The transcriptional network for mesenchymal transformation of brain tumours

Maria Stella Carro1*†, Wei Keat Lim2,3*†, Mariano Javier Alvarez3,4*, Robert J. Bollo8, Xudong Zhao1, Evan Y. Snyder9, Erik P. Sulman10, Sandrine L. Anne1†, Fiona Doetsch5, Howard Colman11, Anna Lasorella1,5,6, Ken Aldape12, Andrea Califano1,2,3,4 & Antonio Iavarone1,5,7

The inference of transcriptional networks that regulate transitions into physiological or pathological cellular states remains a central challenge in systems biology. A mesenchymal phenotype is the hallmark of tumour 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 an 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 tumour aggressiveness. In human glioma, expression of C/EBPβ and STAT3 correlates with mesenchymal differentiation and predicts poor clinical outcome. These results show that the 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 tumours in humans and are essentially incurable1. The defining hallmarks of the aggressiveness of glioblastoma multiforme (GBM) are local invasion and neo-angiogenesis2–4. 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 an uncontrolled ability to invade and stimulate angiogenesis5–7. Gene expression studies have established that overexpression of a ‘mesenchymal’ gene expression signature (MGES) and loss of a proneural signature (PNGES) co-segregate with the poor prognosis group of glioma patients4. Yet, differentiation along the mesenchymal lineage is virtually undetectable in normal neural tissue during development. Thus, it is unclear whether drift towards the mesenchymal lineage is an aberrant event that occurs during brain tumour progression or whether glioma cells recapitulate the rare mesenchymal plasticity of neural stem cells (NSCs)3–7. 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 transcription factors (TFs) that are master regulators (MRs) of specific cancer signatures, on the basis of cellular network models, have yet to produce experimentally validated discoveries, probably because these networks are still poorly mapped, especially within specific mammalian cellular contexts9. Notwithstanding, recent developments in genome-wide reverse engineering were successful in identifying causal, rather than associative interactions2–4, and showed promise in the identification of dysregulated genes within developmental and tumour-related pathways13–17. Thus, we reasoned that context-specific regulatory networks, inferred by unbiased reverse engineering algorithms, may provide sufficient accuracy to allow estimation of (1) the activity of TFs from that of their transcriptional targets or regulons, and (2) the identity of TFs that are MRs of specific eukaryotic signatures18,19 from the overlap between their regulons and the signatures. We applied the above mechanisms to unravel the MRs causally linked to activation of the MGES in malignant glioma (Supplementary Fig. 1).

A transcriptional module is linked to the MGES of HGGs

We first addressed 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 tumours (Supplementary Fig. 2).

We thus used the ARACNe algorithm20 to assemble a genome-wide repertoire of HGG-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). ARACNe is an information theoretical approach for the inference of TF-target interactions from large sets of gene expression profiles18,21, further refined to
determine directed (that is, causal) interactions\(^{12,22}\) (see Methods). ARACNe predicted 92,660 transcriptional interactions, 1,217 of which were between TFs and 102 out of 149 MGES genes\(^4\), represented across all the gene expression profile data.

Next, we applied a new master regulator analysis (MRA) algorithm to the HGG-interactome. The algorithm computes the statistical significance of the overlap between the regulon of each TF (that is, its ARACNe-inferred targets) and the MGES genes (\(P\) values computed by Fisher’s 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 on the basis of the total number of MGES targets they regulated. The top six TFs (STAT3, C/EBP, bHLH-B2 (also known as bHLHE40), RUNX1, FOSL2 and ZNF238) collectively controlled >74% of the MGES genes (Fig. 1a). C/EBP\(^\beta\) and C/EBP\(^\delta\) 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’s correlation analysis showed that five of these TFs are probably activators (STAT3, C/EBP, bHLH-B2, RUNX1 and FOSL2) and one is probably a repressor (ZNF238).

![Figure 1](https://example.com/figure1.png)

**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 of the transcription start site of the target genes identified by ARACNe was analysed for the presence of putative binding sites.

\(b\)–\(e\), Genomic regions of genes containing putative binding sites for specific TFs were immunoprecipitated in SNB75 cells by antibodies specific for STAT3 (\(b\)), C/EBP\(^\beta\) (\(c\)), FOSL2 (\(d\)), and bHLH-B2 (\(e\)). 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.
Overlap between the regulators of these TFs was highly significant (Supplementary Table 4). MRA 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 to determine 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 the promoters of MGES genes. Six TFs were in both lists, for a total of 99 TFs (Supplementary Table 3b). TFs were then ranked on the basis of 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 stepwise linear regression analysis (Supplementary Table 6). Indeed, the three with the highest linear-regression coefficient values were C/EBPβ (r = 0.40), bHLH-B2 (r = 0.41) and STAT3 (r = 0.40), further establishing them as probable MGES-MR candidates. The next strongest TF, ZNF238, had a negative coefficient (r = −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 short hairpin RNA (shRNA) silencing of CEBPB, STAT3, bHLH-B2, FOSL2 and RUNX1 in glioma cells, followed by gene expression profiling and Gene Set Enrichment Analysis (GSEA), showed that after 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 eight out of ten possible intra-module interactions implemented (modularity P = 1.0 × 10−8 by FET, Fig. 2c). Specifically, CEBPB and STAT3 occupy their own promoter (Fig. 2a, b); CEBPB occupies the STAT3, FOSL2, bHLH-B2, CEBPB and CEBPD 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 only occupies the RUNX1 promoter (Supplementary Fig. 3b). CEBPB 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 CEBPB and STAT3 in glioma cells produced >2-fold reduction of the messenger RNAs coding for the second layer TFs in the feed-forward loops (bHLH-B2, FOSL2 and RUNX1; Fig. 2d), supporting their role as MRs. CEBPB and STAT3 also bound the promoters of their MGES targets in primary human GBM (Supplementary Fig. 3c, d).

To validate functionally the role of CEBPB and STAT3 as MRs of the MGES, we conducted gain- and loss-of-function experiments. We transduced v-myc immortalized mouse NSCs (mNSCs) known as C17.2 (refs 26–28), as well as primary mNSCs derived from the mouse telencephalon at embryonic day (E)13.5, with retroviruses expressing C/EBPB and a constitutively active form of STAT3 (STAT3C)29, shRNA-mediated silencing targeted CEBPB and STAT3 in the human glioma cell line SNB19, and in serum-free cultures of tumour cells derived from primary GBM that propel the formation of GBM-like tumours after intracranial transplantation in immunodeficient mice30 (GBM-derived brain tumour initiating cells, GBM-BTICs; see later).

Figure 2 | A hierarchical transcriptional module regulates the MGES.

We generated a global data set 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 PCR with reverse transcription (qRT–PCR) analysis showed that, after CEBPB or STAT3 shRNA silencing in GBM-BTICs and SNB19 cells, the corresponding mRNA levels were significantly reduced compared to non-target control transduced cells (CEBPβ fold ratio = 0.26, P ≤ 0.00108, STAT3 fold ratio = 0.205, P = 0.00109 by U-test). Reciprocal changes followed ectopic expression of the two TFs in C17.2 cells and NSCs (Supplementary Table 9). qRT–PCR values and microarray-based measurements were highly correlated for STAT3 but not for CEBPB mRNA (Supplementary Fig. 4). Thus, we used the qRT–PCR values for CEBPB and STAT3 as more accurate read-outs for their mRNA expression. GSEA 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 (\(P = 2.69 \times 10^{-2}\) and \(P = 2.0 \times 10^{-4}\), respectively, by GSEA). Common targets of C/EBP and STAT3 were eightfold more enriched in MGES genes than targets controlled individually by each TF (Fig. 2e, \(P = 2.25 \times 10^{-3}\)). To test whether the two TFs may be involved in synergistic MGES control, we computed a metagene (\(\text{CEBP} \times \text{STAT3}\)) 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 confirmed that genes ranked by Spearman’s correlation to the CEBP×STAT3 metagene were significantly enriched in MGES genes (Fig. 2g), suggesting that at least a subset of the MGES is synergistically regulated by the CEBP×STAT3 pair.

We sought to establish (1) whether MRs inferred by our procedure would also be inferred when using an independent glioma sample data set, and (2) whether MRs identified on the basis of clinical outcome would overlap significantly with those inferred from MGES analysis. The Atlas-TCGA data set 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 versus best-prognosis samples (at \(P \leq 0.05\) by Student’s t-test, Supplementary Table 11). GSEA confirmed that MGES genes were markedly enriched in the TWPS signature (\(P \leq 1.0 \times 10^{-5}\), Supplementary Fig. 6) indicating that the poor-prognosis group in the Atlas-TCGA data set displays marked mesenchymal features. Despite partial overlap between MGES and TWPS genes (22.8%), five of the six MRs identified by MRA from the original data set were also found among the ten most significant TFs identified by MRA of the Atlas-TCGA data set using the TWPS signature. C/EBPβ was the most significant TF, and STAT3 was in seventh position. C/EBPβ and C/EBPβ had the first and second best linear-regression coefficient by stepwise linear regression analysis, respectively (Supplementary Table 12). These results indicate significant robustness of the approach both to data set and signature selection.

**Mesenchymal reprogramming of NSCs by C/EBPβ and STAT3**

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. The introduction of C/EBPβ and STAT3C into 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 expression of the mesenchymal marker proteins smooth muscle alpha actin (SMA, encoded by Acta2) and fibronectin (Fig. 3a and Supplementary Fig. 7c, d), and with induced expression of the mesenchymal genes Chi3l1 (also known as Ykl40), Acta2, Ctgf and Osmr (Fig. 3b). The individual expression of STAT3C or C/EBPβ was generally insufficient to induce expression of either mesenchymal marker genes or mesenchymal genes (Fig. 3a, b). Removal of mitogens to STAT3C plus C/EBPβ-expressing C17.2 cells resulted in a further increase in the expression of mesenchymal genes, and complete acquisition of mesenchymal properties such as positive alcin blue staining, a specific assay for chondrocyte differentiation (Fig. 3c and Supplementary Fig. 7e, f). The expression of STAT3C and C/EBPβ promoted migration in a wound assay, and triggered invasion through presence of PDGF, \(n = 3\); error bars denote mean ± s.e.m.

**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 cells expressing the indicated TFs. DAPI, 4',6-diamidino-2-phenylindole. **b**, qRT–PCR of the mesenchymal targets in C17.2 cells expressing the indicated TFs, \(n = 3\); error bars are mean ± s.d. **c**, Alcian blue staining of C17.2 cells expressing STAT3C plus C/EBPβ or the empty vector. Green fluorescent protein (GFP) identifies infected cells. **d**, Quantification of GFP + Ctgf + cells. Error bars denote mean ± s.d. of three independent experiments. **e**, qRT–PCR of representative mesenchymal genes in primary NSCs expressing the indicated TFs, \(n = 3\); error bars indicate mean ± s.d. i, qRT–PCR of Tub3, Dcx and Gfap in NSCs expressing STAT3C plus C/EBPβ or the empty vector, \(n = 3\); error bars indicate mean ± s.d. qRT–PCR data are normalized to fold changes in 18S ribosomal RNA. *\(P \leq 0.05\), **\(P \leq 0.01\), ***\(P \leq 0.001\).
the extracellular matrix in a Matrigel invasion assay in the absence or presence of PDGF (Fig. 3d, e and Supplementary Fig. 7g). The combined but not individual expression of STAT3C and C/EBPβ efficiently induced mesenchymal marker proteins and mesenchymal gene expression in primary NSCs (Fig. 3f–h). Conversely, STAT3C and C/EBPβ abolished differentiation along the neuronal and glial lineages (Fig. 3i and Supplementary Fig. 7h). The C/EBPβ- and STAT3C-induced mesenchymal transformation of primary NSCs was associated with withdrawal from the cell cycle (data not shown). Thus, the combined introduction of C/EBPβ and STAT3C in NSCs prevents neural differentiation and triggers reprogramming towards an aberrant mesenchymal lineage.

Mesenchymal aggressiveness of glioma by C/EBPβ and STAT3

Transduction of GBM-BTIC 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 tumour cells of the mesenchymal marker proteins fibronectin, collagen-5α1 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 and 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 cells and GBM-BTICs reduced their ability to invade through Matrigel by >70% (Fig. 4f–i). Next, we determined the effect of CEBPB and STAT3 knockdown on brain tumorigenesis after intracranial injection of SNB19 cells in immunocompromised mice. We observed efficient tumour formation in all mice injected with control or STAT3 shRNA cells. However, only one of four mice from the CEBPB shRNA and one of five mice from the combined CEBPB and STAT3 shRNA groups developed tumours 120 days after the injection (Fig. 5b). The histological analysis demonstrated high-grade tumours, 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 control shRNA tumour group. Conversely, the single tumour in the CEBPB plus STAT3 shRNA group grew well circumscribed and was less angiogenic. Tumours in the STAT3 shRNA group and the single tumour in the CEBPB shRNA group had an intermediate growth pattern and limited angiogenesis. Staining for fibronectin, collagen-5α1 and YKL40 was readily detected in the tumours from the control group but absent or barely detectable in the single tumours from the CEBPB shRNA and CEBPB plus STAT3 shRNA groups (Fig. 5a). Tumours derived from Stat3 shRNA cells showed an intermediate phenotype, with reduced expression of mesenchymal markers compared with tumours in the control shRNA group, but higher than that observed in the tumours in the CEBPB shRNA and CEBPB plus STAT3 shRNA groups (control shRNA > STAT3 shRNA > CEBPB shRNA > CEBPB+STAT3 shRNA). Intracranial transplantation of GBM-BTICs transduced with control shRNA lentivirus produced extremely invasive tumour cell masses extending through the corpus callosum to the contralateral brain. Combined knockdown of CEBPB and STAT3 led to a significant decrease in the tumour area and tumour 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-5α1 (Fig. 5f, g).

Finally, we conducted immunohistochemical analysis for CEBPβ and active, phospho-STAT3 in human tumour specimens, and

**Figure 4** CEBPβ and STAT3 maintain the mesenchymal phenotype of human glioma cells. **a**, Immunofluorescence for fibronectin, collagen-5α1 (COL5A1) and YKL40 in BTSC-3408 infected with lentiviruses expressing STAT3, CEBPB or STAT3 plus CEBPB shRNA. **b**–**d**, Quantification of cells positive for fibronectin (b), COL5A1 (c) and YKL40 (d). n = 3 independent experiments; error bars indicate mean ± s.d. e, qRT–PCR of mesenchymal genes in BTSC-20 infected with lentiviruses expressing STAT3, CEBPB or STAT3 plus CEBPB shRNA. Gene expression was normalized to 18S ribosomal RNA expression. n = 3; error bars indicate mean ± s.d.

f, Microphotographs of invading SNB19 cells infected with lentiviral vectors expressing control or STAT3 plus CEBPB shRNA. g, Quantification of SNB19-invading cells. Error bars indicate mean ± s.d.; n = 6 (two independent experiments, each performed in triplicate). h, Invading BTSC-3408 cells infected with control, STAT3, CEBPB or STAT3 plus CEBPB shRNA lentiviruses. i, Quantification of invading BTSC-3408 cells. Error bars indicate mean ± s.d.; n = 6 (two independent experiments, each performed in triplicate). *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001.
Figure 5 | C/EBPβ and STAT3 are essential for glioma tumour aggressiveness in mice and humans. a, Immunofluorescence staining for human vimentin, CD31, fibronectin, COL5A1 and YKL40 in tumours derived from SNB19 cells infected with lentiviruses expressing shRNA targeting STAT3, CEBPB, or STAT3 plus CEBPB. b, Kaplan–Meier survival curve of NOD/SCID mice transplanted intracranially with SNB19 glioma cells transduced with control shRNA (red), STAT3-shRNA (black), CEBPB-shRNA (green) or STAT3 plus CEBPB-shRNA (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. CC, corpus callosum; St, striatum. d, Quantification of human vimentin-positive area. e, Quantification of Ki67-positive cells. Error bars are s.d. f, Immunostaining for fibronectin (f) and COL5A1 (g) on representative brain sections from mice injected with BTSC-3408 transduced as indicated. h, Kaplan–Meier analysis comparing survival of patients carrying tumours double positive for C/EBPβ and STAT3 (red) and single- or double-negative tumours (black). *P < 0.05, **P < 0.01.

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 model of gene-expression-profile-based cancer research, yielding long lists of differentially expressed genes (that is, 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. Our approach produced virtually identical regulatory MR modules when applied to two completely distinct data sets and

compared their expression with that of YKL40 (a well-established mesenchymal protein expressed in primary human GBM) and with patient outcome in a collection of 62 GBM (Supplementary Fig. 9). Expression of either C/EBPβ or STAT3 was significantly associated with YKL40 expression (C/EBPβ, P = 4.9 × 10^{-5}; STAT3, P = 2.2 × 10^{-5}), with higher association in double-positive tumours (C/EBPβ+STAT3+, P = 2.7 × 10^{-6}) versus double-negative ones (C/EBPβ−STAT3−, Supplementary Table 13). Double-positive tumours were associated with worse clinical outcome than either single- or double-negative tumours (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).
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 analysis8 and relevance network analysis44 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.

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 system35–38. However, although identified, direct experimental validation is feasible.

proposed here. Yet, once a relatively small number of TFs is identified, direct experimental validation is feasible.

models had on average 1–5 TFs. The statistical significance of the enrichment of a ranked list of genes in a section between the TF-regulon and the gene expression signature was computed by identifying the smallest number of TFs that were informative for the expression of MRs sufficient in NSCs and necessary in human glioma cells for mesenchymal transformation. Interestingly, C/EBPβ and STAT3 are expressed in the developing nervous system35–38. However, although STAT3 induces astrocyte differentiation and inhibits neuronal differentiation of neural stem/progenitor cells, C/EBPβ promotes neurogenesis and opposes gliogenesis44–47. A question remains as to how the combined activity of C/EBPβ and STAT3 can reprogram NSCs towards an aberrant lineage (mesenchymal) and oppose the genesis of the normal neural lineages (neuronal and glial). We propose that mesenchymal transformation results from the concurrent activation of two conflicting transcriptional regulators normally operating to funnel opposing signals (neurogenesis versus gliogenesis). This condition 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 (for example, EGFR amplification, PTEN loss, Akt activation)48. Because expression of C/EBPβ and STAT3 in human glioma is essential to maintain the tumour initiating capacity and the ability to invade the normal brain, the two TFs provide important clues for diagnostic and pharmaceutical 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 tumour aggressiveness.

Thus, systems biology methods can be effectively used to infer MRs that choreograph malignant transformation. This model 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 FGF2 and EGF as described42. Differentiation was induced by culturing NSCs in NSC medium without EGF and FGF2. GBM-derived BTICs were grown in Neurobasal media supplemented with FGF2 and EGF.

Generation of transcriptional network, microarrays and qRT–PCR. GBM transcriptional network was generated by ARACNe12. 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 18S ribosomal RNA. 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 FET. Significant genes were ranked on the basis of the number of overlapping genes.

GSEA. The statistical significance of the enrichment of a ranked list of genes in a smaller set of genes was determined as described49.

Stepwise linear regression. 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 data set. 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 1–5 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 mice 48 h after infection with lentiviruses carrying shRNAs using a stereotaxic frame. Animals were monitored and euthanized when they presented with signs of tumour.

Mouse research was approved by the Committee for Animal Care, and conducted in compliance with the Animal Welfare Act Regulations.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 September; accepted 30 November 2009.

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 Res. 65, 1679–1686 (2005).
Acknowledgements  This work was supported by National Institute of Health grants R01CA109755 (A.C.), R01CA101644 (A.L.), R01CA085628 and R01NS061776 (A.I.), NCI Grand Opportunities TDDN Network T1RC2CA148308-01 (A.C.), In Silico Research Centre of Excellence NCI-caBIG 29X5192 (A.C.), National Centers for Biomedical Computing NIH Roadmap Initiative US4CA121852 (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 N. Ramirez-Martinez for technical assistance with mouse husbandry and in vivo procedures.

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 tumour xenograft immunohistochemistry and tumour 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. E.P.S., H.C. and K.A. provided reagents, performed the arrayCGH/expression analysis and primary human tumour immunohistochemistry. S.L.A. performed cell culture immunofluorescence microscopy and analysis. A.L. assisted in primary NSC experiments, performed intracranial injections 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 Information  Gene expression data have been deposited in Gene Expression Omnibus (GEO) with the following accession numbers: GSE19113 for mouse and GSE19114 for human data. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.C. (califano@c2b2.columbia.edu) or A.I. (ai2102@columbia.edu).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.
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 aCGH profiling were performed. Tumours were grouped on the basis of 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 aCGH array as determined by comparison of the sequence annotation of both array platforms, as previously described.

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. 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 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[z; y]}. TF classification. To identify human TFs, we selected the human genes annotated as ‘transcription factor activity’ in Gene Ontology and the list of TFs was computed as follows: the log2 expression of the TFs controlling a larger fraction of the signature will be more likely to determine enriched TFs but not to rank them. TFs are thus ranked based on the total FET. Because FET depends on regulon size, it can be used to assess signature-TF classification.

5

(50 ng ml−1) and B27 supplements (Invitrogen), and human recombinant FGF2 and EGF (20 ng ml−1) were grown as neurospheres in Neurobasal media (Invitrogen) containing N2 (FWER), especially when the number of tests is large.

Enrichment analysis. The log-transformation allows convenient linear representation of regulated, with the average linear-regression coefficient providing further insight. The log-transformation allows convenient linear representation of regulated, with the average linear-regression coefficient providing further insight. The log-transformation allows convenient linear representation of regulated, with the average linear-regression coefficient providing further insight. The log-transformation allows convenient linear representation of regulated, with the average linear-regression coefficient providing further insight.

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 FGF2 and EGF (20 ng ml−1 each) as described.

Differentiation of NSCs was induced by culturing neurospheres on laminin-coated dishes in NSC medium in the absence of growth factors. mNSCs expressing STAT3C and C/EBPβ were differentiated by retroviral infections using supernatant from Phoenix ecotropic packaging cells transduced with pBabe-STAT3C-Flag and/or pLZRS-T7-His-C/EBPβ-2-IRES-GFP.

Promoter analysis and ChIP. Promoter analysis was performed using the MatInspector software (http://www.genomatix.de). A sequence 2 kb upstream and 2 kb downstream from the transcription start site was analysed for the presence of putative binding sites for each TF. Primers used to amplify sequences surrounding the predicted binding sites were designed using the Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/prim3er_www.cgi) and are listed in Supplementary Table 14.

ChIP was performed as described. SNB75 cell lysates were preclared 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, BETHYLAB Laboratories), or normal rabbit immunoglobulins (Santa Cruz). For primary GBM samples, 30 mg of frozen tissue was transferred in a tube with 1 ml of culture medium, fixed with 1% formamide for 15 min and stopped with 0.125 M glycine for 5 min. Samples were centrifuged at 1,500 g 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.

tin (mouse monoclonal, BD Bioscences, 1:100), COL5A1 (rabbit polyclonal, Santa Cruz), FOSL2 (Fra2, sc-604, Santa Cruz), bHLH-B2 (A300-649A, BETHYLAB Laboratories), or normal rabbit immunoglobulins (Santa Cruz). For primary GBM samples, 30 mg of frozen tissue was transferred in a tube with 1 ml of culture medium, fixed with 1% formamide for 15 min and stopped with 0.125 M glycine for 5 min. Samples were centrifuged at 1,500 g 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.

A sequence 2 kb upstream and 2 kb downstream from the transcription start site was analysed for the presence of putative binding sites for each TF. Primers used to amplify sequences surrounding the predicted binding sites were designed using the Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/prim3er_www.cgi) and are listed in Supplementary Table 14.

qRT–PCR and microarray analysis. RNA was prepared with RiboPure kit (Ambion). cDNA was used for further real-time qRT–PCR 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 analysed by the ΔΔCt method using 18S as a housekeeping gene.

RNA amplification for array analysis was performed with Illumina TotalPrep RNA Amplification Kit (Ambion). One-and-a-half micrograms of amplified RNA was hybridized on Illumina HumanHT-12v3c (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 under the R-system.

Immunofluorescence and immunohistochemistry. Immunofluorescence staining was performed as previously described. Primary antibodies and dilutions were: SMA (mouse monoclonal, Sigma, 1:200), fibronectin (mouse monoclonal, BD Biosciences, 1:200), Tau (rabbit polyclonal, Dako, 1:400), βIII-tubulin (mouse monoclonal, Promega, 1:10,000), 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 Axioskop2 FS M0T microscope were used to score positive cells. At least 500 cells were scored for each sample. Quantification of the fibronectin intensity staining in mNSCs was performed using NIH Image J software (http://rsb.info.nih.gov/ij/). 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 tumours, mice were perfused transcardially with 4% paraformaldehyde (PFA), brains were dissected and post-fixed for 48 h in 4% PFA. Immunostaining was performed as previously described. Primary antibodies and dilutions: fibronectin (mouse monoclonal, BD Biosciences, 1:100), COL5A1 (rabbit polyclonal, Santa Cruz, 1:100), YKL40 (rabbit polyclonal, Quidel, 1:100), human vimentin (mouse monoclonal, Sigma, 1:50), and Ki67 (rabbit polyclonal, Novocastra laboratories, 1:1,000). Quantification of the tumour 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/). Five tumours for each group were analysed. For quantification of Ki67, the percentage of positive cells was scored in five tumours per each group. In histograms the 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.). Immunohistochemistry for primary human GBM was performed as previously described. 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 Signalling, 1:25). Scoring for YKL40 was...
based on a three-tiered system, where 0 was <5% of tumour cells positive, 1 was 5–30% positivity, and 2 was >30% of tumour 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 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 1,000 μm was made with a P1000 pipette tip and images were taken every 24 h with an inverted microscope. For the Matrigel invasion assay, mNSCs and SNB19 cells (1 × 10⁵) were added to the upper compartment of a 24-well BioCoat Matrigel Invasion Chamber (BD Biosciences) in serum-free DMEM. The lower compartment of the chamber was filled with DMEM containing either 0.5% horse serum or 20 μg ml⁻¹ PDGF-BB (R&D systems) as a chemoattractant. After 24 h, invading cells were fixed, stained according to the manufacturer’s instructions and counted. For GBM-derived BTICs, 5 × 10⁴ 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⁻¹ 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 described²⁹.

**Intracranial injection.** Intracranial injection of the SNB19 glioma cell line and GBM-derived BTICs was performed in 6–8-week NOD/SCID mice (Charles River laboratories) in accordance with guidelines of the International Agency for Reserch on Cancer’s Animal Care and Use Committee. In brief, 48 h after lentiviral infection, 2 × 10⁵ SNB19 cells or 3 × 10⁵ BTICs were injected 2 mm lateral and 0.5 mm anterior to the bregma, 3 mm below the skull. Mice were monitored daily and euthanized when neurological symptoms appeared. A Kaplan–Meier survival curve of the mice injected with SNB19 glioma cells was generated using the DNA Statview software package (AbacusConcepts).

---