNA control lentivirus produced extremely invasive tumor cell masses extending through the corpus callosum to the contralateral brain. Combined knockdown of C/EBPβ and Stat3 led to significant decrease of the tumor area and tumor cell density as evaluated by human vimentin staining, markedly reduced the proliferation index (Fig. 5c–e), and abolished expression of the mesenchymal markers fibronectin and collagen-5A1 (Fig. 5f, g).

Finally, we conducted an immunohistochemical analysis for C/EBPβ and active, phospho-Stat3 in human tumor specimens, and compared their expression with that of YKL40 (a well-established mesenchymal protein expressed in primary human GBM)[21, 32] and with patient outcome in a collection of 62 GBMs (Supplementary Fig. 9). Expression of either C/EBPβ or Stat3 was significantly associated with YKL40 expression (C/EBPβ, $p = 4.9\times10^{-5}$; Stat3, $p = 2.2\times10^{-4}$), with higher association in double positive tumors (C/EBPβ+/Stat3+, $p = 2.7\times10^{-6}$) vs. double negative ones (C/EBPβ-/Stat3-, Supplementary Table 13). Double positive tumors were associated with worse clinical outcome than either single or double negative tumors (log-rank test, p=0.0002, Fig. 5h). Positivity for either of the two TFs remained predictive of negative outcome but with lower statistical strength than double positivity (C/EBPβ, p=0.0022; Stat3, p=0.0017).

Discussion

We have shown that inference of context-specific regulatory network identifies the transcriptional module that controls expression of the mesenchymal signature associated with poor-prognosis in HGGs. In this approach, the traditional paradigm of gene expression profile based cancer research, yielding long lists of differentially expressed genes (i.e., cancer signatures), becomes only a starting point for a cellular-network analysis, where a causal regulatory model identifies the TFs controlling the signatures and related phenotypes.

Recently, there have been several unsuccessful attempts to identify common expression signatures predictive of the same cellular phenotype[33]. Our approach produced virtually identical regulatory MR modules when applied to two completely distinct datasets and signatures associated with poor-prognosis in HGGs, thus indicating that MRs of mammalian phenotype signatures may be significantly more conserved than the complement of differentially expressed genes. Other methods, including differential expression analysis, DNA-binding-site enrichment analysis[8] and relevance network analysis[34] could not identify C/EBPβ and Stat3 as MRs (see Supplementary Note 2). This suggests that enrichment analysis of ARACNe-inferred TF regulons is specifically useful for the identification of MRs of cellular phenotypes. Our results do not exclude that other graph-theoretical methods such as Bayesian Networks might provide further fine-grain regulatory insight once the number of candidate MRs is reduced to a handful by methods such as those proposed here. Yet, once a relatively small number of TFs is identified, direct experimental validation is feasible.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The experimental follow-up established that C/EBPβ and Stat3 are MRs sufficient in NSCs and necessary in human glioma cells for mesenchymal transformation. Interestingly, C/EBPβ and Stat3 are expressed in the developing nervous system[35–38]. However, while Stat3 induces astrocyte differentiation and inhibits neuronal differentiation of neural stem/progenitor cells, C/EBPβ promotes neurogenesis and opposes gliogenesis[39–41]. How can the combined activity of C/EBPβ and Stat3 reprogram NSCs toward an aberrant lineage (mesenchymal) and oppose the genesis of the normal neural lineages (neuronal and glial)? We propose that mesenchymal transformation results from concurrent activation of two conflicting transcriptional regulators normally operating to funnel opposing signals (neurogenesis vs. gliogenesis). This scenario is intolerable by normal neural stem/progenitor cells whereas it operates to permanently drive the aberrant mesenchymal phenotype in the context of the genetic and epigenetic changes that accompany high-grade gliomagenesis (EGFR amplification, PTEN loss, Akt activation, etc.)[4]. Since expression of C/EBPβ and Stat3 in human glioma is essential to maintain the tumor initiating capacity and the ability to invade the normal brain, the two TFs provide important clues for diagnostic and pharmacological intervention. Consistent with this, the combined expression of C/EBPβ and Stat3 is linked to the mesenchymal state of primary GBM and provides an excellent prognostic biomarker for tumor aggressiveness. Thus, systems biology methods can be effectively used to infer MRs that choreograph malignant transformation. This paradigm will be applicable to the dissection of other phenotypic states.

METHODS SUMMARY

Cell culture

Primary NSCs were isolated from E13.5 mouse telencephalon and cultured in the presence of FGF-2 and EGF as described[42]. Differentiation was induced by culturing NSCs in NSC medium without EGF and FGF-2. GBM-derived BTICs were grown in Neurobasal media supplemented with FGF-2 and EGF.

Generation of transcriptional network, microarrays and qRT-PCR

GBM transcriptional network was generated by ARACNe[12]. Total RNA was reverse transcribed to complementary DNA and amplified using primers specific for human and murine transcripts. Expression values were calculated relative to the β-actin gene. RNA was used for analysis on Illumina HumanHT-12v3 or MouseWG-6 expression BeadChip. Sample information is in Methods.

Master Regulator Analysis

For each TF, the statistical significance of the intersection between the TF-regulon and the gene expression signature was computed by Fisher Exact Test. Significant genes were ranked based on the number of overlapping genes.

Gene Set Enrichment Analysis (GSEA)

The statistical significance of the enrichment of a ranked list of genes in a smaller set of genes was determined as described[43].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Stepwise Linear Regression (SLR)

The regulatory model of each gene was determined by identifying the smallest number of TFs that were informative for the expression of that gene across the dataset. TFs were added to the model one at the time, until the error reduction produced by adding another TF was no longer statistically significant. Models had on average 1 to 5 TFs.

Intracranial injection of glioma cells

SNB19 glioma cell line and GBM-derived BTICs were injected into the brain of 6–8 weeks NOD/SCID 48 h after infection with lentiviruses carrying shRNAs using a stereotaxic frame. Animals were monitored and euthanized when they presented with signs of tumor. Mouse research was approved by the Committee for Animal Care, and conducted in compliance with the Animal Welfare Act Regulations.

Methods

Array comparative genomic hybridization expression correlation

The correlation between gene expression and DNA copy number for the MGES genes was determined using data from 76 high-grade gliomas for which both gene expression array and array comparative genomic hybridization (aCGH) profiling were performed[4]. Tumors were grouped based on molecular subtype (mesenchymal, proneural, or proliferative) and the mean expression for MGES genes were determined in each group. The normalized copy number of each gene was interpolated based on the copy number of the nearest genomic clone on the CGH array as determined by comparison of the sequence annotation of both array platforms, as previously described[21].

ARACNe network reconstruction

ARACNe (Algorithm for the Reconstruction of Accurate Cellular Networks), an information-theoretic algorithm for inferring transcriptional interactions, was used to identify a repertoire of candidate transcriptional regulators of the MGES genes. Expression profiles used in the analysis were previously characterized using Affymetrix HU-133A microarrays and preprocessed by MAS 5.0 normalization procedure[4]. First, candidate interactions between a TF (x) and its potential target (y) are identified by computing pairwise mutual information, $MI[x; y]$, using a Gaussian kernel estimator[12] and by thresholding the mutual information based on the null-hypothesis of statistical independence ($p < 0.05$ Bonferroni corrected for the number of tested pairs). Then, indirect interactions are removed using the data processing inequality, a well-known property of the mutual information. For each TF-target pair (x, y) we considered a path through any other TF (z) and remove any interaction such that $MI[x; y] < \min(MI[x; z], MI[y; z])$.

Transcription Factor classification

To identify human transcription factors (TFs), we selected the human genes annotated as “transcription factor activity” in Gene Ontology and the list of TFs from TRANSFAC. From this list, we removed general TFs (e.g. stable complexes like polymerases or TATA-box-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

binding proteins), and added some TFs not annotated by GO, producing a final list of 928 TFs that were represented on the HU-133A microarray gene set.

Master Regulator Analysis

The MRA has two steps. First, for each TF its signature-enrichment is computed as the p-value of the overlap between the TF-regulon and the signature genes (i.e. the MGES genes in this case), assessed by Fisher Exact Test (FET). Since FET depends on regulon size, it can be used to assess signature-enriched TFs but not to rank them. TFs are thus ranked based on the total number of signature genes included in their regulon, under the assumption that TFs controlling a larger fraction of the signature will be more likely to determine its activity.

Stepwise Linear Regression (SLR) Analysis

A regulatory program for each MGES gene was computed as follows: the $\log_2$ expression of the *i*-th MGES gene was considered as the response variable and the $\log_2$-expression of the TFs as the explanatory variables in the linear model $\log x_i = \sum \alpha_{ij} \log f_j + \beta_{ij}$ [44]. Here, f_j represents the expression of the *j*-th TF in the model and the $(\alpha_{ij}, \beta_{ij})$ are linear coupling coefficients computed by standard regression analysis. TFs are iteratively added to the model, by choosing each time the one producing the smallest relative error $E = \sum |x_i - x_{i0}|/x_{i0}$ between predicted and observed target expression. This is repeated until the decrease in relative error is no longer statistically significant, based on permutation testing. To avoid excessive multiple hypothesis testing correction, TFs were chosen only among the following: (a) the 53 inferred by ARACNe at FDR < 0.05 and (b) TFs whose DNA binding signature was significantly enriched in the proximal promoter of the MGES genes and that are expressed in the dataset, based on the coefficient of variation (CV ≥ 0.5). TFs were then ranked based on the number of MGES target they regulated, with the average Linear-Regression coefficient providing additional insight. The log-transformation allows convenient linear representation of multiplicative interactions between TF activities[44, 45]. TFs were individually added to the model, each time selecting the one contributing the most significant reduction in relative expression error (predicted vs. observed), until error-reduction was no longer significant.

Enrichment analysis

The false discovery rates (FDR) are computed using procedures described by Benjamini and Hochberg[46], where the adjusted p-values, q=p*n/i (p=p-value, n=total number of tests, i=sorted rank of p-value). It is a less conservative procedure to correct for multiple comparisons than familywise error rate (FWER), especially when the number of tests is large.

Cell lines and cell culture conditions

SNB75, SNB19, 293T and Phoenix cell lines were grown in DMEM plus 10% Fetal Bovine Serum (FBS, Gibco/BRL). GBM-derived BTICs were grown as neurospheres in Neurobasal media (Invitrogen) containing N2 and B27 supplements (Invitrogen), and human recombinant FGF-2 and EGF (50 ng/ml each; Peprotech,). Murine neural stem cells (mNSCs) (from an early passage of clone C17.2) (*27–29*) were cultured in DMEM plus 10%

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

heat inactivated FBS, (Gibco/BRL), 5% Horse serum (Gibco/BRL) and 1% L-Glutamine (Gibco/BRL). Neuronal differentiation of mNSCs was induced by growing cells in DMEM supplemented with 0.5% Horse serum. For chondrocyte differentiation, cells were treated with STEMPRO chondrogenesis differentiation kit (Gibco/BRL) for 20 days. Primary mNSCs were isolated from E13.5 mouse telencephalon and cultured in the presence of FGF-2 and EGF (20 ng/ml each) as described[42]. Differentiation of NSCs was induced by culturing neurospheres on laminin-coated dishes in NSC medium in the absence of growth factors. mNSC expressing Stat3C and C/EBPβ, were generated by retroviral infections using supernatant from Phoenix ecotropic packaging cells transfected with pBabe-Stat3C-FLAG and/or pLZRS-T7-His-C/EBPβ-2–IRES-GFP.

Promoter analysis and Chromatin immunoprecipitation (ChIP)

Promoter analysis was performed using the MatInspector software (www.genomatix.de). A sequence of 2kb upstream and 2kb downstream from the transcription start site was analyzed for the presence of putative binding sites for each TFs. Primers used to amplify sequences surroundings the predicted binding sites were designed using the Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and are listed in Supplementary Table 14.

ChIP was performed as described in Ref.[47]. SNB75 cell lysates were precleared with Protein A/G beads (Santa Cruz) and incubated at 4°C overnight with 1 μg of polyclonal antibody specific for C/EBPβ (sc-150, Santa Cruz), Stat3 (sc-482, Santa Cruz), FosL2 (Fra2, sc-604, Santa Cruz), bHLH-B2 (A300–649A, BETHYL Laboratories), or normal rabbit immunoglobulins (Santa Cruz). DNA was eluted in 200 μl of water and 1 μl was analyzed by PCR with Platinum Taq (Invitrogen). For primary GBM samples, 30 mg of frozen tissue was transferred in a tube with 1 ml of culture medium, fixed with 1% formaldehyde for 15 min and stopped with 0.125 M glycine for 5 min. Samples were centrifuged at 4000 rpm for 2 min, washed twice and diluted in PBS. Tissues were homogenized using a pestle and suspended in 3 ml of ice-cold immunoprecipitation buffer with protease inhibitors and sonicated. ChIP was then performed as described above.

QRT-PCR and microarray analysis

RNA was prepared with RiboPure kit (Ambion), and used for first strand cDNA synthesis using random primers and SuperScriptII Reverse Transcriptase (Invitrogen). QRT-PCR was performed using Power SYBR Green PCR Master Mix (Applied Biosystems). Primers are listed in Supplementary Table 15. QRT-PCR results were analyzed by the ΔΔCT method[48] using 18S as housekeeping gene.

RNA amplification for Array analysis was performed with Illumina TotalPrep RNA Amplification Kit (Ambion). 1.5 μg of amplified RNA was hybridized on Illumina HumanHT-12v3 (including 24,385 human genes) or MouseWG-6 (including 20,311 mouse genes) expression BeadChip according to the manufacturer's instructions. Hybridization data was obtained with an iScan BeadArray scanner (Illumina) and pre-processed by variance stabilization and robust spline normalization implemented in the lumi package

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

under the R-system[49]. Gene expression data have been deposited in GEO with the following accession numbers: GSE19113 for mouse and GSE19114 for human data.

Immunofluorescence and Immunohistochemistry

Immunofluorescence staining was performed as previously described[50]. Primary antibodies and dilutions were: SMA (mouse monoclonal, Sigma, 1:200), Fibronectin (mouse monoclonal, BD Biosciences, 1:200), Tau (rabbit polyclonal, Dako, 1:400), βIIITubulin (mouse monoclonal, Promega, 1:1000), CTGF (rabbit polyclonal, Santa Cruz, 1:200), YKL40 (rabbit polyclonal, Quidel, 1:200) and Col5A1 (rabbit polyclonal, Santa Cruz, 1:200). Confocal images acquired with a Zeiss Axioscop2 FS MOT microscope were used to score positive cells. At least 500 cells were scored for each sample. Quantification of the fibronectin intensity staining in mNSC was performed using NIH Image J software (http://rsb.info.nih.gov/ij/, NIH, USA). The histogram of the intensity of fluorescence of each point of a representative field for each condition was generated. The fluorescence intensity of three fields from three independent experiments was scored, standardized to the number of cells in the field and divided by the intensity of the vector. For immunostaining of xenograft tumors, mice were perfused trans-cardially with 4% PFA, brains were dissected and post-fixed for 48h in 4% PFA. Immunostaining was performed as previously described[51]. Primary antibodies and dilutions: fibronectin (mouse moclonal, BD Bioscences, 1;100), Col5A1 (rabbit polyclonal, Santa Cruz, 1:100), YKL40 (rabbit polyclonal, Quidel, 1;100), human vimentin (mouse monoclonal, Sigma, 1:50), Ki67 (rabbit polyclonal, Novocastra laboratories, 1:1000). Quantification of the tumor area was obtained by measuring the human vimentin positive area in the section using the NIH Image J software (http://rsb.info.nih.gov/ij/, NIH, USA). Five tumors for each group were analyzed. For quantification of Ki67, the percentage of positive cells was scored in 5 tumors per each group. In histogram values represents the mean values; error bars are standard deviations. Statistical significance was determined by t test (with Welch's Correction) using GraphPad Prism 4.0 software (GraphPad Inc., San Diego, CA). Immunohistochemistry of primary human GBM was performed as previously described[52]. The primary antibodies and dilutions were: anti-YKL40 (rabbit polyclonal, Quidel, 1:750), anti C/EBPβ, (rabbit polyclonal, Santa Cruz, 1:250) and anti-p-Stat3 (rabbit monoclonal, Cell Signaling, 1;25), Scoring for YKL40 was based on a 3-tiered system, where 0 was <5% of tumor cells positive, 1 was 5–30% positivity and 2 was >30% of tumor cells positive. Scores of 1 and 2 were later collapsed into a single value for display purposes on Kaplan-Meier curves. Associations between C/EBPβ/Stat3 and YKL40 were assessed using the Fisher exact test (FET). Associations between C/EBPβ/Stat3 and patients survival were assessed using the log-rank (Mantel-Cox) test of equality of survival distributions.

Migration and invasion assays

For the wound assay testing migration, mNSCs were plated in 60 mm dishes and grown until 95% confluence. A scratch of approximately 1000 μm was made with a P1000 pipet tip and images were taken every 24 h with an inverted microscope. For the Matrigel invasion assay, mNSCs and SNB19 (1×10^4) were added to the upper compartment of a 24 well BioCoat Matrigel Invasion Chamber (BD Bioscences) in serum free DMEM. The lower compartment of the chamber was filled with DMEM containing either 0.5% horse serum or

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

20 μg/ml PDGF-BB (R&D systems) as chemoattractant. After 24 h, invading cells were fixed, stained according to the manufacturer's instructions and counted. For GBM-derived BTICs, 5×10^4 cells were plated on the upper chamber in the absence of growth factors. In the lower compartment Neurobasal medium containing B27 and N2 supplements plus 20 μg/ml PDGF-BB (R&D systems) was used as chemoattractant.

Lentivirus infection

Lentiviral expression vectors carrying shRNAs were purchased from Sigma. The sequences are listed in Supplementary Table 16. To generate lentiviral particles, each shRNA expression plasmid was co-transfected with pCMV-dR8.91 and pCMV-MD2.G vectors into human embryonic kidney 293T cells using Fugene 6 (Roche). Lentiviral infections were performed as described[51].

Intracranial Injection

Intracranial injection of SNB19 glioma cell line and GBM-derived BTICs was performed in 6–8 weeks NOD/SCID mice (Charles River laboratories) in accordance with guidelines of the International Agency for Reserch on Cancer's Animal Care and Use Committee. Briefly, 48 h after lentiviral infection, 2×10^5 SNB19 or 3×10^5 BTICs were injected 2 mm lateral and 0.5 mm anterior to the bregma, 3 mm below the skull. Mice were monitored daily and sacrificed when neurological symptoms appeared. Kaplan–Meier survival curve of the mice injected with SNB19 glioma cells was generated using the DNA Statview software package (AbacusConcepts, Berkeley CA).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by National Cancer Institute grants R01CA109755 (A.C.), R01CA101644 (A.L.) R01CA085628 (A.I.), National Institute of Allergy and Infectious Diseases grant R01AI066116 (A.C.), National Centers for Biomedical Computing NIH Roadmap Initiative U54CA121852 (A.C.) and National Institute of General Medical Sciences grant P20GM075059 (E.Y.S.). M.S.C. is supported by a fellowship from the Italian Ministry of Welfare/Provincia di Benevento and S.L.A. by a fellowship from Fondation de Recherche Medicale. We thank Nadya Ramirez-Martinez for technical assistance with mouse husbandry and *in vivo* procedures.

References

1. Ohgaki H, Kleihues P. Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. J Neuropathol Exp Neurol. 2005; 64:479–489. [PubMed: 15977639]
2. Demuth T, Berens ME. Molecular mechanisms of glioma cell migration and invasion. J Neurooncol. 2004; 70:217–228. [PubMed: 15674479]
3. Kargiotis O, Rao JS, Kyritsis AP. Mechanisms of angiogenesis in gliomas. J Neurooncol. 2006; 78:281–293. [PubMed: 16554966]
4. Phillips HS, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell. 2006; 9:157–173. [PubMed: 16530701]
5. Tso CL, et al. Primary glioblastomas express mesenchymal stem-like properties. Mol Cancer Res. 2006; 4:607–619. [PubMed: 16966431]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

6. Takashima Y, et al. Neuroepithelial cells supply an initial transient wave of MSC differentiation. Cell. 2007; 129:1377–1388. [PubMed: 17604725]
7. Wurmser AE, et al. Cell fusion-independent differentiation of neural stem cells to the endothelial lineage. Nature. 2004; 430:350–356. [PubMed: 15254537]
8. Rhodes DR, Chinnaiyan AM. Integrative analysis of the cancer transcriptome. Nat Genet. 2005; 37(Suppl):S31–S37. [PubMed: 15920528]
9. Basso K, et al. Reverse engineering of regulatory networks in human B cells. Nat Genet. 2005; 37:382–390. [PubMed: 15778709]
10. Chen Y, et al. Variations in DNA elucidate molecular networks that cause disease. Nature. 2008; 452:429–435. [PubMed: 18344982]
11. Margolin AA, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006; 7(Suppl 1):S7. [PubMed: 16723010]
12. Margolin AA, et al. Reverse engineering cellular networks. Nat Protoc. 2006; 1:662–671. [PubMed: 17406294]
13. Zhao X, et al. The N-Myc-DLL3 Cascade Is Suppressed by the Ubiquitin Ligase Huwe1 to Inhibit Proliferation and Promote Neurogenesis in the Developing Brain. Dev Cell. 2009; 17:210–221. [PubMed: 19686682]
14. Lim WK, Lyashenko E, Califano A. Master regulators used as breast caqncer metastasis classifier. Pac Symp Biocomput. 2009; 14
15. Mani KM, et al. A Systems biology approach to prediction of oncogenes and perturbation targets in B cell lymphomas. Molecular Systems Biology. 2008; 4:169–178. [PubMed: 18277385]
16. Palomero T, et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. Proc Natl Acad Sci U S A. 2006; 103:18261–18266. [PubMed: 17114293]
17. Taylor RC, Acquaah-Mensah G, Singhal M, Malhotra D, Biswal S. Network inference algorithms elucidate Nrf2 regulation of mouse lung oxidative stress. PLoS Comput Biol. 2008; 4:e1000166. [PubMed: 18769717]
18. Hanauer DA, Rhodes DR, Sinha-Kumar C, Chinnaiyan AM. Bioinformatics approaches in the study of cancer. Curr Mol Med. 2007; 7:133–141. [PubMed: 17311538]
19. Lander AD. A calculus of purpose. PLoS Biol. 2004; 2:e164. [PubMed: 15208717]
20. Freije WA, et al. Gene expression profiling of gliomas strongly predicts survival. Cancer Res. 2004; 64:6503–6510. [PubMed: 15374961]
21. Nigro JM, et al. Integrated array-comparative genomic hybridization and expression array profiles identify clinically relevant molecular subtypes of glioblastoma. Cancer Res. 2005; 65:1678–1686. [PubMed: 15753362]
22. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25:25–29. [PubMed: 10802651]
23. Ramji DP, Foka P. CCAAT/enhancer-binding proteins: structure, function and regulation. Biochem J. 2002; 365:561–575. [PubMed: 12006103]
24. Aoki K, et al. RP58 associates with condensed chromatin and mediates a sequence-specific transcriptional repression. J Biol Chem. 1998; 273:26698–26704. [PubMed: 9756912]
25. Fuks F, Burgers WA, Godin N, Kasai M, Kouzarides T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. Embo J. 2001; 20:2536–2544. [PubMed: 11350943]
26. Lee JP, et al. Stem cells act through multiple mechanisms to benefit mice with neurodegenerative metabolic disease. Nat Med. 2007; 13:439–447. [PubMed: 17351625]
27. Park KI, et al. Acute injury directs the migration, proliferation, and differentiation of solid organ stem cells: evidence from the effect of hypoxia-ischemia in the CNS on clonal 'reporter' neural stem cells. Exp Neurol. 2006; 199:156–178. [PubMed: 16737696]
28. Parker MA, et al. Expression profile of an operationally-defined neural stem cell clone. Exp Neurol. 2005; 194:320–332. [PubMed: 15992799]
29. Bromberg JF, et al. Stat3 as an oncogene. Cell. 1999; 98:295–303. [PubMed: 10458605]

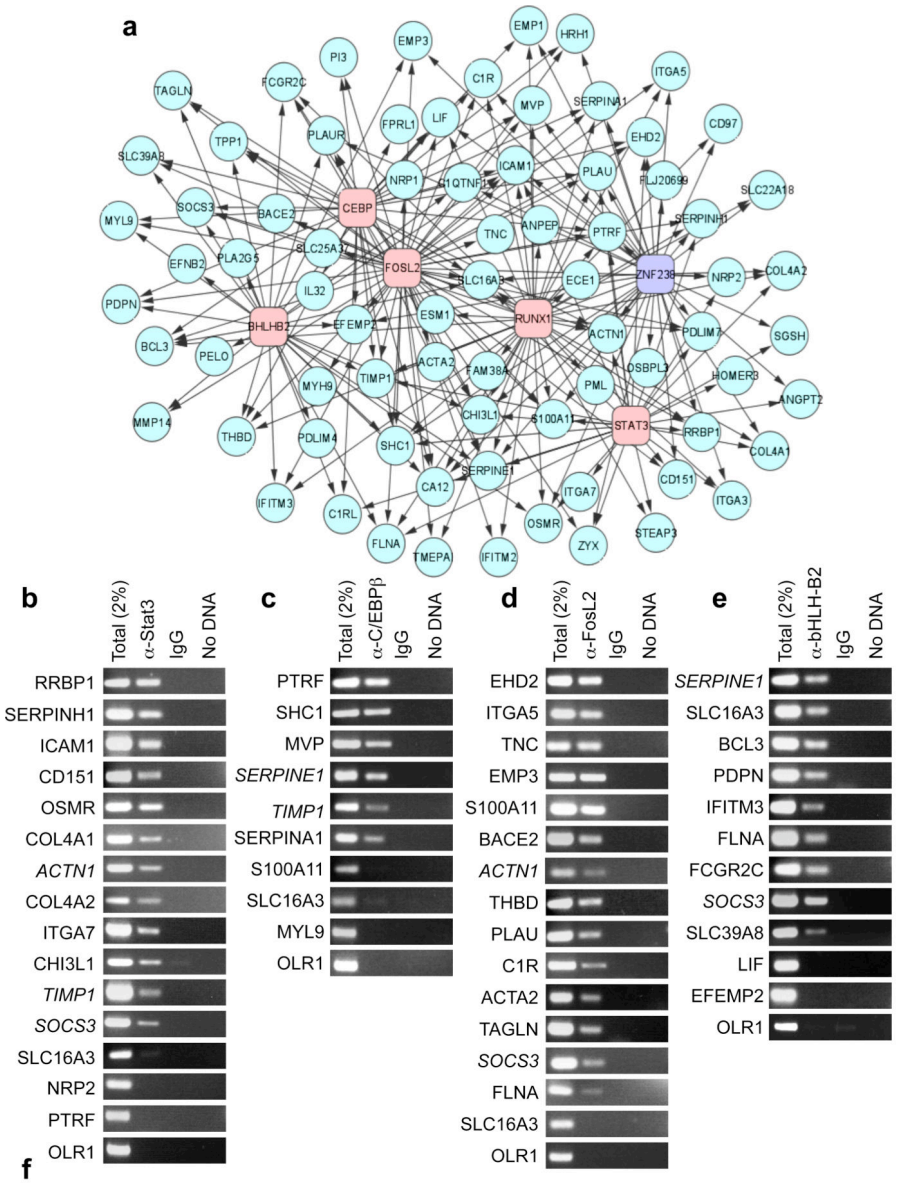
Author Manuscript Author Manuscript Author Manuscript Author Manuscript

30. Lee J, et al. Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. Cancer Cell. 2006; 9:391–403. [PubMed: 16697959]
31. Network A. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–1068. [PubMed: 18772890]
32. Pelloski CE, et al. YKL-40 expression is associated with poorer response to radiation and shorter overall survival in glioblastoma. Clin Cancer Res. 2005; 11:3326–3334. [PubMed: 15867231]
33. Ein-Dor L, Kela I, Getz G, Givol E, Domany L. Outcome signature genes in breast cancer: is there a unique set? Bioinformatics. 2005; 21:171–178. [PubMed: 15308542]
34. Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput. 2000:418–429. [PubMed: 10902190]
35. Barnabe-Heider F, et al. Evidence that embryonic neurons regulate the onset of cortical gliogenesis via cardiotrophin-1. Neuron. 2005; 48:253–265. [PubMed: 16242406]
36. Bonni A, et al. Regulation of gliogenesis in the central nervous system by the JAK-STAT signaling pathway. Science. 1997; 278:477–483. [PubMed: 9334309]
37. Sterneck E, Johnson PF. CCAAT/enhancer binding protein beta is a neuronal transcriptional regulator activated by nerve growth factor receptor signaling. J Neurochem. 1998; 70:2424–2433. [PubMed: 9603207]
38. Nadeau S, Hein P, Fernandes KJ, Peterson AC, Miller FD. A transcriptional role for C/EBP beta in the neuronal response to axonal injury. Mol Cell Neurosci. 2005; 29:525–535. [PubMed: 15936952]
39. Menard C, et al. An essential role for a MEK-C/EBP pathway during growth factor-regulated cortical neurogenesis. Neuron. 2002; 36:597–610. [PubMed: 12441050]
40. Nakashima K, et al. Synergistic signaling in fetal brain by STAT3-Smad1 complex bridged by p300. Science. 1999; 284:479–482. [PubMed: 10205054]
41. Paquin A, Barnabe-Heider F, Kageyama R, Miller FD. CCAAT/enhancer-binding protein phosphorylation biases cortical precursors to generate neurons rather than astrocytes in vivo. J Neurosci. 2005; 25:10747–10758. [PubMed: 16291948]
42. Bachoo RM, et al. Epidermal growth factor receptor and Ink4a/Arf: convergent mechanisms governing terminal differentiation and transformation along the neural stem cell to astrocyte axis. Cancer Cell. 2002; 1:269–277. [PubMed: 12086863]
43. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–15550. [PubMed: 16199517]
44. Tegner J, Yeung MK, Hasty J, Collins JJ. Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. Proc Natl Acad Sci U S A. 2003; 100:5944–5949. [PubMed: 12730377]
45. Bussemaker HJ, Li H, Siggia ED. Regulatory element detection using correlation with expression. Nat Genet. 2001; 27:167–171. [PubMed: 11175784]
46. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B (Methodological). 1995; 57:289–300.
47. Frank SR, Schroeder M, Fernandez P, Taubert S, Amati B. Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. Genes Dev. 2001; 15:2069–2082. [PubMed: 11511539]
48. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25:402–408. [PubMed: 11846609]
49. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008; 24:1547–1548. [PubMed: 18467348]
50. Rothschild G, Zhao X, Iavarone A, Lasorella A. E Proteins and Id2 Converge on p57Kip2 To Regulate Cell Cycle in Neural Cells. Mol Cell Biol. 2006; 26:4351–4361. [PubMed: 16705184]
51. Zhao X, et al. The HECT-domain ubiquitin ligase Huwe1 controls neural differentiation and proliferation by destabilizing the N-Myc oncoprotein. Nat Cell Biol. 2008; 10:643–653. [PubMed: 18488021]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

52. Simmons ML, et al. Analysis of complex relationships between age, p53, epidermal growth factor receptor, and survival in glioblastoma patients. Cancer Res. 2001; 61:1122–1128. [PubMed: 11221842]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



f

TF	Stat3	C/EBPβ	FosL2	bHLH-B2	Total
# MES target	29	20	48	30	
# tested target	14	9	15	12	50
# bound	11 (78.6%)	6 (66.7%)	14 (93.3%)	9 (75%)	40 (80%)

Figure 1. The mesenchymal signature of HGGs is controlled by six TFs

a, TFs involved in activation of MGES targets are shown in pink, those involved in repression are in purple. MGES targets controlled by these TFs are in cyan. Overall, the six TFs control 74% of the genes in the mesenchymal signature of high-grade glioma. A region between 2 kb upstream and downstream the transcription start site of the target genes identified by ARACNe was analyzed for the presence of putative binding sites. Genomic regions of genes containing putative binding sites for specific TFs were immunoprecipitated in SNB75 cells by antibodies specific for **b**, Stat3; **c**, C/EBPβ; **d**, FosL2; **e**, bHLH-B2.

SOCS3 was included as positive control of Stat3 binding. Total chromatin before immunoprecipitation was used as positive control for PCR. The OLR1 gene was used as negative control. **f**, Summary of binding results of the tested TFs to mesenchymal targets.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

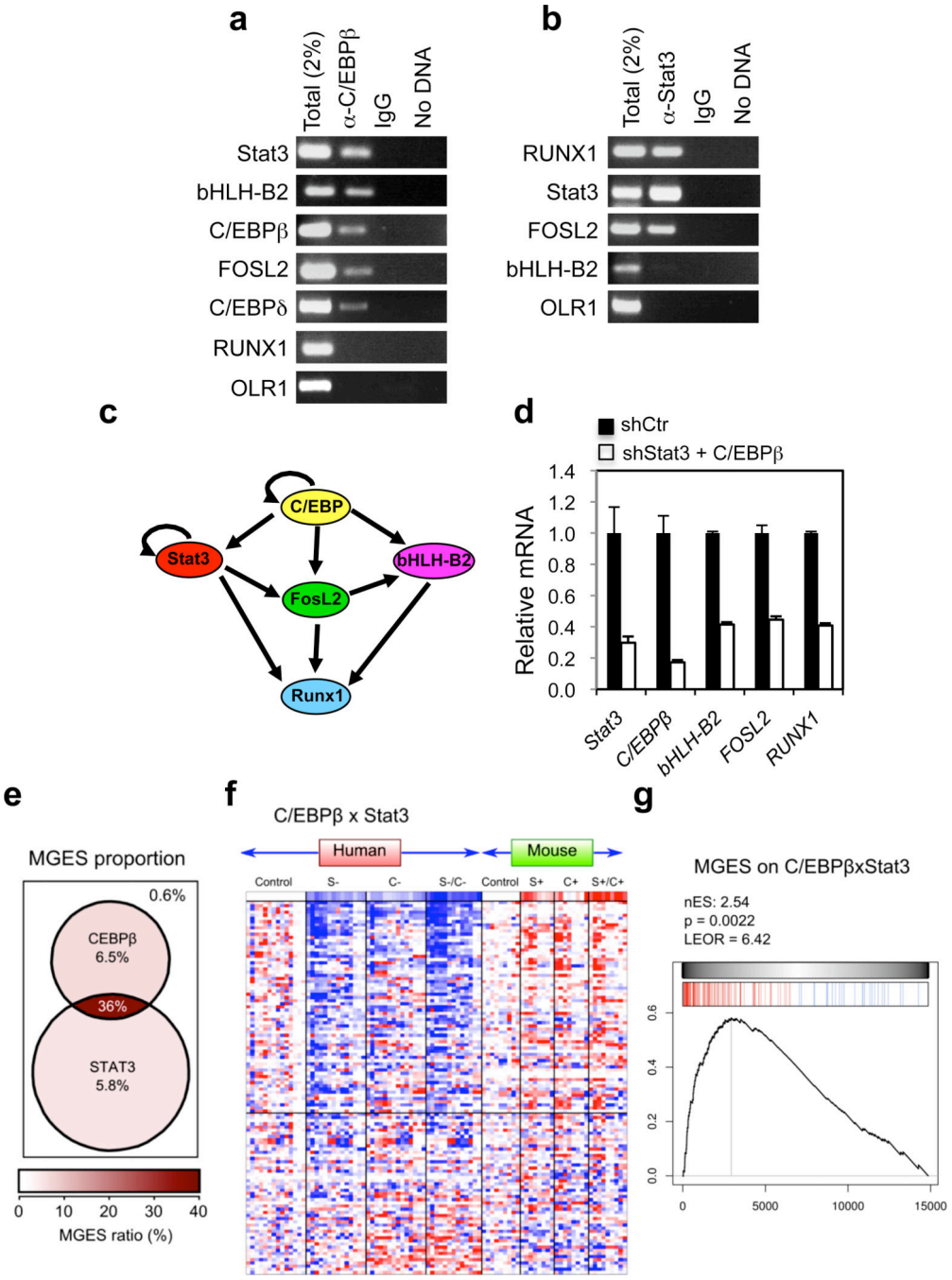


Figure 2. A hierarchical transcriptional module regulates the MGES

ChIP for **a**, C/EBPβ; **b**, Stat3. **c**, Transcriptional network emerging from promoter occupancy analysis. **d**, qRT-PCR of mesenchymal TFs in glioma cells infected with Stat3/C/EBPβ shRNA or controls lentiviruses. **e**, Venn-diagram depicts the proportion of mesenchymal genes identified by ARACNe as targets of only C/EBPβ, Stat3 or both TFs. **f**, Heatmap of MGES gene expression analysis of mouse and human cells carrying perturbations of C/EBPβ plus Stat3. Samples (columns) were grouped according to species and treatment. Control, control shRNA or empty vector; S-, Stat3 knock-down; S+, Stat3

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

overexpression; C-, C/EBPβ knock-down; C+, C/EBPβ overexpression; S-/C-, Stat3/ C/EBPβ knockdown; S+/C+, Stat3/ C/EBPβ overexpression. **g**, GSEA of the MGES on the gene expression profile rank-sorted according to the correlation with the C/EBPβxStat3 metagene. The bar-code plot indicates the position of MGES genes, red and blue colors represent positive and negative correlation, respectively. The gray scale bar indicates the spearman rho coefficient used as weighting score for GSEA. nES, normalized enrichment score; p, sample-permutation-based p-value.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

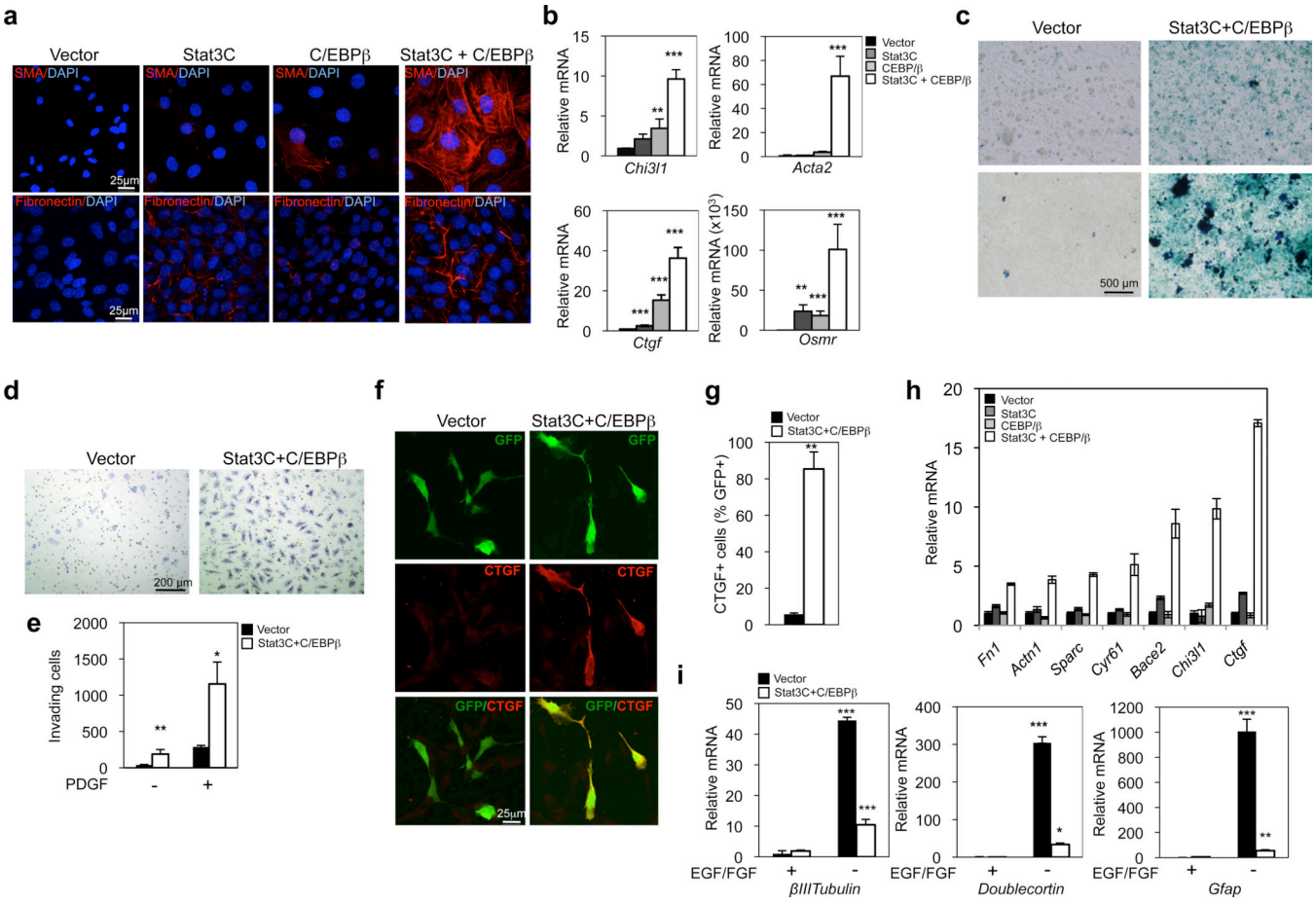


Figure 3. Ectopic expression of C/EBPβ and Stat3C in NSCs induces mesenchymal transformation and inhibits neural differentiation

a, Immunofluorescence analysis for SMA and fibronectin in C17.2 expressing the indicated TFs. **b**, qRT-PCR of mesenchymal targets in C17.2 expressing the indicated TFs. n = 3; Bars indicate Mean±SD. **c**, Alcian blue staining of C17.2 expressing Stat3C and C/EBPβ or the empty vector cultured in growth medium (upper panels), or chondrocyte differentiation medium (lower panels). **d**, Microphotographs of invading C17.2 expressing Stat3C and C/EBPβ or empty vector. **e**, Quantification of invading cell in the absence or in the presence of PDGF. n = 3; Bars indicate Mean±SEM. **f**, Immunofluorescence analysis for CTGF in NSCs expressing Stat3C and C/EBPβ or the empty vector. GFP identifies infected cells. **g**, Quantification of $GFP^+/CTGF^+$ cells. Bars indicate Mean±SD of three independent experiments. **h**, qRT-PCR of representative mesenchymal genes in primary NSCs expressing the indicated TFs. n = 3; Bars indicate Mean±SD. **i**, qRT-PCR of βIII-tubulin, doublecortin and GFAP in NSCs expressing Stat3C plus C/EBPβ or the empty vector. n= 3; Bars indicate Mean±SD. qRT-PCR data are 18S ribosomal RNA normalized fold changes. $*p \leq 0.05$; $**p \leq 0.01$; $***p \leq 0.001$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

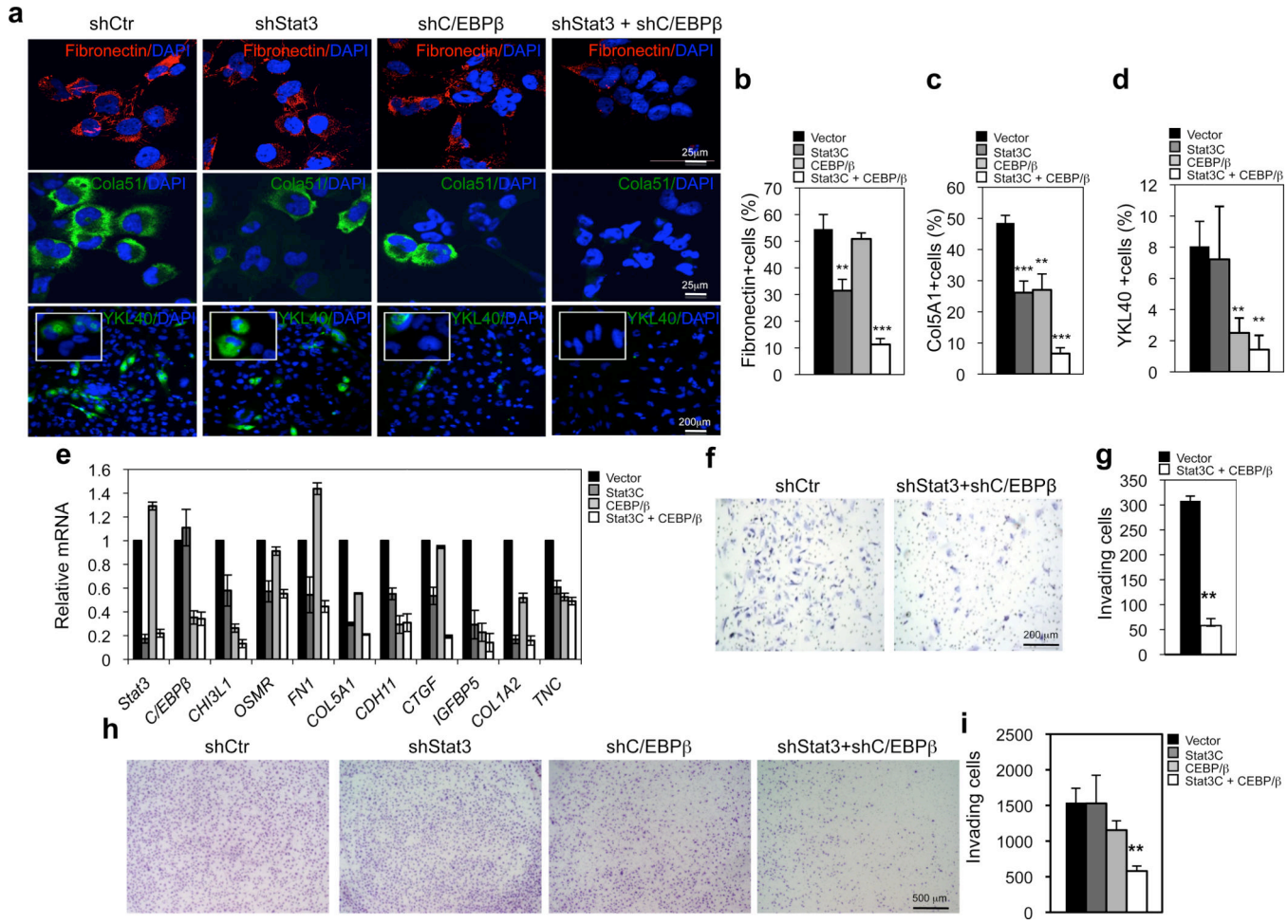


Figure 4. C/EBPβ and Stat3 maintain the mesenchymal phenotype of human glioma cells

a, Immunofluorescence for fibronectin, Col5A1 and YKL40 in BTSC-3408 infected with lentiviruses expressing Stat3, C/EBPβ, or Stat3 plus C/EBPβ shRNA. **b**, Quantification of fibronectin; **c**, Col5A1; and **d**, YKL40 positive cells. n = 3 independent experiments; Bars indicate Mean±SD. **e**, qRT-PCR of mesenchymal genes in BTSC-20 infected with lentiviruses expressing Stat3, C/EBPβ, or Stat3 plus C/EBPβ shRNA. Gene expression was normalized to the expression of 18S ribosomal RNA. n = 3; Bars indicate Mean±SD. **f**, Microphotograps of invading SNB19 cells infected with lentiviral vectors expressing control or shStat3 plus shC/EBPβ. **g**, Quantification of SNB19 invading cells. Bars indicate Mean ±SD; n = 6 (two independent experiments, each performed in triplicate). **h**, Invading BTSC-3408 cells infected with shCtr, shStat3, shC/EBPβ or shStat3 plus shC/EBPβ lentiviruses. **i**, Quantification of invading BTSC-3408 cells. Bars indicate Mean±SD; n = 6 (two independent experiments, each performed in triplicate). $*p \leq 0.05$, $**p \leq 0.01$, $***p \leq 0.001$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

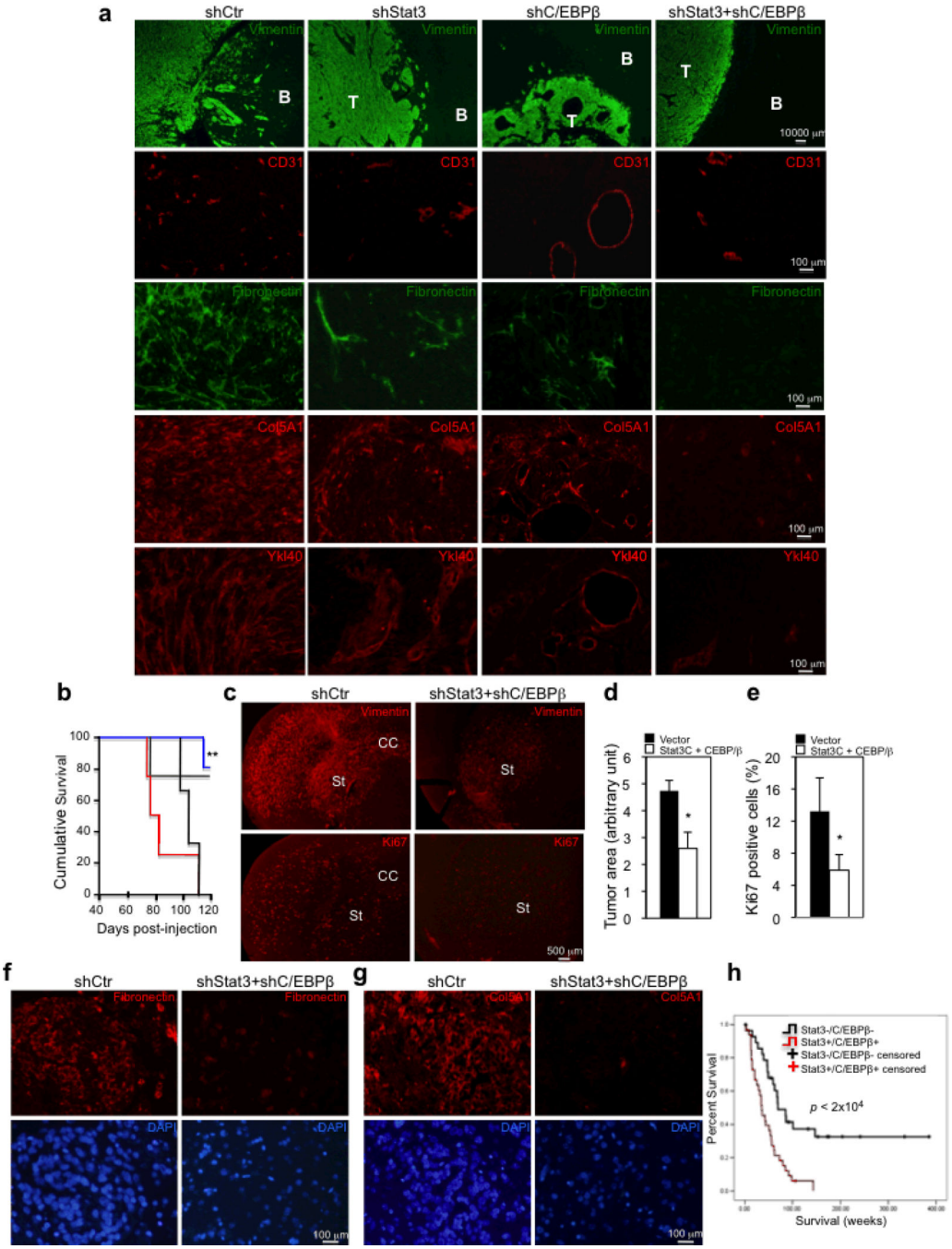


Figure 5. C/EBPβ and Stat3 are essential for glioma tumor aggressiveness in mice and humans

a, Immunofluorescence staining for human vimentin, CD31, fibronectin, Col5A1, and YKL40 in tumors derived from SNB19 cells infected with lentiviruses expressing shRNA targeting Stat3, C/EBPβ, or Stat3 plus C/EBPβ. T, tumor; B, normal brain. **b**, Kaplan-Meier survival curve of NOD SCID mice transplanted intracranially with SNB19 glioma cells transduced with shCtr (red), shStat3 (black), shC/EBPβ (green) or shStat3 plus shC/EBPβ (blue) lentiviruses. **c**, Immunostaining for human vimentin and Ki67 on representative brain sections from mice injected with BTSC-3408 after silencing of C/EBPβ and Stat3. St,

striatum; CC, corpus callosum. **d**, Quantification of human vimentin positive area. **e**, Quantification of Ki67 positive cells. n=5 for each group; Bars indicate Mean±SD. **f**, Immunostaining for fibronectin and **g**, Col5A1 on representative brain sections from mice injected with BTSC-3408 transduced as indicated. **h**, Kaplan-Meier analysis comparing survival of patients carrying tumors double positive for C/EBPβ and Stat3 (red) and single or double negative tumors (black). *$p \leq 0.05$; **$p \leq 0.01$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript