Nrf2 and *SQSTM1*/p62 jointly contribute to mesenchymal transition and invasion in glioblastoma

Petri Pölönen[1], Ashik Jawahar Deen[2]+, Hanna M Leinonen[2]+, Henna-Kaisa Jyrkkänen[2], Suvi Kuosmanen[2], Mimmi Mononen[2], Ashish Jain[3], Tomi Tuomainen[2], Sanna Pasonen-Seppänen[2], Jaana Hartikainen[4], Arto Mannermaa[4], Matti Nykter[5], Pasi Tavi[1], Terje Johansen[3], Merja Heinäniemi[1], Anna-Liisa Levonen[2]

[1] Institute of Biomedicine, School of Medicine, University of Eastern Finland, P.O. Box 1627, FIN-70211, Kuopio, Finland

[2] Department of Biotechnology and Molecular Medicine, A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, P.O. Box 1627, FIN-70211, Kuopio, Finland

[3] Institute of Medical Biology, University of Tromsø, 9037 Tromsø, Norway

[4] Institute of Clinical Medicine, University of Eastern Finland, P.O. Box 1627, FIN-70211, Kuopio, Finland

[5] Faculty of Medicine and Life Sciences and BioMediTech Institute, University of Tampere, P.O. Box 100, FIN-33014, Tampere, Finland

+ equal contribution

Corresponding author:

Anna-Liisa Levonen, MD, PhD

University of Eastern Finland

A.I. Virtanen Institute for Molecular Sciences

PO Box 1627

FIN-70211 Kuopio

E-mail: anna-liisa.levonen@uef.fi

Tel: +358 40 358 9907

Running title: Nrf2 and *SQSTM1*/p62 in glioblastoma

keywords: *SQSTM1*/p62, Nrf2, Glioblastoma, autophagy, mesenchymal

Additional information

Word count: 239 (Abstract), 4338 (Introduction, Results, Discussion), 2042 (Materials and Methods)

Number of Figures: 7

Number of Supplementary Figures: 11

Number of Supplementary tables: 6

Number of References: 53

Disclosure of Potential Conflicts of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

Conception and design: P.P, M.H, A-L.L

Development of methodology: P.P, A.D., H.L., M.H, M.N.

Acquisition of data: P.P, A.D, H.L, H-K.J, S.K

Analysis and interpretation of data: P.P, A.D, H.L

Writing, review, and/or revision of the manuscript: P.P, M.H, A-L.L, A.D, H.L

Administrative, technical, or material support: M.M, A.J, T.T, S.P-S, J.H, A.M, P.T, T.J

Study supervision: A-L.L, M.H

Abstract

Accumulating evidence suggests that constitutively active Nrf2 has a pivotal role in cancer as it induces pro-survival genes that promote cancer cell proliferation and chemoresistance. The mechanisms of Nrf2 dysregulation and functions in cancer have not been fully characterized. Here, we jointly analyzed the Broad-Novartis Cancer Cell Line Encyclopedia (CCLE) and the Cancer Genome Atlas (TCGA) multi-omics data in order to identify cancer types where Nrf2 activation is present. We found that Nrf2 is hyperactivated in a subset of glioblastoma (GBM) patients, whose tumors display a mesenchymal subtype, and uncover several different mechanisms contributing to increased Nrf2 activity. Importantly, we identified a positive feedback loop between *SQSTM1*/p62 and Nrf2 as a mechanism for activation of the Nrf2 pathway. We also show that autophagy and serine/threonine signaling regulates p62 mediated Keap1 degradation. Our results in glioma cell lines indicate that both Nrf2 and p62 promote proliferation, invasion and mesenchymal transition. Finally, Nrf2 activity was associated with decreased progression free survival in TCGA GBM patient samples, suggesting that treatments have limited efficacy if this transcription factor is overactivated. Overall, our findings place Nrf2 and p62 as the key components of the mesenchymal subtype network, with implications to tumorigenesis and treatment resistance. Thus, Nrf2 activation could be used as a surrogate prognostic marker in mesenchymal subtype GBMs. Furthermore, strategies aiming at either inhibiting Nrf2 or exploiting Nrf2 hyperactivity for targeted gene therapy may provide novel treatment options for this subset of GBM.

Introduction

The Kelch-like ECH-associated protein 1 (*KEAP1*/Keap1) – Nuclear factor E2 related factor 2 (*NFE2L2*/Nrf2) pathway is the major regulator of cellular defense against endogenous and environmental oxidative and electrophilic stress. In basal conditions, Nrf2 is located in the cytosol and is bound to the inhibitory Keap1 protein, which facilitates the proteasomal degradation of Nrf2 via the Cullin3 (Cul3) –based E3 ubiquitin ligase complex. Upon activation by oxidative or

3

electrophilic stress, critical thiol residues in Keap1 are modified rendering it unable to function as a substrate adaptor for the E3 ligase, which leads to Nrf2 stabilization and the consequential increased activity. Transcriptional induction of Nrf2 target genes provides resistance against endogenous and exogenous cytotoxic stresses [1,2]. In addition to Keap1-dependent regulation of Nrf2, various alternative Keap1-independent mechanisms contribute to the regulation of Nrf2. These include the binding of disruptor protein p62 [3,4] on Keap1, which inhibits the interaction of Nrf2 and Keap1 causing an increase in cellular Nrf2 protein and its activity.

It has become evident that malignant cells benefit from having increased Nrf2 pathway activity. This was first observed in lung cancer [5,6], as well as subsequently in many other cancer types, such as pancreatic, ovarian, liver and gallbladder cancers [7–9]. Aberrant Keap1-Nrf2 signaling leads to radio- and chemoresistance and provides growth advantage to cancer cells, due to constitutive expression of cytoprotective genes [2,9,10]. Multiple mechanisms for Nrf2 overactivation have been found, such as somatic mutations in either *KEAP1* or *NFE2L2*, deletion of exon 2 of *NFE2L2*, aberrant expression of inhibitory proteins and transcriptional induction by oncogenes and hormones [2,11].

Glioma is a type of tumor in the brain originating from malignant glial cells. Gliomas are classified to WHO grades I-IV by cell type, morphology and mutations. Low-grade gliomas are grade I-II and are well-differentiated whereas high grade gliomas are grade III-IV and are undifferentiated. GBM is a grade IV tumor and is the most common form of brain cancer among adults, and has an extremely poor prognosis, the median survival from diagnosis being only 15 months[12]. GBM is classified into five subtypes: mesenchymal, classical, proneural, neural and G-CIMP, based on the gene expression, DNA methylation and mutation patterns that they possess [13–15]. The mesenchymal subtype is characterized by the expression of mesenchymal and angiogenic genes and is highly invasive. Retaining mesenchymal properties have been proposed to be crucial for the survival of cancer cells, as stem cell-like cells are resistant to treatment and promote tumor regrowth and relapse [16].

In this study, we developed and benchmarked Nrf2 activation signature, based on gene expression profiles, and utilized it to identify cancer types with overactive Nrf2 pathway across 915 samples in the Broad-Novartis Cancer Cell Line Encyclopedia (CCLE) [17]. In addition to cancer cell types having a previously well-established Nrf2 overactivity profile, we found that Nrf2 is also hyperactive in cancer cells of glial origin. We confirmed this in glioma patient samples available from independent studies, namely the Rembrandt, the Cancer Genome Atlas (TCGA) GBM study [18] and the Human Glioma Cell Cultures (HGCC) collection [19]. In addition, we interrogated the mechanisms behind Nrf2 hyperactivity and the functional role of this transcription factor in glioma.

Results

Benchmark of Nrf2 activity inference methods

We combined data from adenoviral overexpression of Nrf2 for 36 and 72 h and short term treatment with the Nrf2 activator nitro-oleic acid (OA-NO$_2$) in HUVEC to form a group with activated Nrf2 [20] and compared this group against samples with basal Nrf2 activation state. We tested a probit regression model [21] to infer Nrf2 activation using a supervised approach using the different Nrf2 activation states as training data. NFE2L2 and KEAP1 mutations are the predominant mechanism that disrupts the Keap1-Nrf2 interaction (causing Nrf2 overactivity) in lung squamous (LUSC) and adenocarcinomas (LUAD) as well as the head-and-neck squamous carcinoma (HNSC), making these cancer types the ideal benchmarking sample sets for which the prediction outcome (Nrf2 activity) can be estimated from mutation data. The Receiver-Operator-Curve (ROC) analysis (Fig. S1A) shows that high Nrf2 activity score (> 80% probability) obtained with the regression model recovers mutated samples with high specificity and sensitivity (81, 76 and 79% sensitivity with 88, 86 and 80% specificity in LUAD, HNSC and LUSC, respectively). Additionally, we tested a gene set-based approach using several existing Nrf2 gene sets and different enrichment methods. The NFE2L2.V2 and BIOCARTA_ARENRF2_PATHWAY gene sets included in MsigDB had poor performance (data not shown). However, SINGH_NFE2L2_TARGETS pathway and GSVA tool were good predictors of

Nrf2 activation (Fig. S1B-D). Therefore, Nrf2 activation could be inferred using both supervised and unsupervised gene set-based methods from patient data.

Nrf2 activity prediction across a panel of cancer cell lines

We applied the regression model to identify cell lines displaying Nrf2 overactivity, from CCLE sample collection which contains gene expression profiles from 915 cell lines representing 36 cancer types (Fig. 1A-B, Table S1). GSVA resulted in a highly similar prediction, as 90% of the cell lines predicted by our approach were significantly enriched (FDR<0.05). We defined 5 Nrf2 activity categories based on the transcription factor (TF) activity probability, as shown in Fig. 1C. In 66 cell lines (constituting about 7% of all cell lines) representing 12 cancer types our analysis predicted to have an over 95% probability of overactive Nrf2. Using a less stringent cutoff of > 80% probability, we identified a total of 174 cell lines (approximately 19% of cell lines examined) (Fig. 1D). The predicted activity in kidney, esophagus, liver and lung cancers was in agreement with previous studies, where mutations of either *Nrf2* or *KEAP1* have been detected [6,22–24]. Nrf2 activity was high also in 10.6% of the glioma cell lines. Since the mechanisms of Nrf2 dysregulation in glioma remained ill understood, we were motivated to further analyze the role of Nrf2 activity in glioma.

Nrf2 is constitutively activated in glioma cell lines and GBM patient samples

First, we determined how glioma WHO grade I-IV correlates to Nrf2 overactivity using 550 patient microarray gene expression profiles from the Rembrandt dataset. Interestingly, Nrf2 activity was gradually higher in grades II-IV, whereas in grade I none of the patient samples had high Nrf2 activity (Fig. 1E). As a confirmation, Nrf2 activity was inferred using RNAseq data from 166 grade IV TCGA GBM samples. Similarly to the Rembrandt grade IV tumors where Nrf2 activity was high in 13% of the samples, 17% of the samples had high Nrf2 activity in TCGA GBM samples and the highest activity with P>95% was detected in 4% in TCGA and 1.6% in the Rembrandt datasets (Fig. 1D). We confirmed that the Nrf2 activity marker genes correlate with Nrf2 activity (Fig. 2A-B). Similarly, patient derived xenograft HGCC 48 cell lines were used to verify Nrf2 activity in GBM cell lines using GSVA (Fig. S1).

As an experimental confirmation of the predicted Nrf2 hyperactivity, we chose three glioma cell lines with > 95% probabilities (H4, T98G and GAMG, referred to as Set 1) and four glioma cell lines with > 80% probabilities of above average Nrf2 activity (LN18, A172, U87 and U251, Set 2) for experimental verification of Nrf2 activity status. Immortalized normal human astrocyte cells (NHA-cells) and a glioma cell line with a predicted low Nrf2 activity (DKMG) were chosen as negative controls, whilst the A549 lung adenocarcinoma cells with known Nrf2 hyperactivity[6] served as a positive control. First, we evaluated the level of Nrf2 activity using the luciferase reporter assay with a lentiviral luciferase reporter construct under the control of minimal SV40 promoter and four Nrf2 binding sites (AREs) derived from the human *GCLM*[25]. Robustly elevated activity was found in A549 and Set 1 cell lines compared to NHA cells, and slightly increased activity in 2 of 4 (LN18 and A172) Set 2 cell lines (Fig. 2C). Next, Nrf2 target levels were analyzed by western blotting (Fig. 2D) and RT-qPCR (Fig. S2). GCLM, NQO1 and HMOX1 proteins were increased in all Set 1 glioma cell lines, with a more variable pattern in Set 2. Moreover, the binding of Nrf2 on ARE on *NQO1* and *HMOX1* promoters was studied using chromatin immunoprecipitation (ChIP) in T98G, GAMG and DKMG cells. The binding of Nrf2 on *NQO1* and *HMOX1* ARE was significantly increased in T98G and GAMG cells compared to DKMG (Fig. 2E). In summary, the cell lines with > 95% probability (H4, T98G and GAMG cells) each showed evidence of overactivated Nrf2 pathway, whereas cell lines selected for validation from Set 2 (displaying lower probability thresholds) had high Nrf2 activity in 2 out of the 4 cell lines examined (LN18, A172) on reporter assay. However, these cell lines had more variable patterns of target gene expression.

Nrf2 activity is linked to Nrf2 and Keap1 gene expression imbalance

As the percentage of Nrf2 overactive samples was the highest in grade IV tumors, we next investigated the potential mechanisms of activation in GBM using multi-level TCGA patient data and cell lines from CCLE and HGCC. GBM samples lacked mutations in *NFE2L2* and *KEAP1* in TCGA

patient data. We obtained similar results when glioma cell lines T98G, GAMG, LN18, DKMG and NHA control cells were sequenced (data not shown).

NFE2L2 and *KEAP1* mRNA expression showed a negative relationship in TCGA and HGCC data and in H4 and T98G cell lines (Fig. 3 A-B, Fig. S3). *KEAP1* copy-number gain was associated with concomitant increase in its mRNA and Nrf2 inactivity, whilst *KEAP1* expression was significantly lower in patients representing Nrf2 overactive samples and not attributed to differential methylation (data not shown). We found a deletion of *KEAP1* in H4 cell line and amplification of *NFE2L2* in T98G and to validate the effect on gene expression, we performed RT-qPCR for *NFE2L2* and *KEAP1*. We found highly elevated expression of *NFE2L2* in T98G cell line, but *KEAP1* mRNA was significantly decreased only in LN18 and significantly upregulated in A172 (Fig. 3C-D). We also compared the protein levels relative to control cell lines with western blotting. Nrf2 was over-expressed in all Set 1 and in one Set 2 cell lines, (Fig. 3E, in H4, T98G, GAMG and LN18). Moreover, H4 and LN18 showed decreased Keap1 protein, whereas in T98G and GAMG Keap1 protein was approximately at the same level or increased compared to controls (Fig. 3F). Taken together, the observed Nrf2 hyperactivity could be linked to expression imbalance of *NFE2L2* and *KEAP1* within the patient cohorts, which was also evident in cell line models.

Pathway analysis reveals central cancer processes correlated to Nrf2 activity

We performed enrichment analysis of oncogenic signatures and biological pathways using GSEA to identify processes linked to Nrf2 activity. Significant pathways (FDR<0.05) from TCGA, CCLE and HGCC datasets were linked to larger biological processes and ranked based on their mean enrichment score (Fig. 4A). The highest ranked cancer-related process was mesenchymal transition, suggesting that increased Nrf2 activity could be linked to the mesenchymal subtype, or contribute to mesenchymal transition. Proliferation was linked to Nrf2 activity through cell cycle and energy metabolism. Importantly, our analysis also associated invasion with Nrf2 activity, together with redox and other stress responses that could be clinically relevant in drug resistance. Several pathways

were also linked to proteasomal degradation and autophagy. We hypothesize that these processes could contribute to Nrf2 activation or function as key Nrf2 regulated pathways in GBM.

Autophagy dysregulation is linked to Nrf2 activation

To further investigate the link between autophagy and Nrf2 overactive cells, we examined the expression of the autophagy-related protein, microtubule-associated protein light chain 3 (LC3) (Fig. 4B). Two bands of LC3 can be separated in Western blots, the upper migrating band (LC3-I) representing cytosolic LC3 and the lower migrating band (LC3-II) is present in isolated membranes and autophagosomes [26]. In addition, the treatment of cells with Bafilomycin A (BafA), which inhibits the fusion of autophagosome with lysosomes [27], was together with the LC3 Western blot used for evaluating whether autophagy was functional. The protein level of LC3-II was accumulated in basal conditions in GAMG and U-87 MG cells. In all cell lines, except for GAMG, there was an increase in LC3-II in response to BafA (Fig. 4B), indicating functional autophagy. We also analyzed H4, T98G, GAMG, LN18 and NHA cells with transmission electron microscopy to confirm the presence of autophagosomal structures (Fig. 4C). In agreement with the LC3-II data, we found that autophagosomal and autolysosomal structures were abundant in GAMG and T98G cells, but not in LN18 or NHA cells. In addition, in GAMG cells, a high number of dark amorphous endosomal/lysosomal-like structures were present, which in combination with the LC3 Western blot results indicates that the autophagosomal pathway is dysfunctional in GAMG cells.

The *SQSTM1*/p62 and Nrf2 positive feedback loop regulates Nrf2 activity in glioma

Sequestosome1 *(SQSTM1*/p62) is a stress-inducible and multifunctional protein, that selectively targets proteins for autophagy and ubiquitin degradation [28]. It has been previously reported that in HCC mTOR activated p62 can bind to Keap1 and promote autophagosomal degradation of Keap1, leading to constitutive activation of Nrf2 [4,24,29]. We observed increased p62 protein and *SQSTM1* mRNA levels in Nrf2 overactive cell lines H4, T98G and GAMG but also in A172 and U-87 MG cell lines (Fig. 5 A-B). Strong significant correlation between *SQSTM1* and Nrf2 activity was observed in

all datasets (GBM shown in Fig. 5C). We also noted strong correlation (FDR = 0.000024) between the protein levels of p62 and Nrf2 activity in 82 TCGA GBM Reverse phase protein lysate microarray (RPPA) patient samples (Fig. 5 D). To compare Nrf2 and p62 association strength to other diseases, we performed a correlation analysis between Nrf2 GSVA score and *SQSTM1*/p62 expression within 33 cancer types included in the TCGA cohorts. The correlation between Nrf2 activity and *SQSTM1*/p62 was particularly strong in GBM and in Kidney renal papillary cell carcinoma (KIRP), whereas the correlation was poor in frequently Nrf2 overactive HNSC, suggesting that the p62-dependent Nrf2 regulation is important in select cancers (Fig. 5 E). Based on the genetic information, elevated *SQSTM1* mRNA in A172 is likely due to *SQSTM1* amplification (Fig. 5 E). Given that *SQSTM1* has been shown to be also transcriptionally regulated by Nrf2[30], we confirmed, using RT-qPCR, that the expression of *NFE2L2* and its target genes were significantly reduced by Nrf2 and p62 siRNA knockdown in these cells and *SQSTM1* gene expression was reduced by Nrf2 knockdown in T98G cells (Fig. S4).

Next, we used co-immunoprecipitation to determine whether Keap1 and p62 are bound to each other in T98G and GAMG cells having upregulated Keap1-Nrf2 pathway, and in A172 in which Nrf2 activity is normal but p62 is upregulated. We found that these proteins interact with each other in all three cell lines, further supporting the importance of p62 in Keap1 protein turnover (Fig. 5G). Moreover, as the Ser-351 phosphorylation of p62 has been shown to increase the binding affinity between Keap1 and p62[29], we also studied the Ser-351 phosphorylation of p62 in all glioma cell lines and found a clear Ser-351 and Ser-403 phosphorylation of p62 in T98G (Fig. 5H). We also studied co-localization of p62 and Keap1 by confocal immunofluorescence analysis. T98G cells contained a high number of large p62-Keap1 aggregates, whereas in other cell lines co-localization was less evident (Fig. 5I, Fig. S5).

Nrf2 and p62 jointly regulate mesenchymal transition in the mesenchymal subtype of GBM

Mesenchymal transition was ranked as one of the top significantly affected pathways (Fig. 4A). Therefore, we examined the prevalence of Nrf2 overactivity in different molecular subtypes of GBM. Our analysis revealed that Nrf2 activity correlated with samples assigned to the mesenchymal subtype, accounting for 83% of Nrf2 active samples in TCGA data (Nrf2 highest, Fig. 6A). All HGCC samples with high Nrf2 activity were of mesenchymal subtype (Fig. S6A). Nrf2 activity anti-correlated with the classical subtype, suggesting that there could be a link between Nrf2 activity and pathway profiles of mesenchymal and classical subtypes. In comparison to the mutation data (Fig. 2), Keap1 copy-number gain (associated with Nrf2 inactivity) was also correlated with the classical subtype. *NF1* is typically mutated in mesenchymal subtype[14], but not in Nrf2 overactive samples (Fig. S6B). We correlated also TCGA protein array data to Nrf2 activity and identified that the expression of endothelial VEGFR2 and PECAM1 and cell adhesion molecule P-Cadherin is significantly correlated, whereas NOTCH1 and EGFR expression significantly anticorrelated with Nrf2 activity (Fig. S7).

To compare Nrf2 and mesenchymal transition association to other diseases, we performed a correlation analysis between Nrf2 GSVA score and epithelial to mesenchymal transition (EMT) associated gene sets within 33 cancer types included in the TCGA cohorts. Correlation between Nrf2 activity and EMT was particularly strong in GBM and LGG, whereas EMT was significantly anticorrelated to Nrf2 activity in HNSC and LUSC, suggesting that the Nrf2 activity correlation to EMT is context dependent and highly specific to gliomas (Fig. 5 E).

We then analyzed whether the mesenchymal phenotype, as defined by mesenchymal markers such as *SNAI1*/Snail, *SNAI2/*Slug, *CTNNB1/*ß-catenin and *VIM/*Vim, are linked to Nrf2 and p62 expression. In both the GAMG and T98G cell lines, Nrf2 and p62 siRNA knockdown resulted in a significant reduction of both the mRNA and protein levels of these markers (Fig. 6C, Fig. S6C-D). We used MAF, BACH1 and Nrf2 ChIP-seq data from the Encyclopedia of DNA Elements at UCSC (ENCODE 3) together with U87 and T98G Global run on sequencing (GRO-seq) data from[31] for finding putative ARE containing enhancer for the mesenchymal markers. We identified putative Nrf2

binding sites in the proximal enhancer regions of *SNAI2* and *CTNNB1* (Fig. S8). The binding of Nrf2 to these regions was significantly higher in Nrf2 active T98G and GAMG cells compared to DKMG with low Nrf2 activity, assessed by ChIP-PCR (Fig. 6D). Taken together, Nrf2 is involved in EMT regulation in glioma and directly regulates SLUG and ß-catenin expression.

Nrf2 overactivity affects progression free survival in GBM patient samples and increases cell proliferation and invasion in glioma cell lines

Next, we analyzed TCGA GBM prognostic clinical data stratified by the Nrf2 activity status. We merged Nrf2 overactive and low activity samples from TCGA microarray and RNAseq data to increase the cohort size to 520 cases. We found no difference in the overall survival of Nrf2 active patient samples. However, patients with the highest Nrf2 activity (22/520) were strongly associated with decreased progression free survival (likelihood ratio test 0.0022, HR 2.6) (Fig. 7A). Multivariate analysis using the RNAseq samples of Nrf2 activity probability and glioblastoma subtypes revealed that Nrf2 activity probability is significantly linked to progression also when considering the disease subtypes (Fig. S9A). We also considered other prognostic factors, such as age, gender, IDH mutation status, recurrence, MGMT methylation status and subtype together with Nrf2 activation categories. The highest Nrf2 activity remained a significant predictor of faster disease progression (Likelihood ratio test, overall model p-value 1e-8, Nrf2 Highest category p-value 0.0013) (Fig. S9B).

The pathway analysis associated Nrf2 activity with proliferation and invasion (Fig. 4A and Table S2). We found significant positive correlation to G6PD protein expression (Fig. S7), a known promoter of cellular proliferation. Accordingly, Nrf2 and p62 siRNA knockdown in GAMG and T98G cells resulted in a significant decrease in cell proliferation at 24 to 96 h, with the most pronounced reduction observed at the 24 h time point (Fig. 7B). T98G cells were able to invade into 3-dimensional type I collagen-cultrex matrix, as judged by actin cytoskeleton and cellular nuclei staining. Upon Nrf2 and p62 siRNA knockdown, cell invasion was significantly impaired (Fig. 7C). Next, we analyzed anchorage-independent growth in GAMG and T98G cells by growing the cells in ultra-low attachment plates. Both cell lines formed 3-dimensional spheres in serum-free growth conditions (Fig. 7D and

Fig S10). In conditions of Nrf2 and p62 siRNA knockdown, the growth of these spheres was significantly inhibited (Fig. 7D and Fig. S10). Results were confirmed using different siRNAs (Fig. S11).

Discussion

In this study, we show that Nrf2 pathway is overactive in a subset of glioma, mainly in GBM mesenchymal subtype. Furthermore, p62 creates a positive feedback loop with Nrf2 in glioma, capable of promoting tumorigenesis and stemness properties, as summarized in Fig 7E.

Nrf2 overactivity was prevalent in the mesenchymal subtype, which could be explained by the ability of Nrf2 and p62 to promote stem cell markers associated with this subtype. Interestingly, mesenchymal subtype had higher Nrf2 expression and lower Keap1 expression, suggesting that Nrf2 upstream regulation promotes increased Nrf2 activity in the mesenchymal subtype. On the contrary, Nrf2 inactivity correlated with the classical subtype, with gene expression imbalance favoring low Nrf2 activity and high NOTCH and EGFR levels.

Autophagy and lysosomal pathways and serine/threonine signaling correlated with Nrf2 activity both in cell lines and in patients. Autophagosomal and autolysosomal structures were found in glioma cell lines, but autophagosomal pathway was defective only in GAMG cells. Thus, in GAMG cells, due to defects in autophagy, p62 accumulates, binds to Keap1 resulting in Nrf2 release and observed hyperactivity (Fig. 5G). Interestingly, Nrf2 also regulates several autophagy genes in addition to *SQSTM1*[32], which is likely to impact the strength of the Nrf2 and p62 feedback and contribute to autophagy dysregulation. In comparison, in T98G cells autophagy was intact, whereas increased p62 -Keap1 interaction could be attributed to p62 Ser-351 and Ser-403 phosphorylation. Nrf2 and p62 dependency on mTOR signaling has been reported in lung cancer[33]. Furthermore, RTK/RAS/PI3K is altered in 88% of glioma patients and drives growth in GBM[18]. Increased activation of the MAPK pathway with increased phosphorylated RAF/MEK/ERK has also been reported in GBM, especially in the mesenchymal subtype[18]. Similarly, we found significant positive correlation

13

between Nrf2 activity and phosphorylated A-RAF and MEK expression (Fig. S7). Therefore, several kinases could influence Nrf2 activity, but functional redundancy in signaling networks and their connectivity to p62 may explain why Nrf2 activity may arise via alternative mechanisms in GBM. Taken together, both autophagy dysregulation and constitutive serine/threonine signaling targeting of Ser-351 and Ser-403 are likely to promote Nrf2 overactivation through p62 in glioma.

A previous study of 10 GBM and 95 anaplastic glioma samples[34] indicated that p62 accumulation is not linked to Nrf2 overactivation. Instead, *IDH1* gene mutations were shown to affect Nrf2 activity in anaplastic glioma, but not in GBM. Accordingly, Nrf2 target genes were shown to correlate with poor prognosis in anaplastic glioma but not in GBM. The different conclusions regarding p62 correlation to Nrf2 activity in GBM may be due to differences in cohort sizes. We have shown in CCLE, TCGA and HGCC data sets that correlation between p62 and Nrf2 exists, and validated the dependency between Nrf2 and p62 expression levels. Furthermore, mutations in the *IDH1* gene are not correlated to Nrf2 activity in our analysis of GBM, confirming previous findings[34].

Our results demonstrate that in GBM cells, inhibition of Nrf2 and p62 decreased tumorigenic properties, such as cell invasion and anchorage-independent growth. This could be partially explained by the fact that mesenchymal proteins such as Snail, Slug, vimentin and beta-catenin are reduced upon siRNA knockdown of Nrf2 and p62. High expression of autophagy regulators including p62 and increased MAPK pathway activation have been previously associated with aggressive properties in mesenchymal subtype of GBM[35]. p38-MAPK and Nrf2 pathway activation in response to ROS has been previously shown to increase glioma survival and promote glioma reprogramming towards the mesenchymal phenotype[36]. In line with these findings, our discovery of Nrf2-p62 joint contribution in promoting mesenchymal properties could reveal how GBM mesenchymal subtype acquire an aggressive phenotype that is resistant to treatments.

On a functional level, our data shows that inhibition of Nrf2 and p62 resulted in defective proliferation in GBM cells, supporting the idea that Nrf2 is involved in metabolic rearrangements to support

proliferation. Metabolic reprogramming is a major hallmark in glioma, which is characterized by upregulated glycolysis, PPP (pentose phosphate pathway) and lipid metabolism, among others[37]. Activation of these metabolic pathways provides molecular building blocks for the high anabolic demands of cancer cells. PPP activation is also an important source for NADPH generation, which is required for antioxidant generation and in anabolic pathways as a reducing agent[38]. Key metabolic genes, such as *G6PD*, *PGD*, *TALDO1* and *ME1* are direct targets of Nrf2 and direct flux from glycolysis to PPP [39]. Our analysis revealed a high correlation between G6PD expression and Nrf2 activity, confirming previous findings. Nrf2 could also function as a key balancing factor in metabolic reprogramming, as Nrf2 is capable of regulating both energy metabolism and antioxidant response to ROS to favor glioma growth and development. Cancer cells tolerate moderate oxidative stress and increased ROS levels contribute to proliferation, genetic instability and evasion from senescence [40]. However, excessive ROS exposure causes oxidative damage and apoptosis[41,42]. The ability of Nrf2 to simultaneously produce NADPH through increased PPP flux to prevent excess oxidative damage could be a crucial protective mechanism. Consequently, the ability of Nrf2 to regulate enzymes involved in the detoxification of drugs would support radio- and chemotherapy resistance in gliomas harboring overactivity of this transcription factor. In line with this, we have shown that patients with the highest Nrf2 activity have decreased progression free survival in TCGA cohort, suggesting that treatments have limited efficacy in these patients. Also, Nrf2 activity increased gradually in glioma grades II-IV. Therefore, to fully assess the extent of Nrf2 activity and its clinical impact, it would be important to study Nrf2 activity in relapsed GBM samples and during treatment.

Nrf2 pathway has been widely studied using multi-omics cancer datasets[11,43,44]. However, these studies are limited to samples with mutations in either the *NFE2L2* or *KEAP1*. In this study, we performed computational predictions of Nrf2 activity in cell line and patient samples, using Nrf2 target gene signatures. This approach could be widely applied to study oncogene activation mechanisms and downstream effects using multi-omics data sets, such as TCGA and CCLE. Our approach has power especially in cancers that have no somatic mutations in the Keap1-Nrf2 pathway components such as GBM and enable investigation of indirect mechanisms of Nrf2 activation.

The discovery of Nrf2 and p62 as key components of the mesenchymal subtype molecular network as well as the evident role of Nrf2 in metabolic reprogramming will facilitate new discoveries in understanding how GBM acquires treatment resistance and invasive phenotype.

Materials and Methods

Reagents

All reagent full names and abbreviations, company, city and country are listed in Table S3.

Cell lines and culture conditions

Human endothelial cells (HUVEC) were isolated and grown as described previously [45]. Human GBM cell lines U87, U251, H4, LN18, A172 and lung carcinoma cell line A549 were cultured in DMEM supplemented with 10% (v/v) FBS and 1% penicillin/streptomycin. T98G cells were cultured in MEM supplemented with 10% FBS, 2 mM L-glutamine, 4x MEM nonessential amino acids and 1% penicillin/streptomycin, DKMG were cultured in RPMI-1640 medium supplemented with 10% FBS, 2mM L-glutamine and 1% penicillin/streptomycin; and NHA cells were grown in DMEM supplemented with 10% FBS, 2mM L-glutamine, 50mg/ml hygromycin and 1% penicillin/streptomycin.

Microarray analysis to define an activity signature for Nrf2

Microarray profiling was performed from adenoviral overexpression of Nrf2 in HUVECs pooled from three separate donors for 36 and 72 h using the Affymetrix U133 plus 2.0 platform. For the analysis, cells were transfected with CMV-promoter-*Nrf2* or CMV-promoter-only containing adenoviral expression vectors using multiplicities of infection (MOI) 50. Triplicate samples were used. Sample collection and microarray analysis were performed as in [45]. Raw data files were normalized using GC-RMA and gene level expression values were obtained using Brainarray probe mapping. These results and microarray data from [45] consisting of transient activation by nitro-oleic acid (OA-NO$_2$) were combined for identifying an Nrf2 dependent gene signature. The Nrf2 negative training set

contained adeno-CMV transduced, or solvent and oleic acid (OA) treated samples from the two experiments, respectively.

Datasets

CCLE data was downloaded from Gene expression Omnibus (GEO, series GSE36133) to infer Nrf2 activity across various cancer types. Copy number results were downloaded from CCLE data portal (Nov 2014). Level 3 TCGA data was downloaded from TCGA data portal and clinical and sample annotation and subtype data was obtained from the publication supplementary table 7[18]. Processed HGCC gene expression and sample annotation data was downloaded from http://www.hgcc.se/. Copy number values for each DNA segment were downloaded using id GSE72209. The Rembrandt study gene expression profiles and clinical data were downloaded using the GEO database (GSE108474) and normalized using RMA and gene expression values obtained using Brainarray probe mapping. TCGA GBM Affymetrix U133 plus 2.0 platform samples processed data matrix was downloaded from http://gliovis.bioinfo.cnio.es[46]. Pan-Cancer Atlas containing RNAseq and RPPA profiles for TCGA 33 diseases were downloaded (June 2019) from https://gdc.cancer.gov/about-data/publications/pancanatlas. U87 and T98G GRO-seq bedGraph files were downloaded from GSE92375.

Predicting Nrf2 activity in cancer data sets

Using the compiled microarray data, the command line version of Binreg2 algorithm and SIGNATURE tool [21] were used to distinguish samples with high probability of Nrf2 activity (Table S1). The gene set variation analysis (GSVA) [47] available in the R/Bioconductor package GSVA 1.22.4, and SINGH_NFE2L2_TARGETS gene set from MsigDB[48] were used to assign a Nrf2 gene set enrichment score in a sample-wise manner with the following settings: mx.diff=F, tau=0.25, rnaseq=F for CCLE, HGCC and TCGA pancan33 datasets. Random gene-set permutations were computed to obtain FDR values to estimate significance of the enrichment scores for HGCC data.

ROC analysis comparing Nrf2 activity predictions on a gold standard gene set

Gene set variation analysis (GSVA) available as an R/Bioconductor package contains several enrichment analysis implementations, including GSVA, ssGSEA and combined Z-score. These tools were used to compare available Nrf2 gene sets (SINGH_NFE2L2_TARGETS, NFE2L2.V2, Nrf2 canonical) to the Bayesian probit regression using the ROCR R package. A gold standard set was used for the comparison consisting of sequenced luad, lusc, hnsc samples (n=483, n=178, n=489, respectively) with high rate of NFE2L2-KEAP1 missense mutations (77, 45, 43, respectively), that are previously described to disrupt the Nrf2-Keap1 interaction and lead to Nrf2 overactivation. Mutated samples were considered as true positives and other samples as true negatives.

GSEA

Gene Set Enrichment Analysis (GSEA) [49] was used to perform pathway analysis. Nrf2 activity was used as a continuous phenotype in CCLE, HGCC and TCGA data using a Pearson correlation as metric and sample permutation to obtain FDR q-values. For this purpose, gene sets from MsigDB v5.0, Wikipathways (06.2015), Recon 1 and Pathway commons 7 were obtained.

Kaplan-Meier survival analysis

Survival and progression time and status for each TCGA GBM sample was obtained from the supplementary table 7 from the publication[18]. The R package 'survival' was used to compute univariate and multivariate cox proportional hazard model to test whether Nrf2 activity affects patient overall and progression free survival. Kaplan Meier curves were used to visualize differential Nrf2 activity groups. Proportional Hazards Assumption of a Cox Regression was verified using the cox.zph weighted residuals method in R.

RNA isolation and quantitative real-time PCR

Total RNA was extracted with TRI-reagent. One microgram of the RNA was used for cDNA synthesis by Transcriptor First Strand cDNA synthesis kit. The relative expression levels were measured by qPCR using relative standard curve method (StepOnePlus Real-Time PCR Systems, Applied Biosystems) with specific primers and probes listed in Table S4.

Western blotting

Cells were lysed, and total protein concentration was measured with BCA assay (Pierce). Western blotting was performed as previously described [25] with the exception that PVDF membrane was used for LC3B detection instead of nitrocellulose membrane.

Luciferase reporter gene assay

Luciferase reporter gene assays were performed as previously described [25]. To test transduction efficiency in each cell line, a lentiviral construct expressing GFP under human PGK promoter was used [50]. MOI yielding >90% of GFP positive cells in FACS analysis were used for transduction. MOI values are shown in Table S5.

siRNA transfections

Cells were transfected with siRNA oligonucleotides targeting *NFE2L2* or *SQSTM1* gene expression, or a non-specific siRNA control (20 nmol/l) using RNAiMax according to manufacturers' instructions. The siRNA sequences used were as follows: siNrf2 (catalog no. s9493 and s9492, Thermo Scientific) and sip62 (catalog no. s16960 (Thermo Scientific) and 5′-GCAUUGAAGUUGAUAUCGAUTT-3' (Invitrogen)). Cells were collected at 48 h time point for q-RT-PCR or for Western blotting.

Fluorescence microscopy

Cells were fixed in 4% paraformadehyde in PBS, permeabilized with 0.2% Triton-X-100 in PBS and blocked with FBS prior to incubation with mouse monoclonal p62 antibody and goat polyclonal Keap1 antibody o/n in +4°C. After washing with PBS, samples were incubated with secondary antibody Alexa Fluor® 488 Donkey anti-goat IgG (H+L) and Alexa Fluor® 647 rabbit anti-mouse IgG (H+L) for one hour and with DAPI for five minutes. Samples were examined with Olympus FV 1000 and Zeiss LSM 700 confocal microscopes.

Transmission electron microscopy

Cells were prefixed in 2.5% glutaraldehyde in 0.1M phosphate buffer (pH 7.4), postfixed in 1% osmium tetraoxide for 1 hour and dehydrated with ethanol. The samples were infiltrated and embedded in Epon LX-112. After polymerization in +60°C for 48 hours, samples were stained with 1% uranyl acetate and lead citrate. Finally, samples were examined with Jeol JEM-1200EX (80 kV).

Co-immunoprecipitation

Co-immunoprecipitation of p62 and Keap1 was performed as in [51].

Cell proliferation assay

GAMG and T98G cells were plated on 6-well plates and treated with siRNAs the following day. Cells were then allowed to grow in reduced serum medium (4% FBS containing growth medium) for 24, 48, 72 and 96 h and counted using a hemocytometer. Results are expressed as cell numbers/ml.

Cell invasion assay in 3D collagen-cultrex matrix

T98G cells, grown and treated with siRNAs in 6-well plates, were trypsinized after 48 h of treatment and plated 60,000 cells/well on 8-well Ibidi chambered μ-slides. 24 h later, the confluent culture of cells was coated with a 1:1 mix of of 2.4 mg/ml each of 3D matrix containing type I collagen and cultrex for an hour at 37°C, according to manufacturer's instructions. Later, growth medium was added on top of the solidified matrix and the cells were allowed to invade the matrix for 24 h. Later, cells were fixed with 4% paraformaldehyde in 0.1M phosphate buffer (PB) for 1.5 h, washed with PB several times, permeabilized with 0.1% TritonX-100 in 1% bovine serum albumin (BSA)-PB for 30 min, blocked with 1% BSA-PB for 40 min, stained with Alexafluor 594-conjugated Phalloidin (Molecular Probes) in PB (according to manufacturer's instructions) for 1.5 h and then with DAPI (1μg/ml) for 15 min. The slides were stored at 4°C until use. 20x optical sections with an interval of 2μm and a range of 55-65 slices were taken in Z direction using a Zeiss LSM 700 confocal microscope and the images were stacked to get a 3D projection. Using ImageJ, the mean fluorescence of Alexafluor 594-Phalloidin and invasion area of the cells were measured. An average of 15 images/group from 3 independent experiments were used.

3-dimensional spheroid formation assay

GAMG and T98G cells, grown and treated with siRNAs in 6-well plates, were trypsinized after 48 h of treatment and plated respectively as 1000 cells/well and 2500 cells/well in serum-free growth medium in Ultra-low attachment 96-well plates. 6 replicates/group were used in a single experiment. Growth of spheres at day 7 was imaged using a 10X objective in a standard light microscope. The experiment was repeated five times and the area of spheres was measured using ImageJ program.

Analysis of EMT markers

The levels of *SNAIL, SLUG* and *ZEB1* mRNA were analyzed by qPCR from three repeats. *GAPDH* expression was used for normalization. Three repeats of western blot analysis were done for mesenchymal markers: snail, slug, vimentin and CTNNB1/beta-catenin, with beta-actin and lamin-B1 used for normalization.

Chromatin Immunoprecipitation

ChIP was performed as previously described in [52] with following modifications: Prior to immunoprecipitations, 20 µl of Magna ChIP magnetic beads (Millipore) per immunoprecipitation were re-suspended to 1 ml of PBS/BSA (5 mg/ml). Beads were washed with PBS/BSA and re-suspended in 2 ml of PBS/BSA. 5 µg of antibody (Nrf2, sc-722, or anti-rabbit IgG, Sc-2027, Santa Cruz Biotechnologies) was added to each tube. The tubes were incubated on rotating platform o/n at + 4 °C. The next day, the beads were washed twice with PBS/BSA and re-suspended in 100 µl of PBS/BSA. T98G, GAMG, DKMG and NHA-empty cells were grown on 10 cm plates. After crosslinking, nuclei were extracted by scraping the cells to 1 ml of MNase buffer (10 mM Tris pH 7.4, 10 mM NaCl, 5 mM $MgCl_2$, 0.1 % NP-40, protease inhibitors) and, after 10 min incubation on ice, the nuclei were pelleted by centrifugation (1500 x g, 5 min, + 4 °C). The extracted nuclei were washed with 1 ml MNase buffer and, after centrifugation, lysed with 0.3 ml SDS lysis buffer (1 % SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.1, protease inhibitors). The lysates were sonicated by Bioruptor UCD-200 (Diagenode, Liege, y a Belgium) to result in DNA fragments of 200 to 1000 bp in length

and cellular debris was removed by centrifugation (full speed, 10 min, 4 °C). Sonicated chromatin was divided in 100 µl aliquots and suspended in 1 ml of ChIP dilution buffer (0.01 % SDS, 1.1 % Triton X-100, 1.2 mM EDTA, 167 mM NaCl, 16.7 mM Tris-HCl, pH 8.1, protease inhibitors). 2.5 ml BSA (100 mg/ml) was added to each tube. 100 µl chromatin sample was removed as input DNA. 100 µl of antibody-bound beads were added to the chromatin samples and the samples were incubated o/n at + 4 °C on a rocking platform. The next day, the beads were washed five times with LiCl wash buffer (100 mM Tris pH 7.5, 500 mM LiCl, 1 % IGEPAL, 1 % Sodium deoxycholate) and twice with TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) and eluted with 200 µl of elution buffer (1 % SDS, 0.1 M NaHCO3). All samples were treated with Proteinase-K (10 mg/ml, Thermo Scientific) and DNA was purified with MinElute PCR Purification Kit (Qiagen). Specific primers are listed in Table S6.

Statistical analysis

All experiments were carried out with at least three independent replicates and One-way ANOVA (Tukey's test) was used unless otherwise stated. Differences were considered statistically significant with a p-value < 0.05. Graphpad Prism version 5.0 was used for the analyses. Spearman's correlation and correlation test of significance was computed in R to find associations between Nrf2 activity and other features. The p-values were adjusted using the Benjamini Hochberg FDR for each data type tested, using the p.adjust function and BH method in R. For survival analysis the likelihood ratio test was used to compute the p-values.

Code availability

Code to reproduce key parts of the analysis can be obtained upon request from the authors.

References

1   Hayes JD, Dinkova-Kostova AT. The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. *Trends Biochem Sci* 2014; **39**: 199–218.

2   Leinonen HM, Kansanen E, Pölönen P, Heinäniemi M, Levonen A-L. Role of the Keap1-Nrf2 pathway in cancer. *Adv Cancer Res* 2014; **122**: 281–320.

3   Lau A, Wang X-J, Zhao F, Villeneuve NF, Wu T, Jiang T *et al.* A noncanonical mechanism of Nrf2 activation by autophagy deficiency: direct interaction between Keap1 and p62. *Mol Cell Biol* 2010; **30**: 3275–3285.

4   Komatsu M, Kurokawa H, Waguri S, Taguchi K, Kobayashi A, Ichimura Y *et al.* The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1. *Nat Cell Biol* 2010; **12**: 213–23.

5   Padmanabhan B, Tong KI, Ohta T, Nakamura Y, Scharlock M, Ohtsuji M *et al.* Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol Cell* 2006; **21**: 689–700.

6   Singh A, Misra V, Thimmulappa RK, Lee H, Ames S, Hoque MO *et al.* Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer. *PLoS Med* 2006; **3**: e420.

7   Lister A, Nedjadi T, Kitteringham NR, Campbell F, Costello E, Lloyd B *et al.* Nrf2 is overexpressed in pancreatic cancer: implications for cell proliferation and therapy. *Mol Cancer* 2011; **10**: 37.

8   Jin YN, Ran KH, Ri KY, Hyeok AC, Hyung LS, Yoo NJ *et al.* Somatic mutations of the KEAP1 gene in common solid cancers. *Histopathology* 2012; **60**: 943–952.

9   Shibata T, Ohta T, Tong KI, Kokubu A, Odogawa R, Tsuta K *et al.* Cancer related mutations in NRF2 impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc Natl Acad Sci U S A* 2008; **105**: 13568–13573.

10   Solis LM, Behrens C, Dong W, Suraokar M, Ozburn NC, Moran CA *et al.* Nrf2 and Keap1 abnormalities in non-small cell lung carcinoma and association with clinicopathologic features. *Clin Cancer Res* 2010; **16**: 3743–3753.

11   Goldstein LD, Lee J, Gnad F, Klijn C, Schaub A, Reeder J *et al.* Recurrent Loss of NFE2L2 Exon 2 Is a Mechanism for Nrf2 Pathway Activation in Human Cancers. *Cell Rep* 2016; **16**: 2605–2617.

12   Johnson DR, O'Neill BP. Glioblastoma survival in the United States before and during the temozolomide era. *J Neurooncol* 2012; **107**: 359–364.

13   Omuro A, DeAngelis LM. Glioblastoma and other malignant gliomas: a clinical review. *JAMA* 2013; **310**: 1842–1850.

14   Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 2010; **17**: 98–110.

15   Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP *et al.* Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer Cell* 2010; **17**: 510–522.

16   Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY *et al.* The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 2009; **463**: 318.

17   Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin A a, Kim S *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012; **483**: 603–7.

18   Brennan CWW, Verhaak RGWGW, McKenna A, Campos B, Noushmehr H, Salama SRR *et al.* The somatic genomic landscape of glioblastoma. *Cell* 2013; **155**: 462–77.

19   Xie Y, Bergström T, Jiang Y, Johansson P, Marinescu VD, Lindberg N *et al.* The Human Glioblastoma Cell Culture Resource: Validated Cell Models Representing All Molecular Subtypes. *EBioMedicine* 2015; **2**: 1351–1363.

20   Kansanen E, Jyrkkänen H-K, Levonen A-L. Activation of stress signaling pathways by electrophilic oxidized and nitrated lipids. *Free Radic Biol Med* 2012; **52**: 973–82.

21   Chang JT, Gatza ML, Lucas JE, Barry WT, Vaughn P, Nevins JR. SIGNATURE: a workbench for gene expression signature analysis. *BMC Bioinformatics* 2011; **12**: 443.

22   Ooi A, Dykema K, Ansari A, Petillo D, Snider J, Kahnoski R *et al.* CUL3 and NRF2 mutations confer an NRF2 activation phenotype in a sporadic form of papillary renal cell carcinoma. *Cancer Res* 2013; **73**: 2044–2051.

23   Kim YR, Oh JE, Kim MS, Kang MR, Park SW, Han JY *et al.* Oncogenic NRF2 mutations in squamous cell carcinomas of oesophagus and skin. *J Pathol* 2010; **220**: 446–451.

24   Inami Y, Waguri S, Sakamoto A, Kouno T, Nakada K, Hino O *et al.* Persistent activation of Nrf2 through p62 in hepatocellular carcinoma cells. *J Cell Biol* 2011; **193**: 275–284.

25   Leinonen HM, Ruotsalainen A-K, Määttä A-M, Laitinen HM, Kuosmanen SM, Kansanen E *et al.* Oxidative stress-regulated lentiviral TK/GCV gene therapy for lung cancer treatment. *Cancer Res* 2012; **72**: 6227–35.

26   Tanida I, Ueno T, Kominami E. LC3 conjugation system in mammalian autophagy. *Int J Biochem Cell Biol* 2004; **36**: 2503–2518.

27   Yamamoto A, Tagawa Y, Yoshimori T, Moriyama Y, Masaki R, Tashiro Y. Bafilomycin A1 prevents maturation of autophagic vacuoles by inhibiting fusion between autophagosomes and lysosomes in rat hepatoma cell line, H-4-II-E cells. *Cell Struct Funct* 1998; **23**: 33–42.

28   Johansen T, Lamark T. Selective autophagy mediated by autophagic adapter proteins. *Autophagy* 2011; **7**: 279–296.

29   Ichimura Y, Waguri S, Sou Y-S, Kageyama S, Hasegawa J, Ishimura R *et al.* Phosphorylation of p62 activates the Keap1-Nrf2 pathway during selective autophagy. *Mol Cell* 2013; **51**: 618–31.

30   Jain A, Lamark T, Sjøttem E, Larsen KB, Awuh JA, Øvervatn A *et al.* p62/SQSTM1 is a target gene for transcription factor NRF2 and creates a positive feedback loop by inducing antioxidant response element-driven gene transcription. *J Biol Chem* 2010; **285**: 22576–22591.

31   Bouvy-Liivrand M, De Sande AH, Pölönen P, Mehtonen J, Vuorenmaa T, Niskanen H *et al.* Analysis of primary microRNA loci from nascent transcriptomes reveals regulatory domains governed by chromatin architecture. *Nucleic Acids Res* 2017; **45**: 9837–9849.

32   Pajares M, Jiménez-Moreno N, García-Yagüe ÁJ, Escoll M, de Ceballos ML, Van Leuven F *et al.* Transcription factor NFE2L2/NRF2 is a regulator of macroautophagy genes. *Autophagy* 2016; **12**: 1902–1916.

33   Shibata T, Saito S, Kokubu A, Suzuki T, Yamamoto M, Hirohashi S. Global downstream pathway analysis reveals a dependence of oncogenic NF-E2-related factor 2 mutation on the mTOR growth signaling pathway. *Cancer Res* 2010; **70**: 9095–105.

34   Kanamori M, Higa T, Sonoda Y, Murakami S, Dodo M, Kitamura H *et al.* Activation of the NRF2 pathway and its impact on the prognosis of anaplastic glioma patients. *Neuro Oncol* 2014; **1**: 1–11.

35   Galavotti S, Bartesaghi S, Faccenda D, Shaked-Rabi M, Sanzone S, McEvoy A *et al.* The

autophagy-associated factors DRAM1 and p62 regulate cell migration and invasion in glioblastoma stem cells. *Oncogene* 2013; **32**: 699–712.

36   Singer E, Judkins J, Salomonis N, Matlaf L, Soteropoulos P, McAllister S *et al.* Reactive oxygen species-mediated therapeutic response and resistance in glioblastoma. *Cell Death &Amp; Dis* 2015; **6**: e1601.

37   Agnihotri S, Zadeh G. Metabolic reprogramming in glioblastoma: the influence of cancer metabolism on epigenetics and unanswered questions. *Neuro Oncol* 2016; **18**: 160–172.

38   Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 2009; **324**: 1029–33.

39   Mitsuishi Y, Motohashi H, Yamamoto M. The Keap1-Nrf2 system in cancers: stress response and anabolic metabolism. *Front Oncol* 2012; **2**: 200.

40   Levine AJ, Puzio-Kuter AM. The Control of the Metabolic Switch in Cancers by Oncogenes and Tumor Suppressor Genes. *Science (80-)* 2010; **330**: 1340 LP – 1344.

41   Jin L, Li D, Alesi GN, Fan J, Kang H-B, Lu Z *et al.* Glutamate Dehydrogenase 1 Signals through Antioxidant Glutathione Peroxidase 1 to Regulate Redox Homeostasis and Tumor Growth. *Cancer Cell* 2015; **27**: 257–270.

42   Rinaldi M, Caffo M, Minutoli L, Marini H, Abbritti RV, Squadrito F *et al.* ROS and Brain Gliomas: An Overview of Potential and Innovative Therapeutic Strategies. *Int J Mol Sci* 2016; **17**: 984.

43   Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* 2018; **173**: 321-337.e10.

44   Collisson E a., Campbell JD, Brooks AN, Berger AH, Lee W, Chmielecki J *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014. doi:10.1038/nature13385.

45   Kansanen E, Jyrkkänen H-K, Volger OL, Leinonen H, Kivelä AM, Häkkinen S-K *et al.* Nrf2-dependent and -independent responses to nitro-fatty acids in human endothelial cells: identification of heat shock response as the major pathway activated by nitro-oleic acid. *J Biol Chem* 2009; **284**: 33233–41.

46   Bowman RL, Wang Q, Carro A, Verhaak RGW, Squatrito M. GlioVis data portal for visualization and analysis of brain tumor expression datasets. *Neuro Oncol* 2017; **19**: 139–141.

47   Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* 2013; **14**: 7.

48   Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011; **27**: 1739–40.

49   Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 2005; **102**: 15545 LP – 15550.

50   Koponen JK, Kekarainen T, E Heinonen S, Laitinen A, Nystedt J, Laine J *et al.* Umbilical cord blood-derived progenitor cells enhance muscle regeneration in mouse hindlimb ischemia model. *Mol Ther* 2007; **15**: 2172–2177.

51   Kansanen E, Bonacci G, Schopfer FJ, Kuosmanen SM, Tong KI, Leinonen H *et al.* Electrophilic nitro-fatty acids activate NRF2 by a KEAP1 cysteine 151-independent mechanism. *J Biol Chem* 2011; **286**: 14019–27.

52   Jyrkkänen H-K, Kuosmanen S, Heinäniemi M, Laitinen H, Kansanen E, Mella-Aho E *et al.* Novel insights into the regulation of antioxidant-response-element-mediated gene expression by electrophiles: induction of the transcriptional repressor BACH1 by Nrf2. *Biochem J* 2011; **440**: 167–74.

53   Rokavec M, Kaller M, Horst D, Hermeking H. Pan-cancer EMT-signature identifies RBM47 down-regulation during colorectal cancer progression. *Sci Rep* 2017; **7**: 1–15.

Figure legends

Figure 1. Prediction of Nrf2 overactivity using SIGNATURE and GSVA tools. A, heatmap of Nrf2 target gene signature expression; warm colors (red and yellow) signify high expression and cool colors (shades of blue), low expression. The signature model training sets consist of microarray samples with inactive and active Nrf2 status. The expression profile of 80 upregulated and 20 downregulated genes relative to inactive samples are shown in rows. B, Cancer cell lines available from the CCLE dataset are shown ranked based on the probability of an active Nrf2 target gene signature. The plot shows individual samples on the X-axis and the probabilities and their confidence intervals on the Y-axis. Also GSVA score for independent Nrf2 target gene signature is shown beside the SIGNATURE tool result. Cell lines were sorted as in the SIGNATURE plot and the corresponding GSVA scores are shown as a heatmap. C, definition of 5 Nrf2 activity categories based on probability

of overactive Nrf2. D, Nrf2 hyperactive cancer types and percentage within tumor type. From the total of 915 samples, 66 (representing about 7 %) belong to the highest activity score (P>95 %). E, percentages of predicted Nrf2 overactive samples categories for Rembrandt glioma and TCGA GBM patient samples, separated by WHO grade are shown as a barplot.

Figure 2. Validation of Nrf2 hyperactivity in glioma. A-B, Nrf2 target gene expression in TCGA patient samples and CCLE glioma cell lines. Samples are sorted according to Nrf2 activity from high activity (left) to low activity (right). Nrf2 target genes are very highly expressed in Nrf2 active H4, T98G, GAMG and SNU-201 cell lines (Set1), highly expressed in set2 and the expression is low in Nrf2 inactive cell lines (P<80 %). *p < 0.05, **p < 0.01, *** p < 0.001, ****p < 0.0001 (Spearman´s correlation). C, ARE activity measured with lentiviral ARE-luciferase reporter assay. Luciferase activity was normalized to proteins and is depicted relative to NHA cells. n=4-8 mean ± SEM, * p < 0.05, **p < 0.01, ***p < 0.001 (One-way ANOVA, Tukey's test). D, Expression of validated Nrf2 target genes GCLM, NQO1, and HMOX1 assessed with Western blotting, using β-actin for normalization. E, ChIP was performed on untreated T98G, GAMG, and DKMG cells using anti-Nrf2 and normal rabbit IgG antibody for measuring specific and unspecific binding, respectively. RT-q-PCR was performed using primers specific for an ARE on NQO1 and HMOX1 gene promoters. IgG control values were subtracted from the output values and the binding is depicted relative to respective input values. Results are depicted as mean ± SEM. (n=9). ***p<0.001, (One-way ANOVA, Tukey's test).

Figure 3. Expression imbalance of *NFE2L2* and *KEAP1* is correlated to Nrf2 activity. A, NFE2L2 and Keap1 gene expression imbalance in Nrf2 active compared to Nrf2 inactive samples from TCGA. TCGA samples with low Nrf2 activity correlate with copy number gains in *KEAP1*. Copy number gain and loss have per gene value of 0.3 to 0.7 or -0.3 to -0.7 respectively. Red star for copy number variation, black for gene expression, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001 (Spearman's correlation). B, Heatmap of *NFE2L2* and *KEAP1* expression in CCLE glioma cell lines reveal amplification in *NFE2L2* T98G and deletion in H4 cells. C-D, Validation of *NFE2L2* and *KEAP1* gene expression in CCLE cell lines. Relative expression of *NFE2L2* and *KEAP1* measured with

qPCR. n=3, mean ± SEM, * p < 0.05, ** p < 0.01, ***p < 0.001 (One-way ANOVA, Tukey's test). E-F, Nrf2, Keap1 and β-actin protein levels of glioma cell lines assessed with Western blotting. DKMG and NHA were used as controls of normal Nrf2 expression, and A549 lung cancer cells with known Nrf2 hyperactivity were used as positive control.

Figure 4. Autophagy dysregulation and p62 protein-protein interactions with Keap1 is linked to Nrf2 overactivation. A, composite pathway analysis in TCGA, CCLE and HGCC links Nrf2 activity with several tumorigenic processes. Ranking from highest to lowest enriched processes was computed as mean normalized enrichment score. Each process consists of several individual pathways, as shown in Table S2. B, Autophagy flux measurement of LC3B and β-actin examined with Western blotting. Flux is impaired in GAMG cell line. C, TEM analysis of H4, T98G, GAMG, LN18 and NHA cells. Autophagosomal structures (black arrows), autolysosomal structures (white arrows) and autophagosomal-like structures (grey arrows) are indicated in the picture.

Figure 5. p62 is overexpressed in samples with high Nrf2 activity and binds to Keap1 in glioma cell lines. A, p62 and β-actin assessed with Western blotting. B, Relative expression of *SQSTM1* measured with qPCR. Data is depicted relative to NHA control cells. Mean ± SEM (n=3). *p < 0.05, **p < 0.01, ***p < 0.001. C, D, *SQSTM1* mRNA (C) and p62 protein (D) are overexpressed in Nrf2 active TCGA samples. E, Correlation coefficient between *SQSTM1* and p62 expression to Nrf2 activity is shown as a heatmap across diseases using TCGA data. ***p < 0.001 and Rho>0.25 for both *SQSTM1* and p62 (Spearman's correlation). F, Heatmap of *SQSTM1* gene expression and gene copy number aberrations in CCLE cell lines. Copy number gain and loss have per gene value of 0.3 to 0.7 or -0.3 to -0.7 respectively. Amplifications and deletions have values of over 0.7 or below -0.7. Red star for copy number variation, black for gene expression, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001 (Spearman´s correlation). G, T98G, GAMG and A172 cell lysates were immunoprecipitated with p62 or Keap1 antibody, and the amount of bound Keap1 or p62 was detected with Western blotting. Total amount of p62 and Keap1 was also analyzed (input). H, p62,

phospho-p62 (Ser-351) phospho-p62 (Ser-403) and β-actin Western blotting of glioma cell lines. I, immunofluorescence staining of p62, Keap1 and nucleus (DAPI) in T98G cells.

Figure 6. Nrf2 and p62 jointly contribute to mesenchymal transition in mesenchymal subtype of glioma. A, GBM subtype classification is shown for each patient sample in TCGA. Also GSVA scores are shown for mesenchymal and classical subtype gene sets. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$ (Spearman´s correlation). B. Correlation coefficient between EMT gene sets and Nrf2 activity is shown as a heatmap across diseases using TCGA data. Significance of correlation is shown as stars for EMT Hallmarks MsigDB gene set, as in A. As a confirmation, mesenchymal signature gene set from Rokavec et. al.[53] is shown. C, Western blot analysis of mesenchymal markers beta-catenin, snail, slug and vimentin in T98G and GAMG cells with Nrf2 and p62 siRNA knockdown. Beta-actin and Lamin-B1 were used as the loading control. Quantitative analysis of the western blot showed that both Nrf2 and p62 siRNA knockdown have significantly reduced the protein expression of mesenchymal markers. n=4, mean ± S.E.M. $*p < 0.05$, $**p < 0.01$, (One-way ANOVA Tukey's test). D, ChIP was performed on untreated T98G, GAMG and DKMG cells using anti-Nrf2 and normal rabbit IgG antibody for measuring specific and unspecific binding, respectively. RT-q-PCR was performed using primers specific for an ARE on SNAI2 and CTNNB1 gene enhancers. IgG control values were subtracted from the output values and the binding is depicted relative to respective input values. n=9, mean ± S.E.M, $*p < 0.05$, $***p < 0.001$ (One-way ANOVA, Tukey's test).

Figure 7. Nrf2 and p62 jointly contribute to proliferation, disease progression and invasion. A, Kaplan-Meyer plot indicate that samples with highest Nrf2 activity significantly increase the risk of disease progression. n=22, p-value 0.0022 (Likelihood ratio test). B, Cell proliferation is decreased with Nrf2 and p62 siRNA knockdown. GAMG and T98G cell proliferation after Nrf2 and p62 knockdown with siRNAs was assessed from 24-96 h growth curves. Quantification of results showed that cell proliferation in both the cell lines was significantly decreased. n=4, Mean ± S.E.M, $*p < 0.05$ (One-way ANOVA, Tukey's test). C, T98G cell invasion in type I collagen-cultrex matrix with Nrf2

and p62 siRNAs and quantification of results showed that Nrf2 and p62 siRNA knockdown significantly inhibits cell invasion in a 3D matrix environment. Scale bar = 100 µm; blue = nuclei, red = actin staining. n=3, mean ± S.E.M, *p < 0.05, ***p < 0.001 (One-way ANOVA, Tukey's test). D, Anchorage-independent growth of T98G cells was inhibited with Nrf2 and p62 siRNA knockdown. T98G (2500 cells/well) were grown in 96-well ultra-low attachment plates in serum-free growth medium and the diameter of 3D growth spheres was measured at day 7. Results showed that Nrf2 and p62 siRNA knockdown significantly decreased the anchorage-independent growth of T98G cell line. n=3, scale bar = 100 µm. Quantification of growth area of 3-dimensional spheres in T98G cells with Nrf2 and p62 siRNA knockdown. n=5, mean ± S.E.M, *p < 0.05, **p < 0.01, (One-way ANOVA Tukey's test). E, schematic illustration of Nrf2 hyperactivity mechanisms and downstream effects in glioma. Nrf2 and p62 form a positive feedback loop in glioma, where p62 binds to Keap1 and promotes Keap1 degradation. p62 mediated degradation of Keap1 is further enhanced by Ser-Thr mediated phosphorylation of p62 and autophagy pathway. Nrf2 and p62 jointly contribute to mesenchymal transition, proliferation and invasion in glioma.
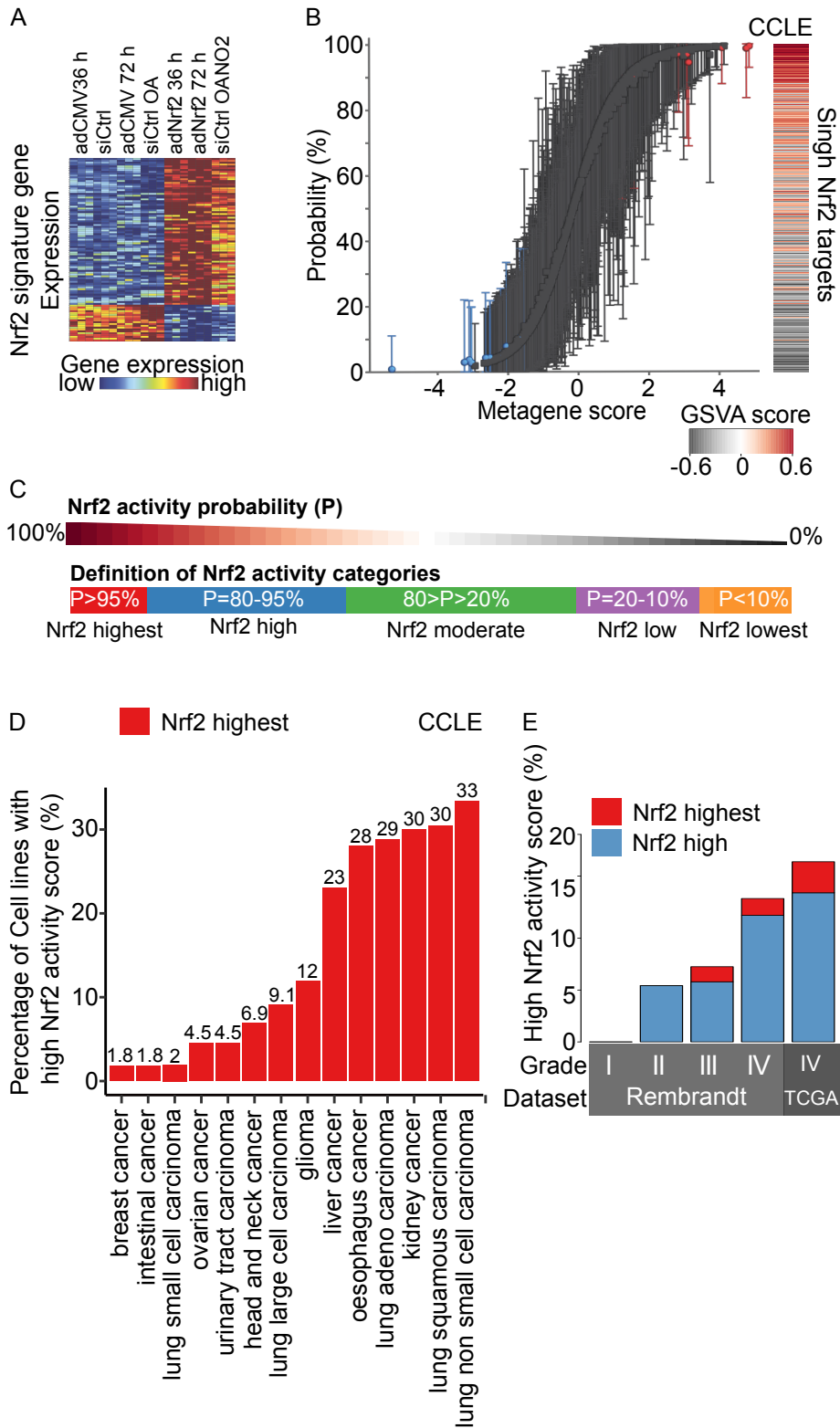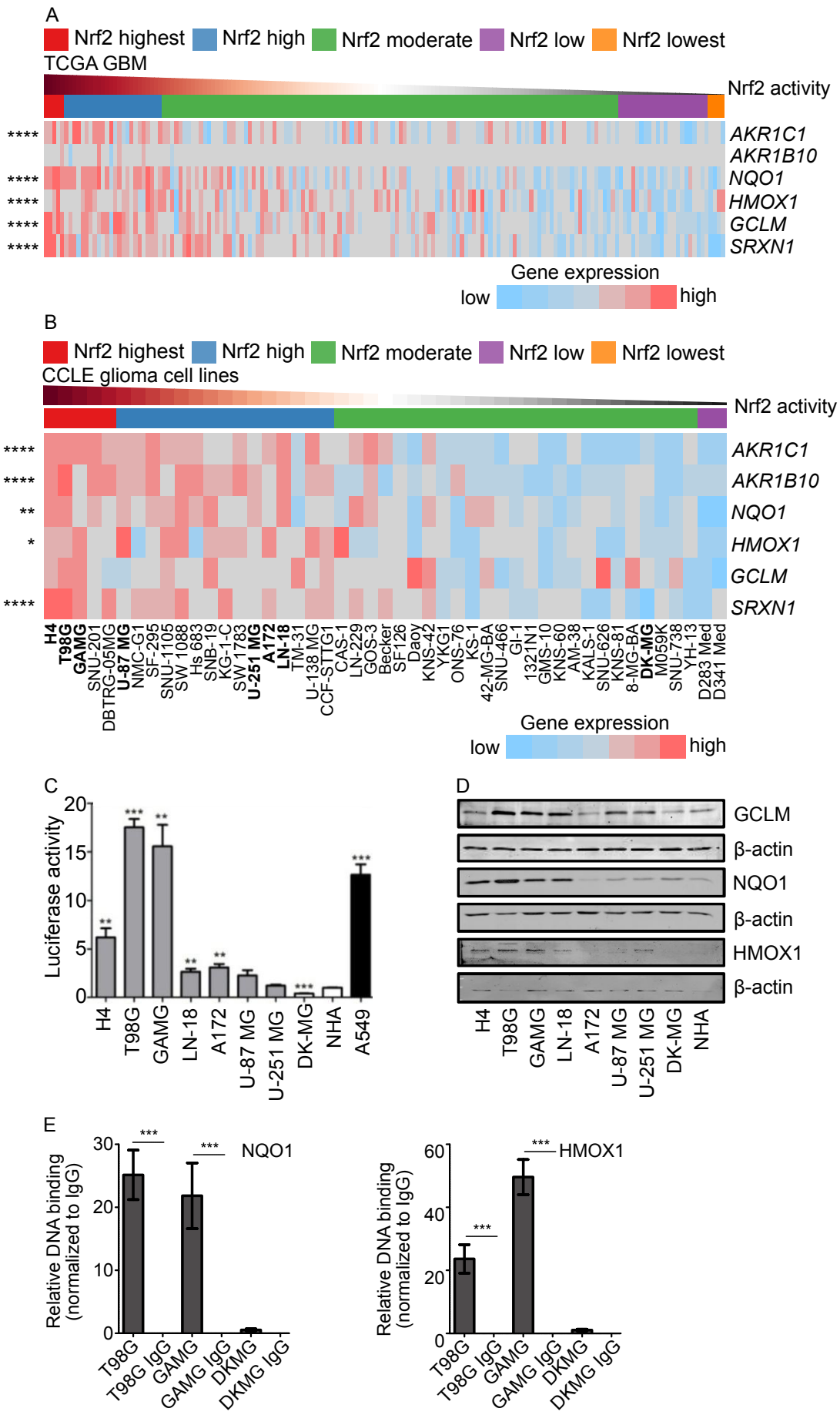
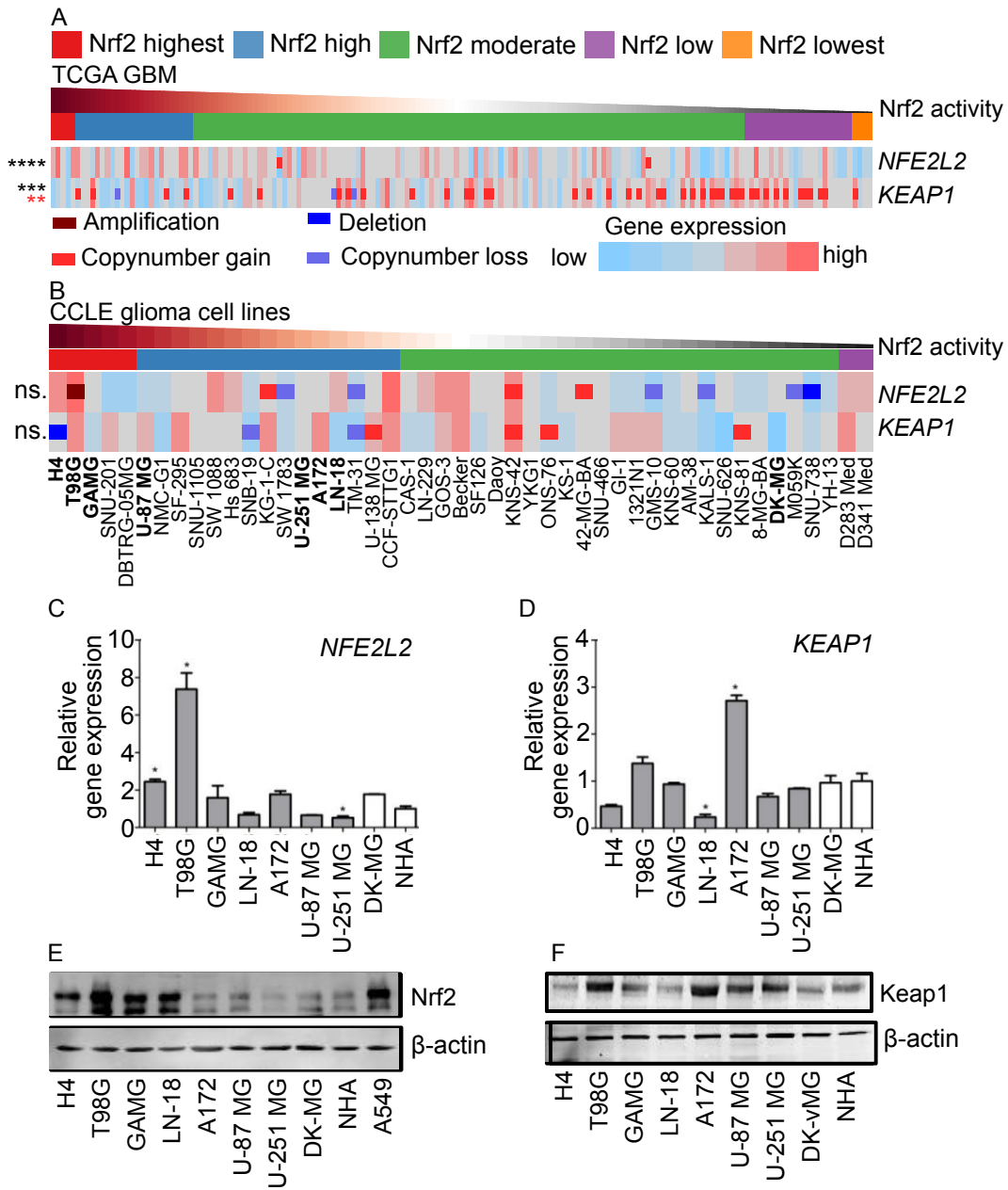Figure 1. Pölönen et al.

Figure 2. Pölönen et al.

Figure 3. Pölönen et al.

A

Joined Gene Set Analysis of GBM Patient and Cell lines link several processes with Nrf2 activity

Process	# Significant Pathways*	Rank
MESENCHYMAL TRANSITION	5	1
CELL CYCLE	4	2
PROTEASOME, AUTOPHAGY, LYSOSOME	22	3
ENERGY METABOLISM	12	4
REDOX PATHWAY	5	5
INFLAMMATION, NFKB, IL-SIGNALLING	28	6
OTHER	37	7
CELL DAMAGE, P53	13	8
MITOCHONDRIA	4	9
GOLGI	2	10
INVASION	4	11
METABOLISM	15	12
PURINE METABOLISM	5	13
CELL SIGNALLING	19	14
MATRIX PROTEIN PROCESSING	4	15
POST-TRANSLATIONAL MODIFICATION	4	16
OTHER PATHWAY	6	17

* FDR q-value < 0.05 in CCLE GBM, TCGA GBM and HGCC GBM

B

LC3I
LC3II
β-actin
BafA

H4   T98G   GAMG   LN-18   A172   U-87 MG   U-251 MG   DK-MG   NHA

C

H4

T98G

GAMG

LN-18

NHA

1µM

→ (white) Autolysosomal structures
→ (black) Autophagosomal structures
→ (grey) Autophagosomal-like/ lysosomal/endosomal structures

Figure 4. Pölönen et al.

A

P62

β-actin

H4 T98G GAMG LN-18 A172 U-87 MG U-251 MG DK-MG NHA

B

Relative gene expression

SQSTM1

H4 T98G GAMG LN-18 A172 U-87 MG U-251 MG DK-MG NHA

C

TCGA GBM

SQSTM1

Gene expression

Nrf2 highest Nrf2 high Nrf2 moderate Nrf2 low Nrf2 lowest

D

TCGA GBM

P62

Protein expression

Nrf2 highest Nrf2 high Nrf2 moderate Nrf2 low Nrf2 lowest

E

SQSTM1/p62 and Nrf2 GSVA score correlation (pan-cancer)

Pearson Rho

KIRP GBM UCS LUSC LIHC MESO SARC BLCA OV BRCA TGCT UVM SKCM DLBC UCEC CHOL LUAD KIRC ACC LGG CESC ESCA STAD PCPG PAAD THCA HNSC COAD PRAD THYM KICH READ

RNA
Protein

F

■ Nrf2 highest ■ Nrf2 high ■ Nrf2 moderate ■ Nrf2 low ■ Nrf2 lowest

CCLE glioma cell lines

Nrf2 activity

SQSTM1

H4 T98G GAMG SNU-201 DBTRG-05MG **U-87 MG** NMC-G1 SF-295 SNU-1105 SW 1088 Hs 683 SNB-19 KG-1-C SW 1783 **U-251 MG A172 LN-18** TM-31 U-138 MG CCF-STTG1 CAS-1 LN-229 GOS-3 Becker SF126 Daoy KNS-42 YKG1 ONS-76 KS-1 42-MG-BA SNU-466 GI-1 1321N1 GMS-10 KNS-60 AM-38 KALS-1 SNU-626 KNS-81 8-MG-BA **DK-MG** M059K SNU-738 YH-13 D283 Med D341 Med

■ Amplification   ■ Deletion
■ Copynumber gain   ■ Copynumber loss

Gene expression

low ▬▬▬▬ high

G

p62 { Keap1 / p62
Keap1 { Keap1 / p62
Input { Keap1 / p62

IP T98G GAMG A172

H

P-Ser351
P-Ser403
p62
β-actin

H4 T98G GAMG LN-18 A172 U-87 MG U-251 MG DK-MG NHA

I

T98G

Keap1 | p62
DAPI | merged

Figure 5. Pölönen et al.

A

Nrf2 highest | Nrf2 high | Nrf2 moderate | Nrf2 low | Nrf2 lowest

TCGA GBM

Nrf2 activity

Subtype

**** GSVA Classical

**** GSVA Mesenchymal

Mesenchymal | Neural | G-CIMP | Classical | Proneural

Subtype Signature

−0.5   0   0.5

B

EMT-gene set and Nrf2 GSVA score correlation (pan-cancer)

Pearson Rho

−0.4   0   0.4

EMT Hallmarks (MsigDB)

Rokavec mesenchymal

UVM LGG GBM SARC KIRP PAAD PCPG MESO DLBC SKCM KICH PRAD ACC LIHC KIRC BLCA CESC THYM BRCA CHOL COAD ESCA READ UCEC LUAD STAD TGCT UCS HNSC LUSC THCA

C

T98G

Cntrl siCntrl siNrf2-1 siNrf2-2 sip62

GAMG

Cntrl siCntrl siNrf2-1 siNrf2-2 sip62

Nrf2
p62
β-catenin
Snail
Slug
Vim
Actin
Lam-B1

T98G

Nrf2 | p62 | Slug | Snail | Bcat | Vim

Fold chang eto control

siCntrl siNrf2-1 siNrf2-2 sip62

GAMG

Nrf2 | p62 | Slug | Snail | Bcat | Vim

Fold chang eto control

siCntrl siNrf2-1 siNrf2-2 sip62

D

SNAI2

Relative DNA binding (normalized to IgG)

T98G | T98G IgG | GAMG | GAMG IgG | DKMG | DKMG IgG

CTNNB1

Relative DNA binding (normalized to IgG)

T98G | T98G IgG | GAMG | GAMG IgG | DKMG | DKMG IgG

Figure 7. Pölönen et al.

Supplementary Materials

Nrf2 and p62 jointly contribute to mesenchymal transition and invasion in glioblastoma

Petri Pölönen, Ashik Jawahar Deen, Hanna M Leinonen, Henna-Kaisa Jyrkkänen, Suvi Kuosmanen, Mimmi Mononen, Ashish Jain, Tomi Tuomainen, Sanna Pasonen-Seppänen, Jaana Hartikainen, Arto Mannermaa, Matti Nykter, Pasi Tavi, Terje Johansen, Merja Heinäniemi, Anna-Liisa Levonen
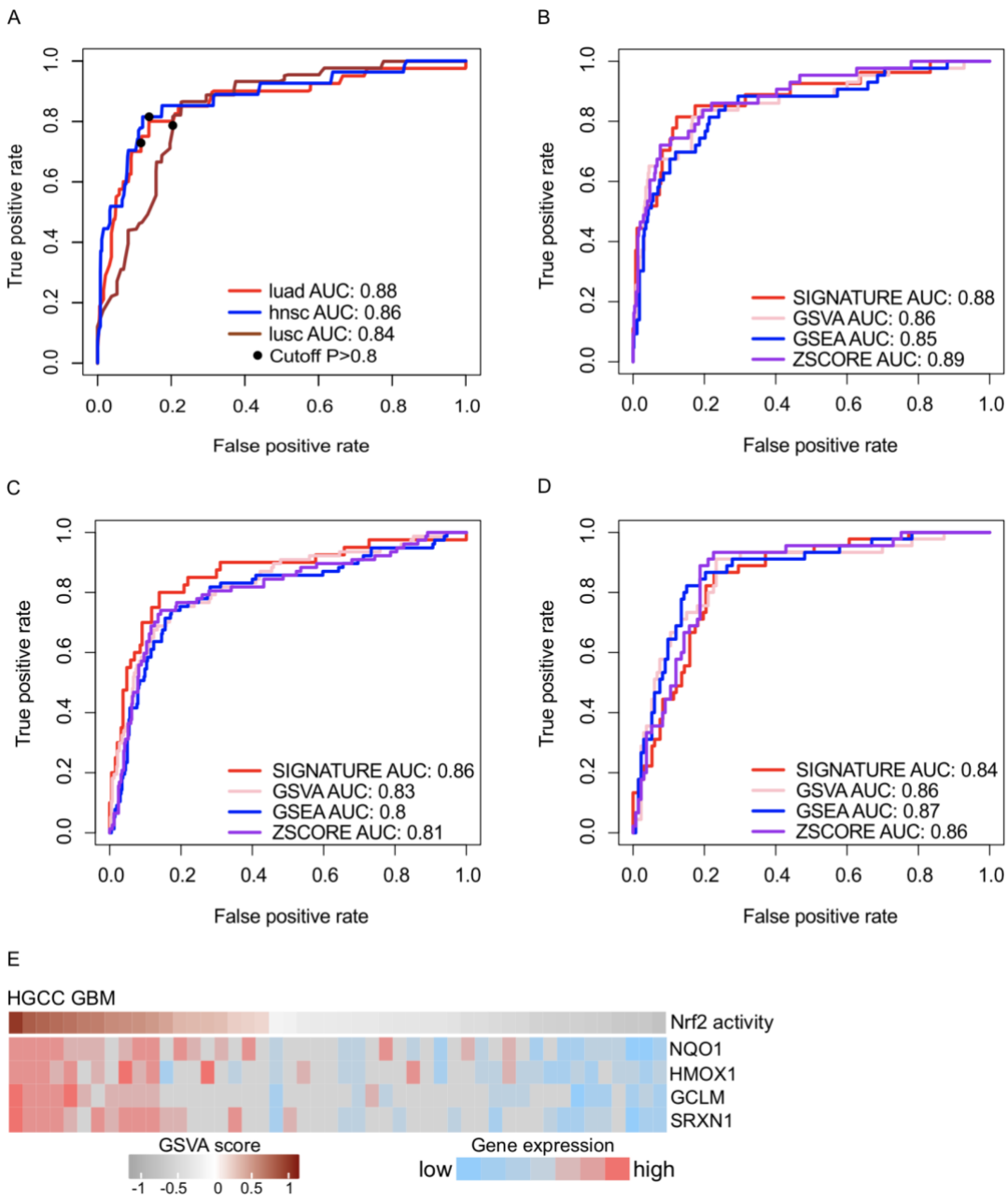
Contents:

Fig. S1. Comparison of Binreg2, GSVA and published Nrf2 signature in activity prediction.
ROC curves established from sequenced samples of lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), head and neck squamous cell carcinoma (HNSC) are shown in A. The missense mutations of *NFE2L2* and *KEAP1* were considered as true positive predictions. Samples from LUAD, HNSC, LUSC Lines are shown in red, blue and brown, respectively. The cutoff of P>80% is indicated by a black dot. The area under the curve (AUC) analysis reveals good classification efficiency of the Binreg2 algorithm. ROC curves for LUAD, LUSC, HNSC with the published gene set (SINGH_NFE2L2_TARGETS) that provided good classification performance for different unsupervised tools (GSVA, single sample GSEA and combined z-score (ZSCORE) are shown in B-D. The NFE2L2.V2 and Nrf2 canonical pathway gene sets had poor performance (data not shown). The ROC curves of binreg2, GSVA, single sample GSEA and combined z-score (ZSCORE) are colored red, pink, blue, purple, respectively. The Binreg2 and GSVA SINGH_NFE2L2_TARGETS gene set methods perform well according to AUC in LUAD and HNSC. E. Nrf2 activity validation in HGCC data sets. Nrf2 target gene expression in HGCC data set. Glioma patient derived cell lines are sorted according to the predicted Nrf2 activity from high activity on left to low activity on right.
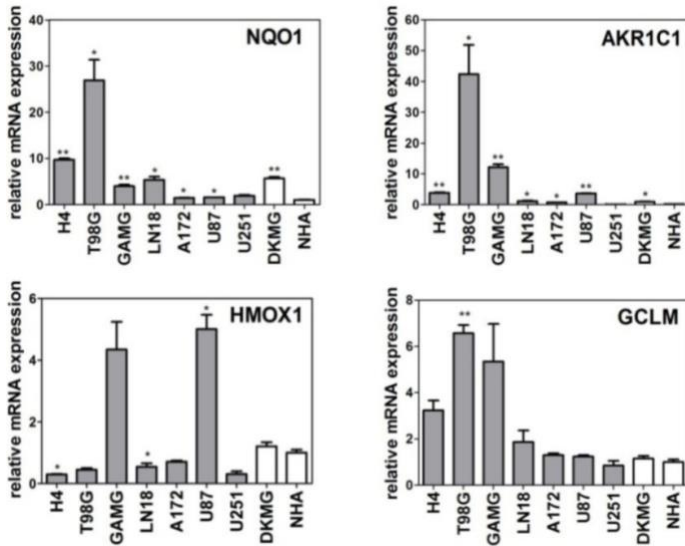
Fig. S2. Relative expression of Nrf2 target genes measured with qPCR. n=3, mean ± SEM.* p<0.05; ** p<0.01; ***p<0.001.
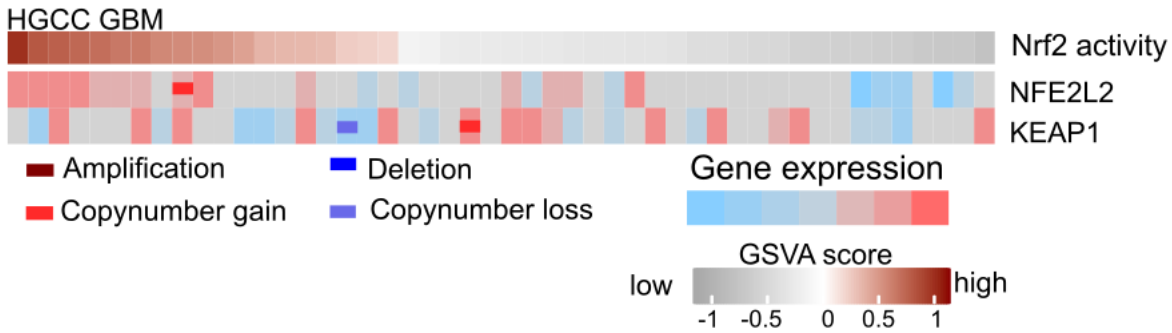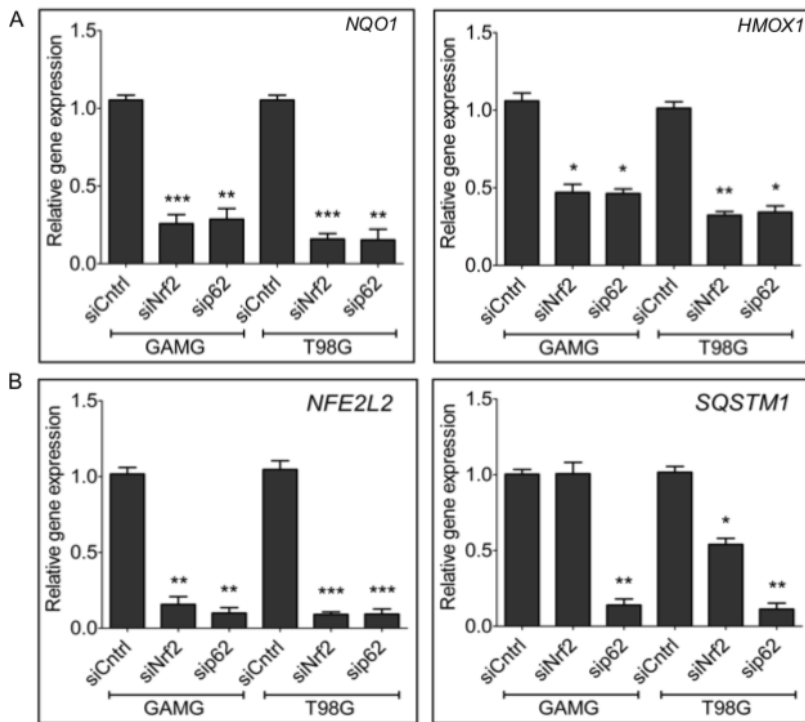

Fig. S3. NFE2L2 and KEAP1 expression and genetic events in HGCC data set. Glioma patient derived cell lines are sorted according to the predicted Nrf2 activity from high activity on left to low activity on right.

Fig. S4. A, GAMG and T98G cells were treated with siRNAs against Nrf2 and p62 for 48 h and the resulting mRNA levels of Nrf2 target genes, NQO1 and HMOX1 were analyzed by qRT-PCR. n = 3, Mean ± S.E.M; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ (One-Way ANOVA, Tukey's test). B, qRT-PCR of NFE2L2 and SQSTM1 in GAMG and T98G cells with Nrf2 and p62 knockdown. n = 3, mean ± S.E.M. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, (One-way ANOVA, Tukey's test).
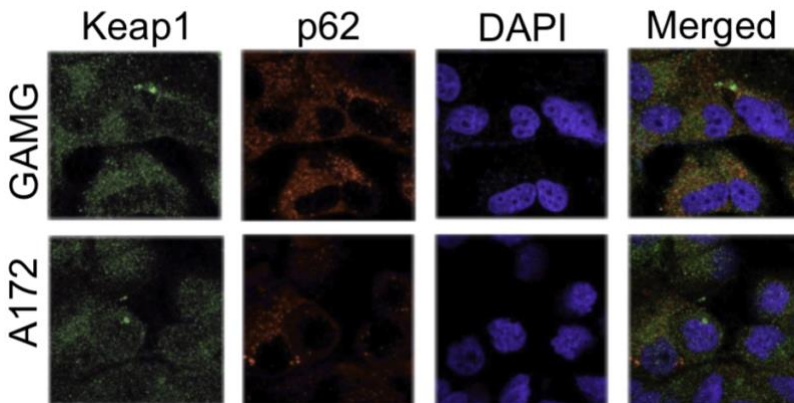

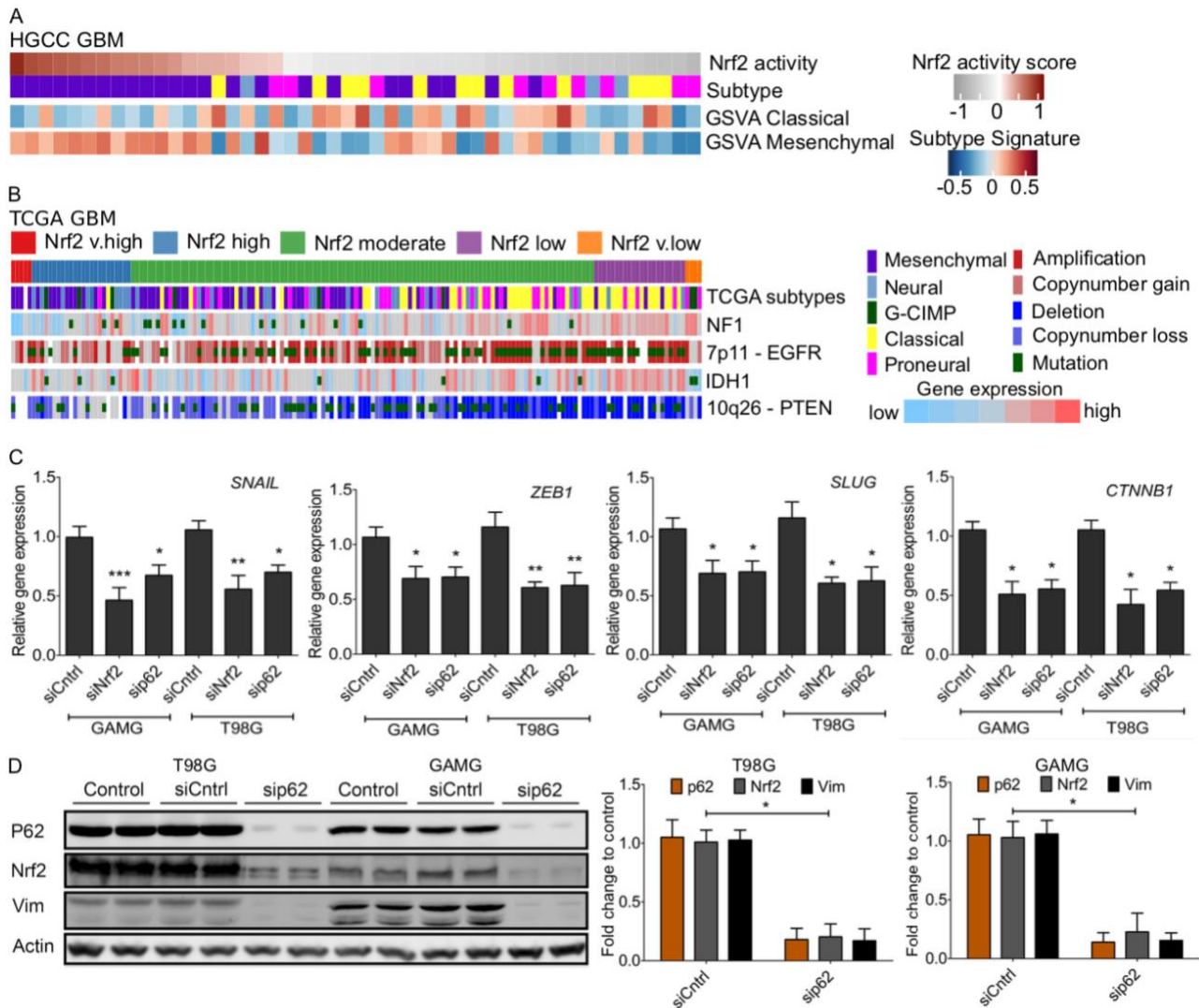Fig. S5. Immunofluorescence staining of p62, Keap1 and nucleus (DAPI) in GAMG and A172 cells.

Fig. S6. A. GBM subtype classification is shown for each patient derived cell line in HGCC data set. B. Nrf2 activity correlation to genetic events enriched in mesenchymal and classical subtypes. Nrf2 activity is correlated to mesenchymal subtype, but not typical mesenchymal subtype NF1 mutations. Nrf2 activity anticorrelates with EGFR and PTEN mutations common in classical subtype. IDH1 mutations are not correlated to Nrf2 activity. C, qRT-PCR of mesenchymal markers SNAIL, ZEB1, SLUG and CTNNB1 in GAMG and T98G cells with Nrf2 and p62 knockdown. D, Western blot analysis of mesenchymal marker vimentin in T98G and GAMG cells with a different p62 siRNA, as in Fig. 6C. Beta-actin was used as the loading control. Quantitative analysis of the western blot showed that p62 siRNA knockdown significantly reduced Nrf2 and vimentin expression. n = 3, mean ± S.E.M, *p < 0.05, **p < 0.01; (One-way ANOVA, Tukey's test).
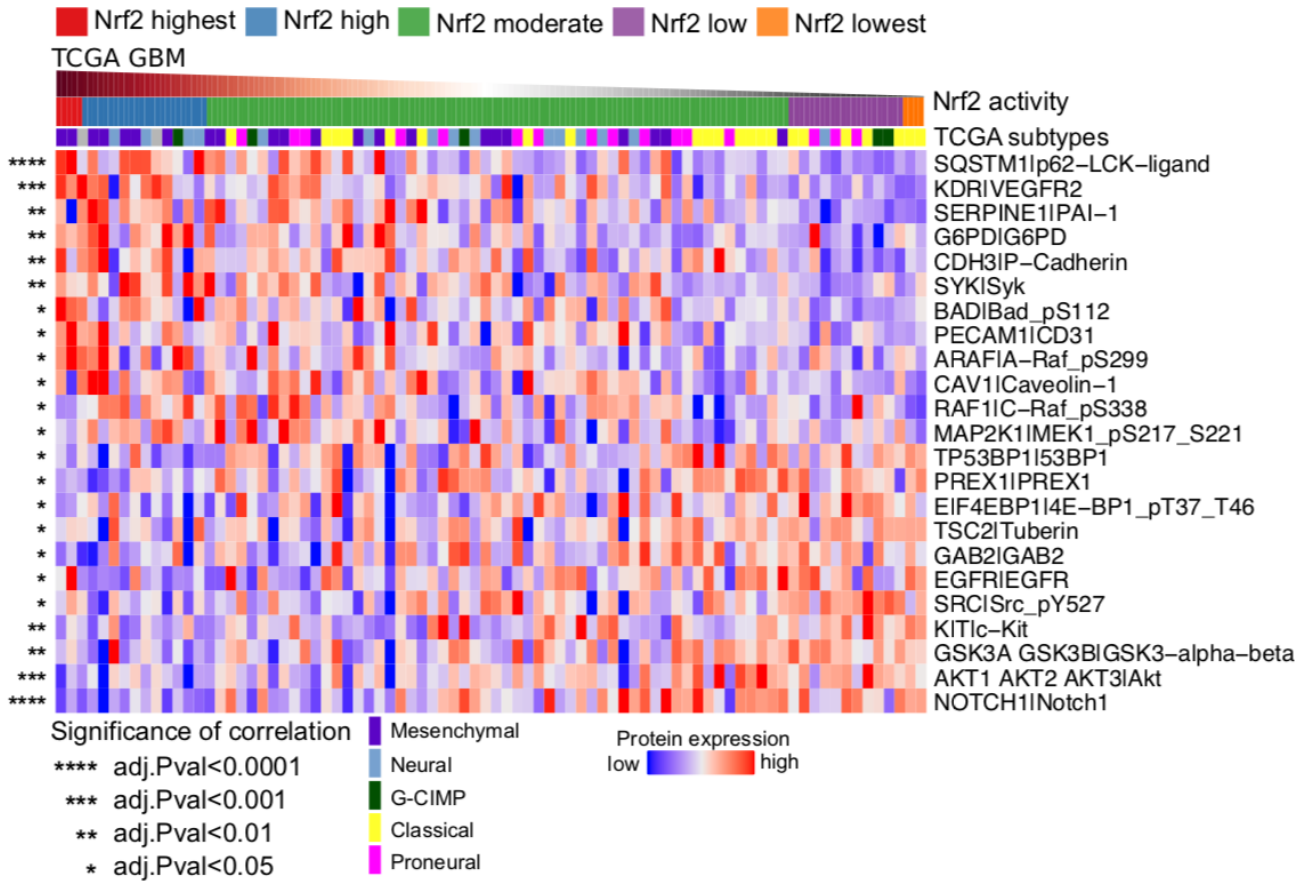
Fig. S7. Nrf2 activity correlation to RPPA protein expression in TCGA GBM data set. Only proteins with significant correlation to Nrf2 activity are shown (FDR < 0.05). *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.001 (Spearman's correlation).
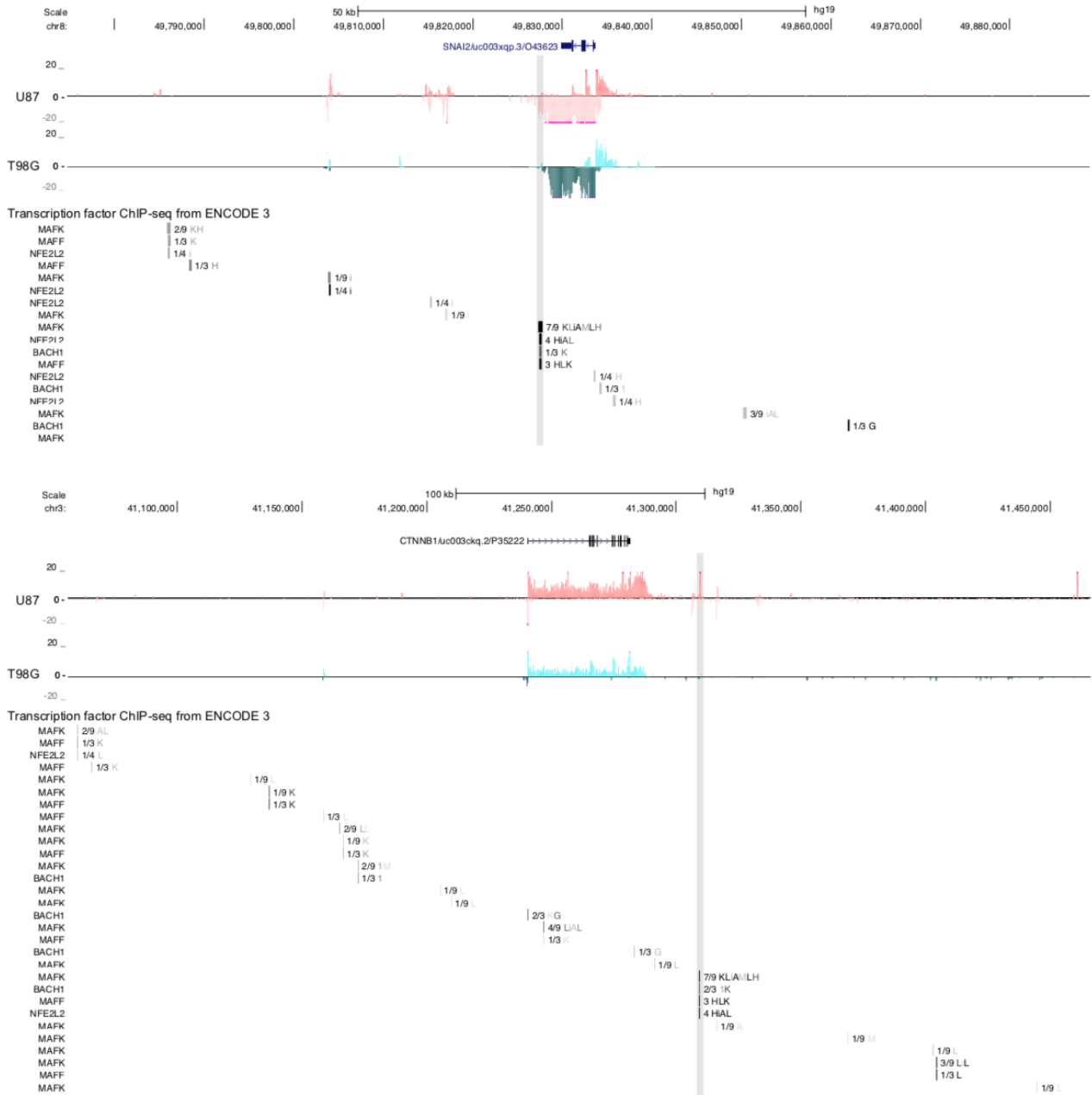
Figure S8. Identification of candidate Nrf2 binding sites on glioma cell line enhancers. GRO-seq nascent RNA expression is shown for U87 and T98G glioma cell lines. Bi-directional signal marks active enhancers and gene promoters. ENCODE 3 transcription factor binding sites for MAFF, MAFK, BACH1 and NFE2L2 are shown in SNAI2 and CTNNB1 locus. Enhancers highlighted were selected for the CHIP analysis in Fig. 6D.

A
TCGA RNA-seq n=166

Cox proportional hazard model, multivariate

	HR	CI .95	p-value	
Nrf2 activity	1.5	1.2-2	0.0019	**
G-CIMP	0.2	0.05-0.75	0.016	*
Neural	0.42	0.19-0.9	0.027	*
Mesenchymal	0.76	0.41-1.4	0.41	
Proneural	0.94	0.5-1.8	0.85	

B
TCGA RNAseq and microarray (n=520)

Cox proportional hazard model, multivariate

	HR	CI .95	p-value	
G-CIMP	0.33	0.19-0.56	0.000056	****
NRF2 Highest	2.5	1.4-4.5	0.0013	**
IDH	0.74	0.58-0.96	0.022	*
Recurrent GBM	0.57	0.31-1	0.062	.
NRF2 Low/Lowest	0.69	0.44-1.1	0.1	
MGMT	0.82	0.64-1.1	0.12	
Secondary GBM	0.46	0.17-1.3	0.14	
Proneural	1.3	0.91-1.8	0.16	
Neural	0.8	0.56-1.1	0.21	
Male	1.1	0.83-1.3	0.64	
Age	1	0.99-1	0.71	
Mesenchymal	0.98	0.73-1.3	0.88	

Likelihood ratio test (complete model) p-value 1e-08

Fig. S9. A. Multivariate cox proportional hazard model analysis of progression free survival using Nrf2 activity probability (numeric feature) and GBM subtypes as predictors. B. Multivariate cox proportional hazard model analysis of progression free survival using Nrf2 activity categories (as in Figure 7A) and prognostic features, age, gender, MGMT methylation status, IDH mutations, recurrence, subtype as predictors.
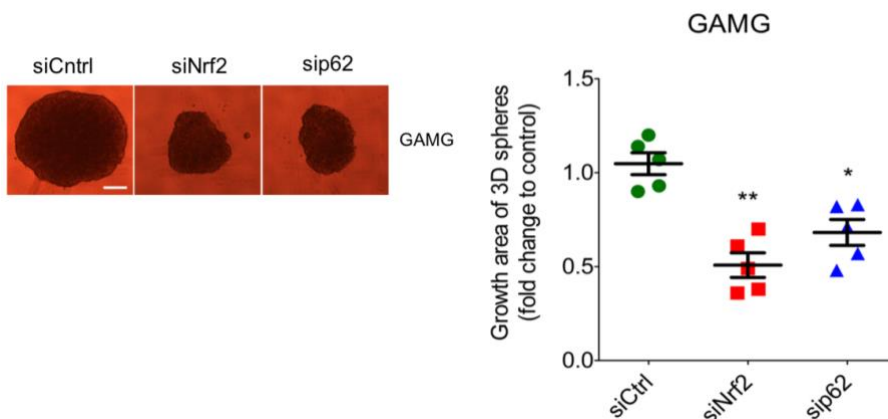

Fig. S10. Anchorage-independent growth of GAMG cells was inhibited with Nrf2 and p62 knockdown. GAMG (1000 cells/well) were grown in 96-well ultra-low attachment plates in serum-free growth medium and the diameter of 3D growth spheres was measured at day 7. Results showed that Nrf2 and p62 knockdown significantly decreased the anchorage-independent growth of GAMG cell line. n = 3, scale bar = 100 µm. Quantification of growth area of 3-dimensional spheres in GAMG cells with Nrf2 and p62 knockdown. n = 5, mean ± S.E.M, *p < 0.05, **p < 0.01, (One-way ANOVA, Tukey's test)).
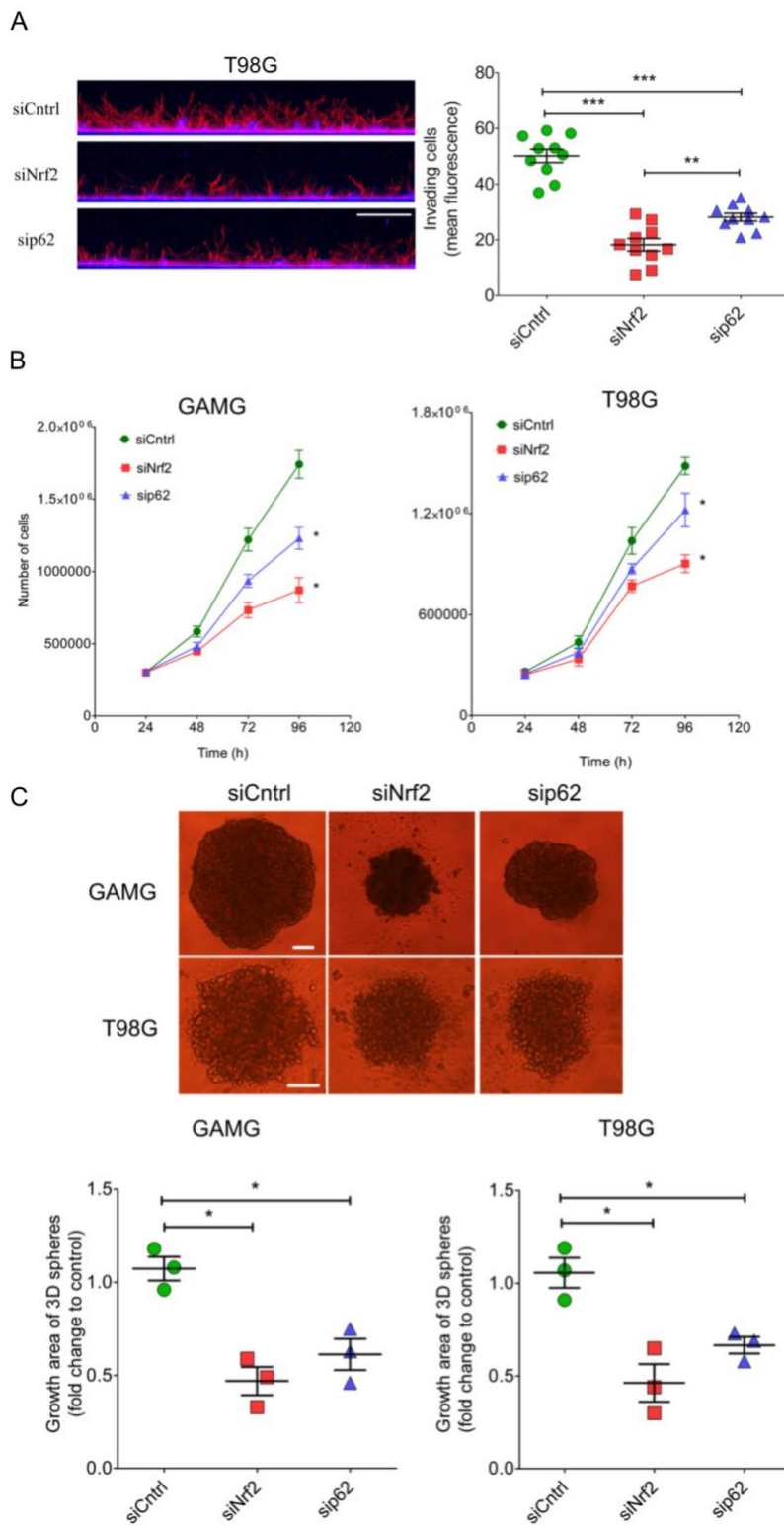
Fig. S11. Verification of Nrf2 and p62 siRNA results using a different siRNA. A-C, as in Fig. 7 B-D. A. T98G cell invasion in type I collagen-cultrex matrix with a different Nrf2 and p62 siRNAs. Results showed that both siNrf2 and sip62 inhibit T98G invasion. Red = Actin, Blue = DAPI nuclei staining. B. Cell proliferation of GAMG and T98G with siNrf2 and sip62 was significantly inhibited. C. Anchorage-independent growth of GAMG and T98G with siNrf2 and sip62 was also considerably reduced. D. Quantification of results from C showed that the growth reduction was significant. n = 3 and n = 10 in A, mean ± SEM, *p < 0.05 (One-way ANOVA, Tukey's test). Scale bar = 100 μm.

Supplementary Table legends and Table S3-6

Table S1. SIGNATURE prediction of Nrf2 activity in CCLE data set. Table contains also annotations for cell lines. Second tab contains Nrf2 signature genes.

Table S2. Joint pathway analysis of CCLE glioma, TCGA and HGCC samples.

Table S3. Reagent sources

Cell culture	Reagent name	Source
	DMEM	Sigma-Aldrich
	FBS	Sigma-Aldrich
	penicillin/streptomycin	Sigma-Aldrich
	Minimum Essential Medium	Sigma-Aldrich
	L-glutamine	Sigma-Aldrich
	MEM nonessential amino acids	Sigma-Aldrich
	RPMI-1640 medium	Sigma-Aldrich
	hygromycin	Sigma-Aldrich
	Bafilomycin A	Sigma-Aldrich
	DMSO	Sigma-Aldrich
Antibodies	rabbit polyclonal Nrf2 antibody	Santa Cruz Biotechnology, sc-13032
	rabbit polyclonal Keap1 antibody	Proteintech Europe, #10503-2-AP
	rabbit polyclonal β-actin antibody	Cell Signaling, #4967L
	mouse monoclonal β-actin antibody	sc-47778, Santa Cruz Biotechnology
	rabbit polyclonal HMOX1 antibody	Stressgen, #SPA-896
	mouse monoclonal p62/SQSTM1 antibody	sc28359, Santa Cruz Biotechnology
	rabbit polyclonal LC3B	L7543, Sigma-Aldrich
	rabbit polyclonal phospho-p62 (Ser-403)	Cell Signaling, #14354
	anti-rabbit phospho-p62 antibody (Ser-351)	gift from Dr. Masaaki Komatsu (Tokyo University, Japan)
	mouse monoclonal NQO1 clones A180 and B771	gifts from Dr. David Ross, University of Colorado, Denver, USA
	rabbit polyclonal GCLM	gifts from Dr. Terrance J. Kavanagh, University of Washington, Seattle, USA
	Antibodies for EMT markers (snail, slug, vimentin and beta-catenin)	Cell Signaling, #9782
	mouse monoclonal p62/SQSTM1 antibody	Santa Cruz Biotechnology, sc28359
	mouse monoclonal vimentin antibody	Proteintech, 60330-1-Ig
	goat polyclonal Keap1 antibody	Santa Cruz Biotechnology, sc-15246
	Alexa Fluor® 488 Donkey anti-goat IgG	Invitrogen

	Alexa Fluor® 647 rabbit anti-mouse IgG	Invitrogen
	TRI-reagent	Sigma-Aldrich
	Transcriptor First Strand cDNA synthesis kit	Roche
RNA isolation and qPCR	non-specific siRNA control (Stealth siRNA control)	Invitrogen
	RNAiMax	Invitrogen
siRNA transfections	Epon LX-112	Ladd Research Industries, Burlington, VT
Other	type I collagen	BD biosciences
	cultrex	Trevigen
	DAPI	Invitrogen, #D1306

Table S4. Primers used in q-RT-PCR together with Roche's probe system.

Gene	Name	Sequence (left)	Sequence (right)
NFE2L2	Nuclear factor (erythroid-derived 2)-like 2	CCCTGTTGATTTAGACGGTA	CTCAACCAGCTTGTCATTTTC
KEAP1	Kelch like ECH associated protein 1	TTTCGTAGCCCCCATGAAG	TGTGACCATCATAGCCTCCA
GCLM	Glutamate-cysteine ligasemodifier subunit	GTTGGAACAGCTGTATCAGTGG	CAGTCAAATCTGGTGGCATC
NQO1	NAD(P)H dehydrogenase quinone 1	CAGCTCACCGAGAGCCTAGT	GAGTGAGCCAGTACGATCAGTG
HMOX1	Heme oxygenase (decycling) 1	GGCAGAGGGTGATAGAAGAGG	AGCTCCTGCAACTCCTCAAA
p62/SQSTM1	Sequestosome 1	TGCCTTGTACCCACATCTCC	AGCCGCCTTCATCAGAGA
AKR1C1	Aldo-keto reductase family 1member C1	CATGCCTGTCCTGGGATTT	AGAATCAATATGGCGGAAGC
MGMT	O6-methylguanine-DNA methyltransferase	GGGTCTGCACGAAATAAAG	CTCCGGACCTCCGAGAAC
SNAIL	Snail Family Transcriptional Repressor 1	CCTCCAGACCCACTCAGATG	AGGGAATTCCATGGCAGTGA
SLUG	Snail Family Transcriptional Repressor 2	TGGCTCAGAAAGCCCCATTA	GCCAGCCCAGAAAAAGTTGA
ZEB1	Zinc Finger E-Box Binding Homeobox 1	CTCTGGGATGCGAAACGC	GGTACACCTTCACAGTCAGC
CTNNB1	Catenin beta 1	ATGATGGTCTGCCAAGTGGG	TCCTGGCCATATCCACCAGA

Table S5. MOI values used for transfections

Cell line	MOI
GAMG	5
NHA	5
A549	10
U87	10
U251	10
A172	10
LN18	10
T98G	10
H4	20
DKMG	40

Table S6. Primers used in ChIP q-RT-PCR

Oligo Name	Sequence (5' - 3')
NQO1 forward	TCCAAATCCGCAGTCACAGT
NQO1 reverse	TTGGCACGAAATGGAGC
HMOX1 forward	TGAGTAATCCTTTCCCGAGC
HMOX1 reverse	GTGACTCAGCGAAAACAGACA
SNAI2 forward	GGTGAGAATGTGGAGAAACGAGATC
SNAI2 reverse	GTGTGGAAGTAAGTCGTCATCTCTCT
CTNNB1 forward	GGACTGGCATCTAATGGGTAGAGCC
CTNNB1 reverse	GCATGGATCAGATGGTCTAGCCTATGG
IgK_neg_ChIP forward	CAGCAGATAGAGGCACCAGA
IgK_neg_ChIP reverse	CTCCCACAGTGTTGCAAGTC
Chr1_neg_ChIP forward	AGCCACCACTACTCCCTCTA
Chr1_neg_ChIP reverse	GCTGCCTGGAATTAAGACGG