

EPIGENETIC SILENCING FOR TREATMENT OF CANCER

FIELD OF THE INVENTION

The present invention relates to gene silencing and/or epigenetic editing, in particular for the treatment of cancer. For example, the invention relates to epigenetic silencer factors (ESFs) comprising a transcription factor DNA-binding domain operably linked to at least one epigenetic effector domain, and their use in the treatment of cancer, such as brain tumours.

BACKGROUND TO THE INVENTION

Gene therapy involves the incorporation of genetic material into a cell to treat or prevent disease. The genetic material may supplement defective genes with functional copies of those genes, inactivate improperly functioning genes, silence genes that may be associated with a disease state (e.g. oncogenic or cancer-associated genes) or introduce new therapeutic genes to a cell.

To date, two main targeting technologies have been used to silence gene expression: RNA interference (RNAi) with single short hairpin RNA (shRNA); and gene targeting with artificial nucleases. Although promising pre-clinical and clinical data have been obtained using these technologies, partial depletion of gene expression with shRNA and the low efficiency by which homozygous disruption occurs in diploid mammalian cells may jeopardise efficacy of these treatments. These disadvantages are particularly relevant in those applications where residual levels of gene activity are sufficient for biological function.

In addition, epigenetic mechanisms have been exploited to silence gene expression. Epigenetics refers to mechanisms that convey heritable changes in the function of the genome without altering the primary DNA sequence. These changes can mediate short-term instructions that can be quickly reverted in response to exogenous stimuli (e.g. histone post-transcriptional modifications; HPTMs). Alternatively, they can constitute long-term instructions that stably contribute to cellular identity and memory (e.g. DNA methylation).

Treatment of cancer may involve a range of approaches including surgery, chemotherapy and radiotherapy. However, even if surgical removal of a tumor is as radical as possible, a few remaining cancer cells with tumor-initiating potential may be sufficient to regrow a tumoral mass in the short term, leading to the reappearance of the disease. In particular, cancer stem cells (CSCs), defined as cells able to self-renew and to initiate or regrowth the tumor, remain quiescent or at very low proliferative activity and may be capable of resisting

certain adjuvant treatments. Thus, there is a significant need to achieve a long-lasting remission, in particular after tumor resection by a more efficient targeting of cancer cells.

Such approaches may be particularly desirable for diseases such as glioblastoma multiforme (GBM), which is the most common and lethal brain cancer in adults with 1-5 cases per 100000 people per year and 12-15 months of median survival. This poor outcome is due to the combination of both the aggressiveness of the disease and the limited efficacy of current therapies that increase the overall survival only marginally. Patients usually undergo surgical resection of the primary tumor mass followed by adjuvant radio- and chemo- (Temozolomide) therapies, although the issues with tumor regrowth described above may lead to recurrence of the cancer.

Attempts have been carried out to restrain cancer (e.g. GBM) development by silencing expression of one or more transcription factors (TFs) with different technologies. For example, TF inactivation has been attempted with several technologies including shRNA, miRNA and TALEN-based epigenetic repressors, but complete and long-term gene silencing has proven challenging. Moreover, cancer cells may rearrange their genetic program to cope with the silencing of a single gene and, thus, maintain unaltered tumorigenic potential.

Accordingly, there remains a significant need for the development of more effective treatments for cancer, in particular treatments that are effective against aggressive cancers such as GBM, and treatments that enable targeting of CSCs

SUMMARY OF THE INVENTION

The present inventors have engineered oncogenic and cancer-associated transcription factors to function as epigenetic repressors, which, for example, enable silencing of downstream tumorigenic networks, thus limiting CSC survival and proliferation.

For example, the inventors engineered the SOX2 transcription factor to generate an epigenetic repressor (referred to as SOX2 Epigenetic Silencer (SES)) that maintains the ability to bind to a large group of its original targets and induce their long-term silencing. By way of example, the inventors deleted the C-terminal domain of SOX2 and fused it with a KRAB domain and the DNA methyltransferase 3A/3L catalytic domains. The inventors found that SES is able to inhibit the SOX2 tumorigenic network (rather than activate it as with the unmodified TF), including genes crucial for tumor maintenance and growth.

The inventors found that SES kills both glioma cell lines and patient-derived CSCs both in vitro and in vivo. Moreover, following in situ viral transduction of GBM xenografts in mice,

SES induced a strong growth inhibition of pre-formed tumor of human origin. The inventors also found that SES is not harmful to neurons and glia, and does not trigger significant transcriptional changes in these cells.

5 The inventors further validated their approach by successfully utilizing other epigenetic effector domains, such as the chromo shadow and YAF2-RYBP domains, and also by engineering additional transcription factors, such as MYC and TEAD1.

10 Compared to previous approaches aimed at silencing the expression of endogenous oncogenic transcription factors, the engineered transcription factors developed by the inventors (referred to as epigenetic silencer factors (ESFs)) may bind to the same targets as the unmodified transcription factors and induce their strong and permanent silencing. The inventors' approach may advantageously lead to repression of the transcriptional cascade of the transcription factor, and strongly reduce any issue of gene reactivation that can lead to tumor recurrence.

15 In one aspect, the invention provides an epigenetic silencer factor (ESF) comprising a transcription factor DNA-binding domain operably linked to at least one epigenetic effector domain, wherein the transcription factor is an oncogenic transcription factor or a cancer-associated transcription factor.

20 In some embodiments, the ESF is a polypeptide comprising the transcription factor DNA-binding domain and the at least one epigenetic effector domain. In some embodiments, the ESF is a fusion protein comprising the transcription factor DNA-binding domain and the at least one epigenetic effector domain.

In preferred embodiments, the transcription factor is an oncogenic transcription factor. In some embodiments, the transcription factor is a cancer-associated transcription factor.

25 In some embodiments, the transcription factor is selected from the group consisting of SOX2, MYC, MYCN, TEAD1, TEAD2, TEAD3, TEAD4, FOXA1, FOXA2, ELK1, ELK3, ELK4, SRF, FOXM1, FOXC1, FOXC2, TWIST1, SALL4, ELF1, HIF1A, SOX9, SOX12, SOX18, ETS1, PAX3, PAX8, GLI1, GLI2, GLI3, ETV1, ETV2, ETV3, RUNX1, RUNX2, RUNX3, MAFB, TFAP2C and E2F1.

In some embodiments, the transcription factor is SOX2.

30 In some embodiments, the transcription factor is MYC.

In some embodiments, the transcription factor is TEAD1.

In preferred embodiments, the ESF does not comprise a transcription factor activation domain.

In some embodiments, the at least one epigenetic effector domain is selected from the group consisting of a KRAB domain, a DNMT3A domain, a DNMT3L domain, a ZIM3-KRAB (Z-KRAB) domain, a Chromo Shadow (CS) domain, a YAF2-RYBP (Y-R) domain, an Engrailed Repressor (En-R) domain, a MeCP2 domain, a GLI3RD domain and a MAD1RD domain.

In some embodiments, the ESF comprises a KRAB domain. In some embodiments, the ESF comprises a DNMT3A domain. In some embodiments, the ESF comprises a DNMT3L domain. In preferred embodiments, the ESF comprises a KRAB domain, a DNMT3A domain and a DNMT3L domain.

In some embodiments, the KRAB domain is a ZNF10 KRAB domain. In some embodiments, the DNMT3A domain is the catalytic domain of DNMT3A. In some embodiments, the CS domain is a CBX5 CS domain.

In some embodiments, the ESF comprises a ZIM3-KRAB (Z-KRAB) domain. In some embodiments, the ESF comprises a Chromo Shadow (CS) domain. In some embodiments, the ESF comprises a YAF2-RYBP (Y-R) domain. In some embodiments, the ESF comprises an Engrailed Repressor (En-R) domain. In some embodiments, the ESF comprises a ZIM3-KRAB (Z-KRAB) domain, a Chromo Shadow (CS) domain and a YAF2-RYBP (Y-R) domain. In some embodiments, the ESF comprises a KRAB domain and a DNMT3A domain. In some embodiments, the ESF comprises a KRAB domain and an MeCP2 domain. In some embodiments, the ESF comprises a GLI3RD domain. In some embodiments, the ESF comprises a MAD1RD domain.

In another aspect, the invention provides a polynucleotide comprising a nucleic acid sequence encoding the ESF of the invention.

In some embodiments, the polynucleotide further comprises a promoter operably linked to the nucleic acid sequence encoding the ESF.

In some embodiments, the promoter is a constitutive promoter. In some embodiments, the constitutive promoter is an Ef1a promoter.

In some embodiments, the promoter is a tissue-specific promoter, preferably a cancer cell-specific promoter.

In some embodiments, the promoter is a proliferating cell-specific promoter.

In some embodiments, the promoter is selected from the group consisting of a Mki67 promoter, a Ccnd1 promoter, a Ccnb2 promoter, a Ccna2 promoter, a Cdc25c promoter, a Cdc2 promoter, a Cks1 promoter, a PCNA promoter, a CDC6 promoter, a POLD1 promoter, a CSK1B promoter, a MCM2 promoter and a PLK1 promoter.

- 5 In some embodiments, the promoter is a Mki67 promoter.

In another aspect, the invention provides a vector comprising the polynucleotide of the invention.

- 10 In some embodiments, the vector is a viral vector. In some embodiments, the vector is a lentiviral vector or adeno-associated viral (AAV) vector. In some embodiments, the vector is a lentiviral vector.

In some embodiments, the vector is an mRNA vector.

In some embodiments, the ESF, polynucleotide or vector is comprised in a nanoparticle. In some embodiments, the nanoparticle is a polymeric nanoparticle, inorganic nanoparticle or lipid nanoparticle. In some embodiments, the nanoparticle is a liposome

- 15 In another aspect, the invention provides a cell comprising the ESF, polynucleotide or vector of the invention.

In another aspect, the invention provides a composition comprising the ESF, polynucleotide, vector or cell of the invention. In some embodiments, the composition is a pharmaceutical composition.

- 20 The composition may be a hydrogel. In some embodiments, the hydrogel is a poly(ethylene glycol) dimethacrylate (PEG-DMA) hydrogel. In some embodiments, the hydrogel further comprises hydroxyapatite nanoparticles.

In another aspect, the invention provides the ESF, polynucleotide, vector, cell or composition of the invention for use in therapy.

- 25 In another aspect, the invention provides the ESF, polynucleotide, vector, cell or composition of the invention for use in the treatment of cancer.

In some embodiments, the cancer is glioma, glioblastoma, medulloblastoma, astrocytoma, neuroblastomas, ependymoma, meningioma, retinoblastoma, rhabdomyosarcoma, lung cancer, prostate cancer, breast cancer, liver cancer, pancreatic cancer, bladder cancer,

oropharyngeal cancer or kidney cancer. In some embodiments, the cancer is a brain tumour. In some embodiments, the cancer is glioblastoma multiforme.

In some embodiments, the treatment reduces tumour size.

In some embodiments, the treatment is as an adjuvant therapy, optionally in combination with surgery. The treatment may reduce the risk of cancer (e.g. GBM) recurrence.

In some embodiments, the ESF, polynucleotide, vector, cell or composition (e.g. hydrogel) is administered locally.

In another aspect, the invention provides a method of treating cancer in a subject, the method comprising administering the ESF, polynucleotide, vector, cell or composition of the invention to a subject in need thereof.

In another aspect, the invention provides use of the ESF, polynucleotide, vector, cell or composition of the invention for the manufacture of a medicament for therapy.

In another aspect, the invention provides use of the ESF, polynucleotide, vector, cell or composition of the invention for the manufacture of a medicament for treatment of cancer.

In another aspect, the invention provides use of the ESF, polynucleotide, vector, cell or composition of the invention for decreasing transcription and/or expression of at least one target gene in a cell.

In another aspect, the invention provides a method of decreasing transcription and/or expression of at least one target gene in a cell, the method comprising introducing the ESF, polynucleotide, vector or composition of the invention into the cell.

In some embodiments, the at least one target gene is silenced, preferably permanently silenced.

DESCRIPTION OF THE DRAWINGS

FIGURE 1. Generation and testing of SOX2 epigenetic silencer (SES).

(a) Constructs generated based on human SOX2 transcription factor and the epigenetic domain KRAB, DNMT3a (3A) and DNMT3L (3L), V5 was added as tag. (b) Infection efficiency of lentivirus carrying the indicated constructs in SNB19 cells. (c) Left, growth curve of SNB19 cells infected with the indicated constructs show that SES was able to kill the cells after 12 days in culture, ****p < 0.0001; statistically compared with two-way ANOVA; right, microphotographs of the cells at the indicated time points from the infection with either mock

(GFP) or SES. (d) Western blot (WB) for V5, SOX2 and H3 (as loading control) in SNB19 cells either not infected or infected with lentivirus carrying GFP or SES. (e) Left, growth curve of U87 cells, ****p < 0.0001; statistically compared with two-way ANOVA; right, microphotographs of the cells at pre-treatment and after 9 days from the infection with either mock (GFP) or SES. f) Left, growth curve of U251 cells, ***p < 0.001; statistically compared with two-way ANOVA; right, microphotographs of the cells at pre-treatment and after 10 days from the infection with either mock (GFP) or SES. (g) Quantification of the indicated SOX2 targets in SNB19 cells 3 days after infection with either GFP or SES. ****p < 0.0001, ***p < 0.001; statistically compared using unpaired t test. (h) Growth curves of the indicated cancer cell lines. ***p < 0.001, ns = not significant; statistically compared with two-way ANOVA.

FIGURE 2. Efficacy of SOX2 epigenetic silencer on patient derived cancer stem cells in vitro.

(a) Microphotographs, growth curve and percentage of death cells of patient derived stem cells (CSCs) of classical (left) and mesenchymal (right) GBM subtype infected with either GFP (both subtypes), SES (both) or binding defective SES(R74P-L97P) (Classical subtype only). ****p < 0.0001; ***p < 0.001; statistically compared with two-way ANOVA. CSCs (classical type) either not infected or infected with lentivirus carrying GFP or SES were assayed for their clonogenic potential using sphere number and dimension (percentage of spheres below 100 um in diameter) at the indicated time points as parameters. ****p < 0.0001; statistically compared with two-way ANOVA.

FIGURE 3. Molecular consequences of SES.

(a) SES caused massive gene deregulation in both U87 and SNB19 cells as assessed by RNA-seq. (b) IGV snapshots of RNAseq tracks within Sox2 locus in both cell lines in each condition (Mock and SES infected) indicate the overexpression of the initial part of SOX2 (included in SES construct, see Fig 1a) both in SOX2 negative (U87) and SOX2 positive (SNB19) cell lines. (c) Gene Ontology analysis indicated that gene associated with apoptosis (upregulation) and cell cycle regulation (downregulation) were impaired by SES expression. (d) Genes predicted to be regulated by SOX2 were affected, with the majority being downregulated. (e) Density plot of ChIP-seq normalized signals (SOX2 and SES) on the SOX2 peaks indicated that SES binds the same regions at genome wide level. (f) Density plot of MeDIP-seq normalized signals (Mock and SES) on the SOX2 peaks indicated that SES is able to increase DNA methylation levels.

FIGURE 4. SES functionality *in vivo*: heterotopic xenograft.

- (a) Heterotopic xenograft by subcutaneous injection of 1 million of GBM cells, pre-infected with either mock (GFP) or SES in NSG mice. (b) After 4 weeks from the injection, mock U87 cells always generated huge masses while in only one case we retrieved one small nodule from SES infected cells. (c) Evaluation of volume of tumors generated using the indicated cells after the indicated time windows (Mo = months).

FIGURE 5. SES functionality *in vivo*: orthotopic xenograft.

- (a) Orthotopic xenograft by injection in the striatum of 300000 cells, pre-infected with either mock (GFP) or SES in the brain of NSG mice. (b-c) After 25-30 days from the injection, mock U87 cells always generated huge GFP positive tumors that were also able to invade the cortex, while we never detected any tumors in brains transplanted with SES U87 cells by either Nissl histological staining (b) or V5 antibody (c). (d) Tumors arising from mock cells were formed by human nuclei (HuN) positive, proliferating (PH3 positive) cells while virtually no human cells were retrieved in brain injected with SES cells. Of note, the few PH3 positive cells present, were located at the level of the lateral ventricles, being presumably mouse neural precursors in active division. (e) Estimation of tumor volume. (f) Kaplan-Meier curve shows that mice injected with mock cells died within a month from the surgery while SES receiving animals were in good health and the time of sacrifice (n = 5 animals per group). (g) Xenografts of mock infected CSCs gave rise huge tumors while mice injected with SES CSCs displayed small tumors after 6 weeks from the surgery. (h) Estimation of tumor volume. (i) Kaplan-Meier curve shows that mice injected with mock cells died within 6 weeks from the surgery while SES receiving animals were better condition at the time of sacrifice (n = 5 animals per group).

FIGURE 6. GBM-cortex organoids.

- (a) Early patterned cortical organoids were seeded with GBM floating spheres (labelled with RFP) to obtain a fusion. (b) Fused GBM-cortical organoids using either mock or SES-infected spheres (1 week after seeding) were fixed, cut and stained. Notably SES limits the growth as well as the infiltration of GBM cells (RFP positive) in normal cortical parenchyma (DAPI only staining).

FIGURE 7. SES functionality *in vivo*: treatment of preformed orthotopic xenograft.

- (a) The orthotopic xenograft was generated by injection of 75000 naïve U87 cells in the striatum of NSG mice; after 4 days the animals were reoperated to inject lentivirus carrying either mock (GFP) or SES and evaluated 26 days after (30 days in total). (b) Histological staining at the time of the lentiviral injection (pre-treatment) and at the end of the protocol

(post-treatment) showed that the growth of the tumor is limited in SES treated animals. (c) Estimation of tumor volume after 26 days from the lentiviral injection (treatment). (d) Kaplan-Meier curve shows that mice which tumors were injected with mock virus died within two months from the injection of the cells while all but one SES receiving animals reached 3 months (n = 4 animals per group). (e) Immunohistochemistry indicated that the resulting tumors were GFP positive in the case of the mock treatment, and were V5 negative in the case of the SES injections (asterisk) suggesting a negative selection effect for those tumor cells that were infected by the SES virus. Of note V5 labelling (arrowheads) were present in the parenchyma of the mouse brain surrounding the tumor. (f) The orthotopic xenograft was generated by injection of 100000 naïve CSCs in the striatum of NSG mice; after 7 days the animals were reoperated to inject lentivirus carrying either mock (GFP) or SES and evaluated by MRI scanning and sacrificed 6 weeks after the treatment (7 weeks from the grafting of cells). (g) Examples of MRI scanning 3, 4 and 5 weeks post infection (p.i.) (two slices for both one mock and one SES treated mouse). (h) Histological staining at the end of the protocol (endpoint) showed that the growth of the tumor is limited in SES-treated animals. (i) Evaluation of the tumor volume as measured by hyper-intensity of T2-weighted imaging (MIPAV software). (j) Estimation of tumor volume by histological measurements at the endpoint and thus after 6 weeks from the lentiviral injection (treatment), 7 from the CSC injection.

20 **FIGURE 8. SES effect on cultured neurons.**

(a) Infection and death evaluation of primary murine hippocampal neurons infected with either mock or SES indicated no neurodegeneration induced by SES. (b) SES caused only marginal gene deregulation in mouse primary neurons as assessed by RNA-seq. (c) Human iPSC-derived neurons were infected with either mock or SES. (d) The evaluation of neuronal loss through the staining for PI, V5 and MAP2 indicated that the presence of SES was not increasing the neuronal death at least at 21 days after the infection.

FIGURE 9. SES effect on normal mouse brain.

(a) Mock or SES lentiviral injections in the hippocampi of WT c57bl/6 mice. (b) Example of viral transduction in the murine hippocampus, using GFP as reporter, after 4 weeks from the injection. (c) Quantification of both viral genome and mRNA of exogenous transgenes in infected hippocampi by qPCR show no difference between the conditions. (d) Quantification of cleaved Caspase 3 positive cells within infected hippocampi shows no difference between the conditions indicating that SES was not toxic for murine neural cells in this setting. (e) Spontaneous alternation test suggested no difference between mock and SES injected mice

as assessed by: the percentage of the entries in the different arms, and both the percentage of spontaneous alternation performance (SAP), the percentage of alternate arm return (ARR) and the percentage of same arm return (SAR) on the total entries. Statistically compared with Mann-Whitney test. (f) Radial maze test indicated no difference in the time to accomplish the task or particular tendency of SES treated animals in committing errors during the entire protocol of the test compared to the mock injected animals. Statistically compared with two-way ANOVA. (g) Morris water maze test. On the left the protocol we used and the scheme of the platform (black square) positioning. On the right, the plot of the time used to accomplish the task (up) and the quantification of the time spent in the platform area or in the opposite one (bottom) indicate no differences between conditions. Statistically compared with Mann-Whitney test.

FIGURE 10. Vector amelioration.

(a) Scheme depicts the original SES (v1) and the further version (v1.1) carrying a different promoter (KI67 promoter expressed in proliferating cells, v1.1). (b) The SES v1.1 test in U251 cells indicated that the KI67 promoter guides the expression of a transgene (either GFP or SES) in a very high percentage of KI67 positive proliferating cells. The SES v1.1 effect on the growth was similar compared to the original version. (c) Both a constitutive GFP and pKI67-GFP were used on mouse primary cortical cultures that contain mainly post-mitotic neurons but also glial types, both proliferating and post-mitotic. Immunofluorescence with the indicated antibodies, indicated that GFP when guided by pKI67 was found only in KI67⁺ proliferating cells (red arrows and quantification) while never in MAP2⁺ neurons (white arrows, and quantification). Constitutive GFP was observed in virtually all cells as expected.

FIGURE 11. Generation of other ESFs.

(a) Construct generated based on human SOX2 transcription factor fused with the epigenetic domain chromo shadow (CS) and V5 was named SES v2. Construct generated based on human SOX2 transcription factor fused with the epigenetic domain YAF2-RYBP (Y-R) and V5 was named SES v3. (b) Infection efficiency of lentivirus carrying the indicated constructs (see a panel) in U251 cells. (c) Growth curve of U251 cells infected with the indicated constructs show that SESv2 and v3 were able to kill the cells after 9 days in culture, ****p < 0.0001; statistically compared with two-way ANOVA. (d) Construct generated based on human TEAD1 transcription factor and the epigenetic domain KRAB, DNMT3a (3A), DNMT3L (3L) and V5 was named TES, while the construct generated based on human MYC transcription factor and the epigenetic domain KRAB, DNMT3a (3A), DNMT3L (3L) and V5 was named MES. (e) Infection efficiency of lentivirus carrying the indicated constructs (see a

panel, right) in U251 cells. (e) Growth curve of U251 cells infected with the indicated constructs show that MES was able to kill the cells after 9 days in culture, ****p < 0.0001.

FIGURE 12. Efficacy of SESv3 on patient-derived cancer stem cells *in vitro*

(a) Microphotographs and (b) growth curve of patient-derived stem cells (CSCs) of classical GBM subtype infected with either GFP or SESv3. ****p < 0.0001; ***p < 0.001; statistically compared with two-way ANOVA.

FIGURE 13. Efficacy of TES and MES on patient-derived cancer stem cells *in vitro*

(a-c) Microphotographs (a), growth curve and percentage of death cells (b-c) of patient-derived stem cells (CSCs) of classical (left) and mesenchymal (right) GBM subtype either not infected (NI) or infected with GFP or TES. *p < 0.05; ***p < 0.001; ****p < 0.0001; ns = not significant; statistically compared with two-way ANOVA. (d-e) Growth curve and percentage of death cells of patient-derived stem cells (CSCs) of classical (left, d) and mesenchymal (right, e) GBM subtype either not infected (NI) or infected with GFP or TES. *p < 0.05; ***p < 0.001; ****p < 0.0001; ns = not significant; statistically compared with two-way ANOVA.

FIGURE 14. TES/MES functionality *in vivo*: heterotopic xenograft

(a) Heterotopic xenograft by subcutaneous injection of 3,000,000 of classical CSCs, pre-infected with either mock (GFP) or TES or MES in NSG mice. After 4 weeks from the injection, mock CSCs always generated huge masses while smaller tumors emerged from TES (b-c) or MES infected cells (d-e).

FIGURE 15. TES/MES functionality *in vivo*: orthotopic xenograft

(a) Orthotopic xenograft by injection in the striatum of 300,000 classical CSCs, pre-infected with either mock (GFP) or TES or MES in brain of NSG mice. (b) After 5 weeks from the injection (WPI), mock CSCs always generated huge tumors able to invade also the cortex while TES cells formed smaller tumors in brains. Evaluation by Nissl histological staining. (c) Estimation of tumor volume (n = 4 animals per group). (d) After 3 weeks from the injection (WPI), mock CSCs already showed important tumor masses while MES cells formed smaller tumors. Evaluation by DAPI staining. (e) Estimation of tumor volume (n = 4 animals per group).

FIGURE 16. TES functionality *in vivo*: treatment of preformed orthotopic xenograft

- (a) The orthotopic xenograft was generated by injection of 300.000 classical CSCs in the striatum of NSG mice; after 7 days the animals were re-operated to inject lentivirus carrying either mock (GFP) or TES and evaluated 3 weeks after (WPT) (4 weeks in total). (b) DAPI staining at the end of the protocol (post-treatment) showed that the growth of the tumor is limited to the site of injection in TES treated animals while huge masses are present in mock treated mice. (c) Estimation of tumor volume after 3 weeks from the lentiviral injection (treatment) (n = 3 animals per group).

DETAILED DESCRIPTION OF THE INVENTION

- The terms “comprising”, “comprises” and “comprised of” as used herein are synonymous with “including” or “includes”; or “containing” or “contains”, and are inclusive or open-ended and do not exclude additional, non-recited members, elements or steps. The terms “comprising”, “comprises” and “comprised of” also include the term “consisting of”.

EPIGENETIC SILENCER FACTOR (ESF)

- In one aspect, the invention provides a polypeptide comprising a transcription factor DNA-binding domain operably linked to at least one epigenetic effector domain, preferably wherein the transcription factor is an oncogenic transcription factor or a cancer-associated transcription factor. The polypeptide may be for decreasing transcription and/or expression of one or more target gene. The polypeptide may be referred to as an epigenetic silencer factor (ESF). The polypeptide may be a multimeric polypeptide, for example comprised of two, three or more polypeptide chains. For example the polypeptide may be a dimer, such as a heterodimer. The polypeptide may be a comprised of a single polypeptide chain. The polypeptide may be a fusion protein.

- In another aspect the invention provides an epigenetic silencer factor (ESF) comprising a transcription factor DNA-binding domain operably linked to at least one epigenetic effector domain, preferably wherein the transcription factor is an oncogenic transcription factor or a cancer-associated transcription factor.

- ESFs are agents that may decrease the transcription and/or expression of one or more target gene (e.g. silence one or more target gene). ESFs of the invention may comprise at least part of a transcription factor that binds to DNA, wherein the part is operably linked to an epigenetic effector domain. The effector domain may have transcriptional repression activity and may enable silencing (e.g. permanent silencing) of one or more target gene of the transcription factor. In particular, when the transcription factor is an oncogenic transcription

factor or cancer-associated transcription factor the ESF may block a gene expression cascade involved in tumor growth.

ESFs may be chimeric or fusion proteins that are comprised of a DNA-binding domain operably linked to an effector domain (e.g. a KRAB domain, a DNMT3A domain and/or a DNMT3L domain). The effector domain may harbour a catalytic activity which represses transcription and/or expression of one or more target gene. Alternatively, or additionally, the effector domain may recruit additional agents within a cell to one or more target gene, which may repress transcription and/or expression of the target gene(s).

By "operably linked", it is to be understood that individual components are linked together in a manner which enables them to carry out their function (e.g. binding to DNA, catalysing a reaction or recruiting additional agents from within a cell) substantially unhindered. For example, a DNA-binding domain may be conjugated to an effector domain, for example to form a fusion protein. Methods for conjugating polypeptides are known in the art, for example through the provision of a linker amino acid sequence connecting the polypeptides (e.g. a linker comprising glycine and/or serine residues). Alternative methods of conjugating polypeptides known in the art include chemical and light-induced conjugation methods (e.g. using chemical cross-linking agents). Preferably, the DNA-binding domain and effector domain of the ESF form a fusion protein.

In some embodiments, the ESF is a fusion protein comprising a transcription factor DNA-binding domain and at least one epigenetic effector domain.

The ESF may be formed by separate polypeptide chains that bind together to form a complex, for example a heterodimeric complex. For example, the binding may be enabled by an epitope (e.g. a Suntag) comprised on a first chain and an epitope-binding molecule, such as a single-chain variable fragment (scFv), comprised on a second chain.

In some embodiments, the ESF comprises two epigenetic effector domains, for example fused to the same DNA-binding domain. In some embodiments, the ESF comprises three epigenetic effector domains, for example fused to the same DNA-binding domain. The ESF may comprise four, five, six or more epigenetic effector domains, for example fused to the same DNA-binding domain.

Where the ESF comprises more than one epigenetic effector domain, the effector domains may be different. Where the ESF comprises more than one epigenetic effector domain, the effector domains may be the same.

In preferred embodiments, the ESF comprises a KRAB domain, a DNMT3A domain and a DNMT3L domain.

EPIGENETIC EFFECTOR DOMAIN

5 The term “epigenetic effector domain”, is to be understood as referring to a part of the ESF which provides for an epigenetic effect on a target gene, for example by catalysing a reaction on DNA or chromatin (e.g. methylation of DNA, methylation or acetylation of a histone, or demethylation or deacetylation of a histone), or by recruiting an additional agent, resulting in the repression of the transcription of the gene.

10 “Domain” is to be understood in this context as referring to a part of the ESF that harbours a certain function. The domain may be an individual domain (e.g. a catalytic domain) isolated from a natural protein or it may be an entire, full-length natural protein. Put another way, either the full-length protein or a functional fragment thereof can be used as an effector domain. Therefore, for example, “KRAB domain” may refer to a part of the ESF that comprises an amino acid sequence with the function of a KRAB domain.

15 Chromatin remodelling enzymes that are known to be involved in the permanent epigenetic silencing of endogenous retroviruses (ERVs; Feschotte, C. et al. (2012) Nat. Rev. Genet. 13: 283-96; Leung, D.C. et al. (2012) Trends Biochem. Sci. 37: 127-33) may provide suitable effector domains for exploitation in the invention.

20 In some embodiments, the epigenetic effector domain represses transcription and/or expression of at least one target gene. In some embodiments, the epigenetic effector domain is a repressor domain.

In some embodiments, the epigenetic effector domain catalyses chemical modification of chromatin and/or chromatin remodelling.

25 In some embodiments, the epigenetic effector domain catalyses DNA modification, such as DNA methylation. In some embodiments, the epigenetic effector domain is a DNA methyltransferase and/or is capable of recruiting a DNA methyltransferase.

30 In some embodiments, the epigenetic effector domain catalyses histone modification, such as histone methylation or histone acetylation. In some embodiments, the epigenetic effector domain is a histone methyltransferase or histone acetyltransferase. In some embodiments, the epigenetic effector domain catalyses histone demethylation or histone deacetylation. In some embodiments, the epigenetic effector domain is a histone methylase or histone acetylase.

Krüppel-associated box (KRAB) domain

The family of the Krüppel-associated box containing zinc finger proteins (KRAB-ZFP; Huntley, S. et al. (2006) Genome Res. 16: 669-77) plays an important role in the silencing of endogenous retroviruses. These transcription factors bind to specific ERV sequences through their ZFP DNA-binding domain, while they recruit the KRAB Associated Protein 1 (KAP1) with their conserved KRAB domain. KAP1 in turn binds a large number of effectors that promote the local formation of repressive chromatin (Iyengar, S. et al. (2011) J. Biol. Chem. 286: 26267-76). For example, they may induce repressive chromatin modification (e.g. H3K9me3) and/or remove active marks (e.g. H3K4ac)

10 In some embodiments, the ESF comprises a KRAB domain.

Various KRAB domains are known in the family of KRAB-ZFP proteins. For example, an ESF of the invention may comprise the KRAB domain of human zinc finger protein 10 (ZNF10; Szulc, J. et al. (2006) Nat. Methods 3: 109-16). An example sequence of a KRAB domain of human zinc finger protein 10 is:

15 ALSPQHSVAVTQGSIIKNKEGMDAKSLTAWSTRILVTFKDVFDFTREEWKLLDTAQQIVYRNV
MLENYKNLVSLGYQLTKPDVILRLLEKGEEPWLVEREIHQETHPDSETAFEIKSSV

(SEQ ID NO: 1)

Further examples of suitable KRAB domains for use in the invention include:

20 MNNSQGRVTFEDVTVNFTQGEWQRLNPEQRNLYRDVMLENYSNLVSVGQGETTKPDVILRLLE
QGKEPWLEEEVVLGSGRAEKNGDIGGQIWKPKDVKESL

(the KRAB domain of the ZIM3 protein; SEQ ID NO 2)

ITLEDVAVDFTWEEWQLLGAAQKDLYRDVMLENYSNLVAVGYQASKPDALFKLEQGEQLWTI
EDGIHSGACS

(the KRAB domain of the ZNF350 protein; SEQ ID NO: 3)

25 VMFEEVSVCFSTSEEWACLGPIQRALYWDVMLENYGNVTSLEWETMTENEVTSKPSSSQRAD
SHKGTSKRLQG

(the KRAB domain of the ZNF197 protein; SEQ ID NO: 4)

VSFKDVAVDFTQEEWQQLDPEKITYRDVMLENYSHLVSVGYDTTKPNVVIKLEQGEEPWIM
GGEFPCQHSP

30 (the KRAB domain of the RBAK protein; SEQ ID NO: 5)

VKIEDMAVSLILEEWGCQNLARRNLSRDNRQENYGSAFPQGGENRNENEESTSKAETSEDSA
SRGETTGRSQKE

(the KRAB domain of the ZKSCAN1 protein; SEQ ID NO: 6)

LTFKDVFVDFTLEEWQQLDSAQKNLYRDVMLENYSHLVSVGYLVAKPDVIFRLGPGEESWMA
DGGTPVRTCA

(the KRAB domain of the KRBOX4 protein; SEQ ID NO: 7)

5 VTFEDVTLGFTPEEWGLLDLKQKSLYREVMLENYRNLSVVEHQLSKPDVVSQLEEAEDFWPV
ERGIPQDTIP

(the KRAB domain of the ZNF274 protein; SEQ ID NO: 8)

An example nucleotide sequence encoding a KRAB domain is:

ATGGATGCTAAGTCACTGACTGCCTGGTCCCGGACACTGGTGACCTTCAAGGATGTGTTTGT
GGACTTCACCAGGGAGGAGTGGAAGCTGCTGGACACTGCTCAGCAGATCCTGTACAGAAATG
10 TGATGCTGGAGAACTATAAGAACCTGGTTTCCTTGGGTTATCAGCTTACTAAGCCAGATGTG
ATCCTCCGTTGGAGAAGGAGAGAGCCCTGGCTGGTGGAGAGAGAAATTCACCAAGAGAC
CCATCCTGATTGAGAGACTGCATTTGAAATCAAATCATCAGTT

(SEQ ID NO: 9)

DNA methyltransferase (DNMT) domain

15 In some embodiments, the ESF comprises a DNA methyltransferase (DNMT) domain. In
some embodiments, the ESF comprises a DNMT3A domain, a DNMT3B domain and/or a
DNMT1 domain. In some embodiments, the ESF comprises a DNMT3A domain.

An ESF of the invention may, for example, comprise a domain of human DNA
methyltransferase 3A (DNMT3A; Law, J.A. et al. (2010) Nat. Rev. Genet. 11: 204-20),
20 preferably the catalytic domain. An example DNMT3A sequence is:

TYGLLRREDWPSRLQMFFANNHDQEFDPKVPYPPVPAEKRKPIRVLSLFDGIATGLLVLKD
LGIQVDRIYASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIEWGPFDLVIGGSPCNDLS
IVNPARKGLYEGTGRLFFEFYRLLDARPKEGDDRPFFWLFENVVAMGVSDKRDISRFLESN
PVMIDAKEVSAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTITTRSN
25 SIKQGKDQHFPVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQLLGRSWSVPVIRH
LFAPLKEYFACV

(SEQ ID NO: 10)

DNA methyltransferases 3B and 1 (DNMT3B and DNMT1), similarly to DNMT3A, are also
responsible for the deposition and maintenance of DNA methylation, and may also be used
30 in an ATR of the present invention. Example sequences are:

CHGVLRNRKDNVRLQAFFTSDTGLEYEAPKLYPAIPAARRRPIRVLSLFDGIATGYLVLKE
LGIKVGKYVASEVCEESIAVGTVKHEGNIKYVNDVRNITKKNIEEWGPFDLVIGGSPCNDLS
NVNPARKGLYEGTGRLFFEFYHLLNYSRPKEGDDRPFFWMFENVVAMKVGDKRDISRFLECN
PVMIDAIKVSAHRARYFWGNLPGMNRPIASKNDKLELQDCLEYNRIAKLKKVQTITTKSN
35 SIKQGNQLFPVVMNGKEDVLWCTELERIFGFPVHYTDVSNMGRGARQKLLGRSWSVPVIRH
LFAPLKDYFACE

(the catalytic domain of human DNMT3B; SEQ ID NO: 11)

5 MVAELISEEDLEFMKGDTRHLNGEEDAGGREDSSILVNGACSDQSSDSPPILEAIRTPAIRGR
 RSSSRSLSKREVSSLLSYTQDLTGDDGDDGSDTPVMPKLFRETRTRSESPAVRTRNNNSV
 SSRERHRPSRSTRGRQGRNHVDESPVEFPATRSRRLRRATASAGTPWPSPSSYLTIDLTDD
 TEDTHGTPQSSSTPYARLAQDSQQGGMESPOVEADSGDGSSEYQDGKEFGIGDLVWGKIKG
 10 FSWWPAMVVSWKATSKRQAMSGMRWVQWFGDGKFSEVSADKLVALGLFSQHFNLATFNKLVS
 YRKAMYHALEKARVRAGKTFPSSPGDSLEDQLKPMLEWAHGGFKPTGIEGLKPNNTQPENKT
 RRRTADDSATSDYCPAPKRLKTNCYNNGKDRGDEDQSRQMASDVANNKSSLEDGCLSCGRK
 NPVSFHLPLFEGGLCQTCRDRFLELFYMYDDDGYQSYCTVCCEGRELLCSNTSCCRCFCVEC
 15 LEVLVGTGTAAEAKLQEPWSCYMCLPQRCHGVLRRRKDWNVRLQAFFTSDTGLEYEAPKLYP
 AIPAARRRPIRVLSLFDGIATGYLVVLKELGKVGKYVASEVCEESIAGTVKHENIKYVND
 VRNITKKNIEEWGPFDLVIGGSPCNDLSNVNPARKGLYEGTGRLFFEFYHLLNYSRPKEGDD
 RPFWMFENVVAMKVGDKRDISRFLECNPMIDAIVSAHRARYFWGNLPGMNRPVIAASKN
 DKLELQDCLEYNRIAKLKKVQTITTKSNSIKQGKNQLFPVVMNGKEDVLWCTELERIFGFPV
 HYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE
 (DNMT3B: SEQ ID NO: 12)

20 LRTLDVFSGCGGLSEGFHQAGISDTLWAIEMWDPAQAFLNNPGSTVFTEDCNILLKLVMA
 GETTNSRGQRLPQKGDVEMLCGGPPCQGFSGMNRFNSTYSKFKNSLVVSFLSYCDYYRPRF
 FLLENVRFVVSFKRSMVLKLTLRCLVRMGYQCTFGVLQAGQYGVATRRRAIILAAAPGEKL
 PLFPEPLHVFAPRACQLSVVDDKKFVSNITRLSSGPFRTITVRDTMSDLPEVRNGASALEI
 SYNGEPQSWFQRLRGAQYQPIILRDHICKDMSALVAARMRIPLAPGSDWRDLPIEVRSLD
 GTMARKLRYTHHDKNGRSSSGALRGVCSCVEAGKACDPAARQFNTLIPWCLPHTGNRHNHW
 AGLYGRLEWDGFFSTVTNPEPMGKQGRVLHPEQHRVSVRECARSQGFPTYRLFGNILDK
 HRQVGNVAVPPPLAKAIGLEIKLCMLAKARESASAKIEEEAAKD
 (the catalytic domain of human DNMT1; SEQ ID NO: 13)

25 DNMT-like domain

In some embodiments, the ESF comprises a DNMT-like domain. A “DNMT-like” domain refers to a protein which is a member of a DNMT family, but which does not possess DNA methylation activity. The DNMT-like protein typically activates or recruits other epigenetic effector domains.

30 An ESF of the invention may, for example, comprise DNA (cytosine-5)-methyltransferase 3-like (DNMT3L), a catalytically inactive DNA methyltransferase that activates DNMT3A by binding to its catalytic domain. An example DNMT3L sequence is:

35 MAAIPALDPEAEPSMDVILVGSSELSSSVSPGTGRDLIAIEVKANQRNIEDICICCGSLQVH
 TQHPLFEGGICAPCKDKFLDALFLYDDDGYSYCSICCSGETLLICGNPDCTRCYCFECVDS
 LVGPGTSGKVHAMSNNWCYLCLPSSRSGLLQRRRKWRSQKAFYDRESENPLEMFETVPVWR
 RQPVRLSLFEDIKKELTSLGFLSGSDPGQLKHVVVDVTDVTRKDVVEEWGPFDLVYGATPPL
 GHTCDRPPSWYLFQFHRLLQYARPKGPSRPFWMFVDNLVLNKEDLDVASRFLEMEPTIP
 DVHGGSLQNAVVRVWSNIPAIRSRHWALVSEELSLLAQNKQSSKLAAKWPTKLVKNCFLPLR
 EYFKYFSTELTSSL
 40 (SEQ ID NO: 14)

An example nucleotide sequence encoding DNMT3A and DNMT3L domains is:

AACCATGATCAAGAGTTCGATCCTCCTAAAGTGTATCCCCCTGTCCCTGCCGAAAAGCGGAA
 ACCGATCAGGGTTCTCAGTCTCTTCGATGGAATTGCCACAGGCCTCTTGGTCCCTTAAGGACC
 TGGGCATCCAAGTgGAcCGCTACATCGCCTCAGAAGTATGTGAgGACAGCATCACAGTCGGC
 ATGGTGCGCCACCAGGGGAAAATAATGTACGTGCGCGACGTGAGATCAGTTACTCAGAAACA
 5 TATTCAGGAATGGGGGCCTTTTGATCTCGTAATTGGCGGCAGCCCCCTGCAATGATCTGAGTA
 TTGTGAACCCCGCCCGGAAAGGTTTGTACGAAGGTACGGGCAGACTGTTCTTCGAGTTTTAC
 AGACTGTTGCACGACGCACGACCAAAGGAGGGTGAcGAtcGACCCTTCTTCTGGCTGTTCGA
 GAACGTGGTCGCTATGGGCGTGTCTGACAAGCGGGACATTTctAGATTCCCTGGAGAGCAATC
 CAGTGATGATTGATGCAAAGGAAGTAtccGCTGCCCACCGCGCCAGATACTTCTGGGGCAAT
 10 CTGCCCCGGCATGAATCGACCCTTGGCAAGCACCGTGAATGATAAGTTGGAGTTGCAGGAGTG
 TCTgGAGCATGGGCGGATCGCAAAGTTTAGCAAGGTGCGGACCATCACGACCCGAAGCAATT
 CAATTAAGCAGGGAAAAGGACCAACATTTTCCAGTGTTTATGAACGAGAAAAGAGGAcATCCTG
 TGGTGcACAGAAATGGAGCGGGTTTTCGGATTCCCCGTACATTATACTGATGTGTCCAATAT
 GTCTAGGTTGGCTAGACAGAGGCTGtTGGGACGCTCATGGAGTGTTCTGTCTATCCGCCACC
 15 TGTTCTGCTCCTCTGAAGGAGTATTTTCGcTGCGTtTCATCCGGgAATTCAAACGCAAACAGC
 AGAGGCCCATCCTTTTCTTCCGGCCTGGTgCCACTTAGTCTGCGCGGCTCTCACATGGGACC
 TATGGAAATATACAAAACAGTGAGCGCCTGGAAGAGGCAGCCTGTGCGGGTCTTGAGTCTGT
 TCAGAAACATTGACAAGGTGCTTAAGTCCCTTGGATTTCTGGAGTCTGGCAGCGGTTCTGGG
 GGTGgcACCCCTTAAGTATGTGGAGGACGTGACAAACGTGCTGAGGAGAGATGTGGAGAAATG
 20 GGGGCCATTTGACCTGGTATACGGGAGCACTCAACCTCTCGGTTCTTCTTGTGATCGCTGCC
 CAGGGTGGTATATGTTTCAGTTCCATCGcATCCTGCAATACGCCCTTCCGCGGCAGGAGAGT
 CAGCGACCATTTCTTCTGGATATTCATGGACAATCTCCTGCTGACAGAgGACGACCAAGAGAC
 TACGACTAGATTTCTTTCAGACCGAGGCCGTTACTCTCCAGGACGTTAGAGGTAGGGACTACC
 AGAATGCCATGCGAGTGTGGAGTAACATCCCCGGACTCAAGTCAAAGCAcGCACCCCTGACC
 25 CCCAAGGAAGAGGAATACCTcCAGGCGCAGGTGAGGAGCCGAGTAAGCTCGATGCTCCTAA
 AGTCGATTTGTTGGTGAAGAATTGCTTGCTGCCCCTGAGAGAGTACTTTAAGTACTTTAGCC
 AGAACAGCTTGCCACTG

(SEQ ID NO: 15)

Chromo Shadow (CS) domain

30 In some embodiments, the ESF comprises a Chromo Shadow (CS) domain. The CS domain may be a CS domain of CBX5.

An example CS domain sequence is:

MLEPEKIIIGATDSCGDLMLMKWKDTEADLVLAKEANVKCPQIVIAFYEEERLTWHAYPE

(SEQ ID NO: 16)

35 An example nucleotide sequence encoding a CS domain is:

ATGCTGGAACCAGAAAAGATCATTTGGGGCAACAGATTCTGTGGTGATTTAATGTTTCCTAAT
 GAAATGGAAAGACACAGATGAAGCTGACCTGGTTCTTGCAAAGAAGCTAATGTGAAATGTC
 CACAAATTGTGATAGCATTTTATGAAGAGAGACTGACATGGCATGCATATCCTGAG

(SEQ ID NO: 17)

40 YAF2-RYBP (Y-R) domain

In some embodiments, the ESF comprises a YAF2-RYBP (Y-R) domain.

An example Y-R domain sequence is:

MRPRLKNVDRSTAQQQLAVTVGNVTVIITDFKEKTRSSSTSSSTVTSSAGSEQQNQ
(SEQ ID NO: 18)

An example nucleotide sequence encoding a Y-R domain is:

5 ATGAGGCCCCGGCTGAAAAACGTGGACAGGAGCACTGCACAGCAGTTGGCAGTAAGTGTGGG
CAACGTCACCGTCATTATCACAGACTTTAAGGAAAAGACTCGCTCCTCATCGACATCCTCAT
CCACAGTGACCTCCAGTGCAGGGTCAGAACAGCAGAACCAG
(SEQ ID NO: 19)

Further epigenetic effector domains

10 Example sequences of further suitable epigenetic effector domains are:

TF	Sequence	SEQ ID NO
EnR	MALEDRCSPQSAPSPITLQMQLHHLHQQQQQQQQQQMQLHLHL QQQLQLHQQQLAAGVFHHFAMAFDAAAAAAAAAAAAHAHAA ALQQLRLSGSGSPASCSTPASSTPLTIKEESDSVIGDMSFHNQ THTTNEEEAEEDDDIDVDVDDTSAGGRLPPPAHQQQSTAKPS LAFSISNILSDRFGDVQKPGKSMENQASIFRPFEASRSQTATP SAFTRVDLLEFSRQQQAAAAAATAAMMLERANFLNCFNPAAYP RIHEEIVQSRLRRSAANAVIPPPMSSKMSDANPEKSAL	21
MECP2	SPKKKRKVEASVQVKRVLEKSPGKLLVKMPFQASPGGKGE ATTSAQVMVIKRPGRKRKAEADPQAIKKRGRKPGSVVAAAA EAKKKAVKESSIRSVQETVLPKKRKTRETVSIEVKEVVKPLL VSTLGEKSGKGLKTKCKSPGRKSKESSPKGRSSSASSPPKEHH HHHHHAESPKAPMPLLPPEPQSSDPISPPEPQDLSSSI CKEEKMPRAGSLES DGC PKEPKTPQPMVAAAATTTTTTTTVA EKYKHRGEGEKDIVSSSMRPNREEPVDSTPVTERTVS	22
GLI3RD	GTVFAMDPRNGYMEPHYHPPHLPFAFHPVPIDARHHEGRYHY DPSPIPPLHMTSALSSSPTYDLPFIRISPHRNPTAASESPFS PPHPYINPYMDYIRSLHSSPSLSMISATRGLSPTDAPHAGVSP AE	23
MAD1RD	RMNIQMLLEAADYLER	24

Example nucleotide sequences encoding further suitable epigenetic effector domains are:

TF	Sequence	SEQ ID NO
ZIM3-KRAB	ATGAACAATCCCAGGGAAGAGTGACCTTCGAGGATGTCACTG TGAACCTCACCCAGGGGAGTGGCAGCGGCTGAATCCCGAACA GAGAACTTGTACAGGGATGTGATGCTGGAGAATTACAGCAAC CTTGCTCTGTGGGACAAGGGGAAACCACCAACCCGATGTGA TCTTGAGGTTGGAACAAGGAAAGGAGCCATGGTTGGAGGAAGA GGAAGTGCTGGGAAGTGGCCGTGCAGAAAAAATGGGGACATT GGAGGGCAGATTTGGAAGCCAAAGGATGTGAAAGAGAGTCTC	25
EnR	ATGGCCCTGGAGGATCGCTGCAGTCCACAGTCAGCGCCAGCC CCATTACCCTACAAATGCAGCATCTTACCACCAGCAACAGCA GCAGCAGCAACAGCAGCAGCAATGCAGCACCTCCACCAGCTG CAGCAACTGCAGCAGTTGCACCAACAGCAACTGGCCGCCGGTG	26

	TCTTCCACCATCCGGAATGGCCTTCGATGCCGCTGCAGCCGC CGCTGCTGCAGCTGCTGCTGCGGCCGCCACGCTCATGCTGCT GCACTGCAGCAGCGCCTCAGTGGCAGTGGATCGCCGCATCCT GCTCCACGCCCCGCTCGTCCACGCCGCTGACCATCAAGGAGGA GGAAAGCGACTCCGTGATCGGTGACATGAGTTTCCACAATCAG ACGCACACCACCAACGAGGAGGAGGAGGCGGAGGAGGATGACG ACATTGATGTGGATGTGGATGATACGTCGGCGGGCGGACGCCT GCCACCACCCGCCACCAGCAGCAGTCCGACGGCCAAGCCCTCG CTGGCCTTTTCCATCTCCAACATCCTGAGCGATCGTTTCGGAG ATGTCCAGAAGCCGGGCAAGTCGATGGAGAACCAGGCCAGCAT ATTCGCCCCCTTCGAGGCGAGTCGTTCCCAGACTGCCACGCC TCCGCCCTTTACAAGAGTGGATCTGCTGGAGTTTAGCCGGCAAC AGCAGGCTGCCGCCGAGCCGCTACTGCGGCCATGATGCTGGA ACGGGCCAACTTCCTTAACTGCTTCAATCCGGCTGCCTATCCC AGGATACACGAGGAAATCGTGCAGAGTCGGCTGCGCAGGAGTG CAGCCAATGCCGTCATCCCGCCGCCATGAGCTCCAAGATGAG CGATGCCAATCCAGAGAAATCTGCTCTG	
MECP2	AGCCCAAAGAAGAAACGGAAGGTGGAAGCCTCAGTGCAGGTGA AAAGGGTGCTGGAAAAATCCCCGGCAAACTCCTCGTGAAGAT GCCCTTCCAGGCTTCCCCTGGCGGAAAAGGTGAAGGGGTGGC GCAACCACATCTGCCCAGGTCATGGTCATCAAGCGACCTGGAA GGAAAAGAAAGGCCGAGGCTGACCCCTCAGGCCATTCCAAAGAA ACGGGGACGCAAGCCAGGGTCCGTGGTCGCAGCTGCAGCAGCT GAGGCTAAGAAAAAGGCAGTGAAGGAAAGCTCCATCCGCAGTG TGCAGGAGACTGTCCTGCCCATCAAGAAGAGGAAGACTAGGGA GACCGTGTCATCGAGGTCAAAGAAGTGGTCAAGCCCCTGCTC GTGTCCACCCTGGGCGAAAAATCTGGAAAGGGGCTCAAAACAT GCAAGTCACCTGGACGGAAGCAAGGAGTCTAGTCCAAAGGG GCGCTCAAGCTCCGCTTCTAGTCCCCCTAAAAAGGAACACCAT CACCATCACCATCACGCCGAGTCTCCTAAGGCTCCTATGCCAC TGCTCCCACCACCTCCACCACCTGAGCCACAGTCAAGCGAAGA CCCCATCAGCCCACCCGAGCCTCAGGATCTGTCTCTAGTATT TGCAAAGAGGAAAAGATGCCCAGAGCAGGCAGCCTGGAGAGTG ATGGCTGTCCAAAAGAACCCGCCAAGACCCAGCCTATGGTGGC AGCCGCTGCAACTACCACCACAACCACAACCTACCACAGTGGCC GAAAAATACAAGCATCGCGCGAGGGCGAACGAAAGGACATTG TGTCAAGCTCCATGCCCAGACCTAACCGGGAGGAACAGTCTGA TAGTAGGACACCCGTGACTGAGAGAGTCTCA	27
GLI3RD	GGGACGGTGTTTGCCATGGACCCAGGAATGGTTACATGGAGC CCCCTACCACCCTCCTCATCTTTTCCCTGCCTTCCATCCTCC TGTACCAATTGATGCCAGACATCATGAGGGCCGTTACCATTAC GATCCATCTCCGATTCTCCATTGCATATGACTTCCGCCTTAT CTAGTAGCCCTACGTATCCGGACCTGCCCTTCATTAGGATCTC CCCACACCGGAACCCCACTGCTGCTTCCGAGTCTCCCTTCAGC CCTCCACATCCCTACATTAATCCCTACATGGACTATATCCGCT CCTTGACAGCAGCCCATCGCTCTCCATGATCTCAGCAACCCG TGGGCTGAGCCCTACAGATGCGCCCCATGCAGGAGTCAGCCCA GCAGAA	28
MAD1RD	CGGATGAACATCCAGATGCTGCTGGAGGCGGCCGACTATCTGG AGCGG	29

The ESF of the invention may, for example, comprise an amino acid sequence that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identity to any one of SEQ ID

NOs: 1-8, 10-14, 16, 18 or 21-24, preferably wherein the amino acid sequence substantially retains the natural function of the protein represented by SEQ ID NO: 1-8, 10-14, 16, 18 or 21-24, respectively.

5 The ESF of the invention may, for example, be encoded by a polynucleotide comprising a nucleic acid sequence which encodes a protein that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid identity to any one of SEQ ID NOs: 1-8, 10-14, 16, 18 or 21-24, preferably wherein the amino acid sequence substantially retains the natural function of the protein represented by SEQ ID NO: 1-8, 10-14, 16, 18 or 21-24, respectively.

10 The ESF of the invention may, for example, be encoded by a polynucleotide comprising a nucleic acid sequence that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% nucleotide identity to any one of SEQ ID NOs: 9, 15, 17, 19 or 25-29, preferably wherein the encoded amino acid sequence substantially retains the natural function of the protein encoded by SEQ ID NO: 9, 15, 17, 19 or 25-29, respectively.

TRANSCRIPTION FACTOR

15 A transcription factor (TF) may control the transcription of DNA through binding to certain DNA sequences. Transcription factors typically function to regulate genes to control transcription and/or expression, for example dependent on cell type and timing. Groups of transcription factors may work in a coordinated manner to direct cell division, cell growth and cell death; cell migration and organization during embryonic development; and in response
20 to extracellular signals.

Transcription factors comprise at least one DNA-binding domain, which may target the transcription factor to certain sequences to direct their regulatory function.

The polypeptides and ESFs of the invention comprise at least one transcription factor DNA-binding domain. The skilled person is readily able to identify DNA-binding domains from
25 transcription factors using well known methods, for example using sequence comparison tools and/or databases.

The polypeptides and ESFs of the invention may comprise a minimum transcription factor sequence that retains its function in binding to DNA. However, it is preferred that as much transcription factor sequence as possible is retained in the polypeptides and ESFs of the
30 invention, without adversely impacting the function of the ESF in decreasing transcription and/or expression. Without wishing to be bound by theory, transcription factor sequence in

addition to the DNA-binding domain may enable recruitment of additional factors within a cell.

Unmodified transcription factors may comprise an activation domain (AD; also referred to as a trans-activation domain), which may function to activate gene transcription and/or expression. The polypeptides and ESFs of the invention preferably do not comprise a functional activation domain. The transcription factor sequence may be modified (e.g. mutated or truncated) to disrupt activation domain function. The transcription factor sequence incorporated into the polypeptide or ESF may lack an activation domain.

In preferred embodiments, the ESF does not comprise a functional transcription factor activation domain. The ESF may, for example, not comprise a transcription factor activation domain. The ESF may, for example, comprise a fragment of a transcription factor that lacks a functional transcription activation domain, and which comprises a functional DNA-binding domain.

In preferred embodiments, the transcription factor is an oncogenic transcription factor.

The term "oncogenic transcription factor" as used herein refers to a transcription factor that may transform a healthy cell into a cancer cell, for example through causation of inappropriate gene expression patterns, which can for example promote tumor initiation and progression.

In some embodiments, the transcription factor is a cancer-associated transcription factor.

The term "cancer-associated transcription factor" as used herein refers to a transcription factor that may induce tumorigenic properties, but cannot transform a healthy cell into cancer cell.

In some embodiments, the transcription factor is selected from the group consisting of SOX2, MYC, MYCN, TEAD1, TEAD2, TEAD3, TEAD4, FOXA1, FOXA2, ELK1, ELK3, ELK4, SRF, FOXM1, FOXC1, FOXC2, TWIST1, SALL4, ELF1, HIF1A, SOX9, SOX12, SOX18, ETS1, PAX3, PAX8, GLI1, GLI2, GLI3, ETV1, ETV2, ETV3, RUNX1, RUNX2, RUNX3, MAFB, TFAP2C and E2F1.

In some embodiments, the transcription factor is SOX2. The ESF may comprise a fragment of human SOX2 consisting of amino acids 1-179.

In some embodiments, the transcription factor is MYC. The ESF may comprise a fragment of human MYC consisting of amino acids 144-454.

In some embodiments, the transcription factor is TEAD1. The ESF may comprise a fragment of human TEAD1 consisting of amino acids 1-166.

Example sequences comprising suitable transcription factor DNA-binding domains are:

TF	Sequence	SEQ ID NO
SOX2	MYNMMETELKPPGPQQTSGGGGGNSTAAAAGGNQKNSPDRVKR PMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETE KRPFIDEAKRLRALHMKEHPDYKYRPRRKTCTLMKKDKYTLP GLLAPGGNSMASGVGVGAGLGAGVNRMDSYAHMNGWSNGSYS MMQDQLG	30
MYC	IIQDCMWSGFSAAAKLVSEKLASYQAARKDSGSPNPARGHSVC STSSLYLQDLASAASECIDPSVVFYPLNDSSSPKSCASQDSS AFSPSSDLSLSTESSPQGSPEPLVLHEETPPTTSSDSEEEQE DEEIDVVSVEKRQAPGKRSESGSPSAGGHSKPPHSPPLVKRC HVSTHQHNYAAPPSTRKDYPAAKRVKLDSDRVLRQISNNRKCT SPRSSDTEENVKRRTHNVLERQRRNELKRSFFALRDQIPELEN NEKAPKVILKKATAYILSVQAEEQKLISEEDLLRKRREQLKH KLEQLRNSCA	31
MYCN	MPSCSTSTMPGMICKNPDLEFDSLQPCFYDDEDDFYFGGPDST PPGEDIWKKFELLPTPPLSPSRGFAEHSSEPPSWVTEMLLENE LWGSPEEDAFGLGGLGGLTPNPVILQDCMWSGFSAREKLERA VSEKLQHGRGPPTAGSTAQSPGAGAAAPAGRGHGAAGAGRAG AALPAELAHPAECVDPVAVFPFVFNKREPAPVPAAPASAPAA GPAVASGAGIAAPAGAPGVAPPRPGGRQTSGGDHKALSTSGED TLSDSDDEDEDEDEDEDEDEIDVVTVEKRRSSSNTKAVTTFTITV RPNAAALGPGRAQSSELILKRCLPIHQHNYAAPSYPVESEDA PPQKKIKSEASPRPLKSVIPKAKSLSPRNSDSEDSERRRNHN ILERQRRNDLRSSFLTLRDHVPPELVKNEKAAKVILKKATEYV HSLQAEEHQLLLEKEKLQARQQQLLKKIEHARTC	32
TEAD1	IEPSSWSGSESPAENMERMSDSADKPIDNDAEGVWSPDIEQSF QEALAIYPPCGRRKIILSDEGKMYGRNELIARYIKLRTGKTRT RKQVSSHQVLRARRKSRDFHSLKDQTAQDKALQHMAAMSSAQ IVSATAIHNLGLPGIIPRPTFPAGPFWPGMIQTGQP	33
TEAD2	MGEPRAGAALDDGSGWTGSEEGSEEGTGGSEGAGGDGGPDAEG VWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRNELIARY IKLRTGKTRTRKQVSSHQVLRARRKSREIQSKLDQVSKDKAF QTMATMSSAQLISAPSLQAKLGPTGPGQASELFQFWSGSGPPW NVPDVKPFSTPFTLSLTPPSTDLPGYEPPQALSPLPPPTPSP PAWQARGLTARLQLVEFSFAFVEPPDAVDSYQRHLFVHISQHC PSPGAPPLESVDVRQIYDKFPEKKGGLRELYDRGPPHAFFLVK FWADLNWGPSGEEAGAGGSISSGGFYGVSSQYESLEHMTLTCS SKVCSFGKQVVEKVETERAQLEDGRFVYRLLRSPMCEYLVNFL HKLRLQLPERYMMNSVLENFTILQVVTNRDTQELLCTAYVFEV STSERGAQHIIYRLVRD	34
TEAD3	IASNSWNASSSPGEAREDGPEGLDKGLDNDAGVWSPDIEQSF QEALAIYPPCGRRKIILSDEGKMYGRNELIARYIKLRTGKTRT RKQVSSHQVLRARRKVVREYQVGIKAMNLDQVSKDKALQSMAS SSAQIVSASVLQNKFSPPSPLPQAVFSTSSRFWSSPPLLGQOP GPSQDIKPFAPAYPIQPPLPPTLSSYEPLAPLPSAAASVPVW QDRTIASSRLRLLEYSAFMEVQRDPDTYSKHLFVHIGQTNPAF SDPPLEAVDVRQIYDKFPEKKGGLKELYEKGPNAFFLVKFWA DLNSTIQEGPGAFYGVSSQYSSADSMTISVSTKVCSFGKQVVE KVETEYARLENGRFVYRIHRSPMCEYMINFIHKLKHLPEKYMM	35

	NSVLENFTILQVVTSRDSQETLLVIAFVFEVSTSEHGAQHVVY KLVKD	
TEAD4	LEGTAGTITSNEWSSPTSPEGSTASGGSQALDKPIDNDAEGVW SPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRNELIARIYIK LRTGKTRTRKQVSSHQVLARRKAREIQAKLKFWQGALPGQAG TSHDVKPFSSQQTAYAVQPPLPLPGFESPAGPAPSPSAPPAPPWQ GRSVASSKLWMLFEFSAFLEQQQDPDTYNKHLFVHIGQSSPSYS DPYLEAVDIRQIYDKFPEKKGGLKDLFERGPSNAFFLVKFWAD LNTNIEDEGSSFYGVSSQYESPENMIITCSTKVCSEFGKQVVEK VETERYARYENGHYSYRIHRSPICEYMINFIHKLKHLPEKYMNN SVLENFTILQVVTNRDTQETLLCIAVFEVSASEHGAQHVIYR LVKE	36
FOXA1	MLGTVKMEGHETSDWNSYADTQEAYSVPVSNMNSGLGSMNS MNTYMTNTMTTSGNMTPASFNMSYANPGLGAGLSPGAVAGMP GGSAGAMNSMTAAGVTAMGTALSPSGMGAMGAQQAASMNGLGP YAAAMNPCMSPMAYAPSNLGRSRAGGGGDAKTFKRSYPHAKPP YSYISLITMAIQQAPSKMLTLSEIYQWIMDLFPYRQNRQQRWQ NSIRHSLSFNDCFVKVARSPDKPGKGSYWTLPDPSGNMFENG CYLRRQKRFKCEKQPGAGGGGGSGSGSGAKGGPESRKDPGAS NPSADSPLHRGVHGKTGQLEGAPAPGPAASPQTLDHSGATATG GASELKTASSTAPPISSGPGALASVPASHPAHGLAPHESQLH LKGDPHYSFNHPFSINNLMSSSEQQHKLDKAYEQALQYSPYG STLPASLPLGSASVTTRSPIEPSALEPAYYQGVYSRPLVNTS	37
FOXA2	MHSASSMLGAVKMEGHEPDSWSSYYAEPEGYSSVSNMNAAGLM NGMNTYMSMSAAAAMGSGSGNMSAGSMNMSSYVGAGMSPSLAGM SPGAGAMAGMGGSAGAAGVAGMGPPLSPSLPLGGQAAGAMGG LAPYANMNSMSPMYGQAGLSRARDPKTYRRSYTHAKPPYSYIS LITMAIQQSPNKMLTLSEIYQWIMDLFPYRQNRQQRWQNSIRH SLSFNDCFLKVPRSPDKPGKGSFWTLHPDPSGNMFENGCYLRRQ KRFKCEKQLALKEAAGAAGSGKAAAGAAQASQAQLGEAAGPAS ETPAGTESPHSSASPCQEHKRGGLGELKGTAAAALSPPEPAPS PGQQQQAHAHLLGPPHHPGLPPEAHLKPEHHYAFNHPFSINN LMSSSEQQHHHSHHHHQPHKMDLKAYEQVMHYPGYGSPMPGSLAM GPVTNKTGLDASPLAADTSYYQGVYSRPLMNSS	38
ELK1	MDPSVTWQFLLQLLREQNGHIIISWTSRDGGEFKLVDAEEVA RLWGLRKNKTNMNYDKLSRALRYYYDKNIIKRVSGQKFVYKFV SYPEVAGCSTEDCPPQPEVSVTSTMPNVAPAAIHAAPGDTVSG KPGTPKGAGMAGPGGLARSSRNEYMRSGLYSTFTIQSLQPQPP PHPRPAVVLPSAAPAGAAAPPSGSRSTSPSPLEACLEAEAEAGL PLQVILTPPEAPNLKSEELNVEPGLGRALPPEVKVEGPKEELE VAGERGFVPETTKAEPEVPPQEGVPARLPVVMMDTAGQAGGHA ASSPEISQPQKGRKPRDLELPLSPSLGGPGPERTPGSGSGSG LQAPGPALTPSLLPHTLTPVLLTPSSLPPSIHFWSTLSPIAP RSPAKLSFQFPSSGSAQVHIPSISVDGLSTPVVLSPPGPQKP	39
ELK3	MESAITLWQFLLQLLDQKHEHLICWTSNDGEFKLLKAEVAK LWGLRKNKTNMNYDKLSRALRYYYDKNIIKRVIGQKFVYKFVS FPEILKMDPHAVEISRESLLLQSDCKASPEGREAHKHGLAAL RSTSRNEYIHSGLYSSFTINSLQNPPDAFKAIKTEKLEPPED SPPVEEVRTVIRFVTNKTDKHVTRPVVSLPSTSEAAAAASAFLA SSVSAKISSMLPNAASISSASPFSSRSPSLSPNSPLPSEHRS LFLEAACHSDSLEPLNLSSGSKTKSPSLPPKAKKPKGLEISA PPLVLSGTDIGSIALNSPALPSGSLTPAFFTAQTPNGLLLTPS PLLSSIHFWSSLSPVAPLSPARLQGPSTLFQFPTLLNGHMPVP IPSLDRAASPVLLSSNSQKS	40
ELK4	MDSAITLWQFLLQLLQKQPNKHMICWTSNDGQFKLLQAEVAVAR	41

	LWGIKRNKPNMNYDKLSRALRYYYVKNIIKKVNGQKFVYKFVS YPEILNMDPMTVGRIEGDCESLNFSEVSSSSKDVENGKDKPP QPGAKTSSRNDYIHSGLYSSFTLNSLNSSNVKLFKLIK TENPA EKLAEEKSPQEPTPSVIKFVTTPSKKPPVEPVAATISIGPSIS PSSEETIQALETLVSPKLPSLEAPTSASNVMTAFATTPPISSI PPLQEPPRTPSPPLSSHDPIDIDTIDSVASQPMELPENLSLEPK DQDSVLLEKDKVNNSRSRKKPKGLELAPTLVITSSDPSPLGIL SPSLPTASLTPAFFSQTPPIILTPSPLLSSIHFWSTLSPVAPLS PARLQGANTLFQFPVLSNSHGPF TLSGLDGPSTPGPFSPDLQK T	
SRE	MLPTQAGAAAALGRGSALGGSINRTPTGRPGGGGGTRGANGGR VPGNGAGLGPGRLEREAAAAAATTAPTAGALYSGSEGDSESG EEEEELGAERRGLKRSLSEMEIGMVVGGPEASAAATGGYGPVSG AVSGAKPGKKTRGRVKIKMEFIDNKLRRYTTFSKRKTGIMKKA YELSTLTGTQVLLLVASETGHVYTFATRKLQPMITSETGKALI QTCLNSPDSPPRSDPTTDQMSATGFEETDLTYQVSESDSSGE TKDTLKPFTVTNLP GTTSTIQTAPSTSTTMQVSSGSPFPITN YLAPVSASVSPSAVSSANGTVLKSTGSGPVSSGGLMQLPTSFT LMPGGAVAQQVPVQAIQVHQAPQQASPSRDSSTDLTQTSSSGT VTLPATIMTSSVPTTVGGHMYPSPHAVMYAPTSGLGDGSLTV LNAFSQAPSTMQVSHSQVQEPGGVPQVFLTASSGTVQIPVSAV QLHQMAVIGQQAGSSSNLTQLQVNLDTAHSTKSE	42
FOXMI	MKTSPPRRPLILKRRRLPLPVQNA PSETSEEPKRSPAQQESNQ AEASKEVAESNSCKFPAGIKIINHPTMPNTQVVAIPNNANIHS IITALTAKGKESGSSGPNKFILISCGGAPTQPPGLRPQTQTSY DAKRTEVTLET LGPKPAARDVNLPRPPGALCEQKRETCADGEA AGCTINNSLSNIQWLRKMSSDGLGSRSIKQEMEKENCHLEQR QVKVEEPSRPSASWQNSVSEPPSYMMAMIQFAINSTERKRMT LKDIYTWIEDHFPYFKHIAKPGWKNSIRHNLSLHDMFVRETS NGKVSFWTIHPSANRYLTLDQVFKPLDPGSPQLPEHLESQQR PNPELRRNMTIKTELPLGARRKMKPLLPRVSSYLVPIQFPVNQ SLVLQPSVKVPLPLAASLMSELARHSKRVR IAPKVFGQVVF GYMSKFFSGDLRDFGTPITSLFNFI FLCLSVLLAEEGIAPLSS AGPGKEEKLLFGEGFSPLLPVQTIKEEEIQPGEEMPHLARPIK VESPPLEEWSPAPSFKEESSHWEDSSQSPTPRPKKSYSGLR SPTRCVSEMLVIQHRERRERSRSRRKQHLLPPCVDEPELLFSE GPSTSRWAAELPFPADSSDPASQLSYSQEVGGPFKTPIKETLP ISSTPSKSVLPRTPESWRLTPPAKVGGGLDFSPVQTSQGASDPL PDPLGLMDLSTTPLQSAPPLESPQRLLSSEPLDLISVPFGNSS PSDIDVPKPGSPEPQVSGLAANRSLTEGLVLDTMNDSL SKILL DISFPGLDEDPLGPDNINWSQFIPELQ	43
FOXC1	MQARYSVSSPNSLGVPYLGGEQSYRAAAAAAGGGYTAMPAP MSVYSHPAHAEQYPGGMARAYGPYTPQPQPKDMVKPPYSYIAL ITMAIQNAPDKKITLNGIYQFIMDRFPFYRDNKQGWQNSIRHN LSLNECFVKVPRDDKKPGKGSYWTLDPD SYNMFENG SFLRRR RFLKKDAVKDKKEEDRLHLKEPPPPGRQPPAPPEQADGNAPG PQPPPVRIQDIKTENGTCPSPPQPLSPAALGSGSAAAVPKIE SPDSSSSSLSSGSSPPGSLPSARPLSLDGADSAPPPAPSPAPP PHHSQGFSDNIMTSLRGSPQSAAAELSSGLLASAAASSRAGI APPLALGAYSPGQSSLYSSPCSQTSSAGSSGGGGGAGAAGGA GGAGTYHCNLQAMSLYAAGERGGHLQAGPGAGGS AVDDPLPD YSLPPVTSSSSSSLSHGGGGGGGGGQ EAGHHPAAHQGR LTSW YLNQAGGDLGHLASAAAAAAGYPGQQQNFHVSREMFE SQRI GLNNSPVNGNSSCQMAFPSSQSLYRTSGAFVYDCSKF	44
FOXC2	MQARYSVSDPNALGVVPYLSEQNYRAAGSYGGMASPMGVYSG	45

	HPEQYSAGMGRSYAPYHHHQPAAPKDLVKPPYSYIALITMAIQ NAPEKKITLNGIYQFIMDRFPFYRENKQGWQNSIRHNLNLNEC FVKVPRDDKKPGKGSYWTLDPSYNTMFENGSLRRRRRFFKKKD VSKEKEERAHLKEPPPAASKGAPATPHLADAPKEAEKKVVIKS EAASPALPVITKVETLSPESALQGSPRSAASTPAGSPDGLPE HHAAAPNGLPGFSVENIMTLRTSPPGGELSPGAGRAGLVVPPL ALPYAAAPPAAYGQPCAQGLEAGAAGGYQCSMRAMSLYTGAER PAHMCVPPALDEALSDHPSGPTSPLSALNLAAGQEGALAATGH HHQHGHGHHHPQAPPPPPAPQPQPTPQFGAAAAQAASWYLNHSG DLNHLPGHTFAAQQTTFPNVREMFSNHLGIENSTLGESQVSG NASCQLPYRSTPPLYRHAAPYSYDCTKY	
TWIST1	MMQDVSSSPVSPADDSLSNSEEPDRQQPPSGKRGGRRKRSSR RSAGGGAGPGGAAGGGVGGDEPGSPAQGKRGKKSAGCGGGGG AGGGGGSSSGGGSPQSYEELQTQRMANVRERQRTQSLNEAFA ALRKIIPTLPSDKLSKIQTCLKLAARYIDFLYQVLQSDDELDSKM ASC SYVAHERLSYAFSVWRMEGAWSMSASH	46
HIF1A	MEGAGGANDKKKISSERRKEKSRDAARSRRSKESEVFYELAHQ LPLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLIEDDMKA QMNC FYLKALDGFVMVLTDDGDMIIISDNVNKYMGLTQFELTG HSVDFDTHPCDHEEMREMLTHRNGLVKKGKEQNTQRSFFLRMK CTLTSRGRMTNIIKSATWKVLHCTGHIHVDYDTSNQPQCQYKKP PMTCLVLICEPIPHPSNIEIPLDSKTFLSRHSMDMKFSYCDER ITELMGYEPEELLGRSIEYEHALDSDLTKTHHDMFTKGQVT TGQYRMLAKRGGYVWVETQATVIYNTKNSQPQCIVCVNYVVS IIQHDLIFSLQQTECVLKPVESSDMKMTQLFTKVESEDTSLSF DKLKKEPDALTLAPAAAGDTIISLDGFSNDTETDDQQLLEEVL YNDVMLPSPNEKLQNINLAMSPLPTAETPKPLRSSADPALNQE VALKLEPNPESLELSFTMPQIQDQTPSPSDGSTRQSSPEPNP SEYCFYVSDSMVNEFKLELVEKLEFAEDTEAKNPFSTQDLDL EMLAPYIPMDDDFQLRSFDQLSPLESSSASPESASPQSTVTVF QQTQIQEPTANATTTTATTTDELKTVTKDRMEDIKILIASPSPT HIHKETTSATSSPYRDTQSRTASPNRAGKGVIEQTEKSHPRSP NVLSVALSQRRTVPEEELNPKILALQNAQRKRKMEHDGSLFQA VGII	47
SALL4	MSRRKQAKPQHINSEEDQGEQQPQQQTPEFADAAPAAPAAGEL GAPVNHGPNDEVASEDEATVKRLRREETHVCEKCCAEFFSISE FLEHKKNCTKNPPVIMNDSEGPVPSSEDFSGAVLSHQPTSPGS KDCHRENGGSSSEDMKEKPDASVYVLTETALPPTPQDISYLA KGKVANTNVTLQALRGTKVAVNQRSADALPAPVPGANSIPWVL EQILCLQQQLQQIQLTEQIRIQVNMWASHALHSSGAGADTLK TLGSHMSQQVSAVALLSQKAGSQGLSLDALKQAKLPHANIPS ATSSLSPGLAPFTLKPDPGTRVLPNVMSRLPSALLPQAPGSVLF QSPFSTVALDTSKKGKGPKNISAVDVKPKDEAALYKHKCKYC SKVFGTDSSLQIHLRSHTGERPFVCSVCGRHFTTKGNLKVHFFH RHPQVKANPQLFAEFQDKVAAGNGIPYALSVPDPIDEPSLSLD SKPVLVTTSVGLPQNLSSTGTPNPKDLTGGSPLPGDLQPGPSPSE GGPTLPVGPNYNSPRAGGFQGSSTPEPGSETLKLQQLVENID KATTDPNELICHRVLSQSSSLKMHYRTHTGERPFQCKICGRA FSTKGNLKTGLGVHRTNTSIKTQHSCPICQKKFTNAVMLQQHI RMHMGQIPNTPLPENPCDFTGSEPMTVGENGSTGAICHDDVI ESIDVEEVSSQEAPSSSSKVPTPLPSIHSASPTLGFAMMASLD APGKVGPPAPFNLQRQGSRENGSVESDGLTNDSSSLMGDQEQS RSPDILETTSFQALSPANSQAESIKSKSPDAGSKAESSENSRT EMEGRSSLPTFIRAPPTYVKVEVPPTFVGPSTLSPGMTPLLA AQPRRQAKQHGCRTCGKNFSSASALQIHERHTHTEKPFVCNIC	48

	GRAFTTKGNLKVHYMTHGANNNSARRGRKLAIENTMALLGTDG KRVSEIFPKEILAPSVNVDPVVWNQYTSMLNGGLAVKTNEISV IQSGGVPTLPVSLGATSVVNNATVSKMDGSQSGISADVEKPSA TDGVPKHQFPHFLEENKIAVS	
ELF1	MAAVVQQNDLVFEFASNVMEDERQLGDPAlFFAVIVEHVPGAD ILNSYAGLACVEEPNDMITESSLDVAEEEEIIDDDDDITLTVE ASCHDGETIETIEAAEALLNMDSPGPMLDEKRINNNIFSSPE DDMVVAPVTHVSVTLDGIPVEMETQQVQEKYADSPGASSPEQP KRKKGRKTKPPRPDSPATTPNISVKKKNKDGKNTIYLWEFL ALLQDKATCPKYIKWTQREKGIKFLVDSKAVSRLWGKHKNKPD MNYETMGRALRYYYQRGILAKVEGQRLVYQFKEMPKDLIYIND EDPSSSIESSDPSLSSSATSNRNQTSRSRVSSSPGVKGGATT LKPGNSKAAKPKDPVEVAQPPSEVLRTVQPTQSPYPTQLFRTVH VVQPVQAVPEGEAARTSTMQDETINSSVQSIRTIQAPTQVPVV VSPRNQQLHTVTLQTVPLTTVIASTDPSAGTGSQKFIQAIPS SQPMTVLKENVMLQSQKAGSPPSIVLGPAQVQVLTSSNVQTIC NGTVSVASSPSFSATAPVVTFSRSSLVAHPPGTVITSVIKT QETKTLTQVEKKESEDHLKENTEKTEQQPQPYVMVSSSNGF TSQVAMKQNELLEPNSF	49
SOX9	MNLLDPFMKMTDEQEKGSLGAPSPMTSEDSAGSPCPSGSGSDT ENTRPQENTFPKGEPDLKKESEEDKFPVCIREAVSQVLKGYDW TLVPMFVRVNGSSKNKPHVKRPMNAFMVWAQAARRKLADQYPH LHNAELSKTLGKLWRLNESEKRPFVEEAERLRVQHKKDHPDY KYQPRRRKSVKNGQAEAEATEQTHISPNAIFKALQADSPHSS SGMSEVHSPGEHSGSQSGPPTPPTTPKTDVQPGKADLKREGR LPEGGRQPPIDFRDVIDIGELSSDVISNIETFDVNEFDQYLPPN GHPGVPATHGQVITYTGSYGISSTAATPASAGHVWMSKQQAAPP PPQPPQAPPAPQAPPQQAAPPQQAAPPQQAHTLTTLSS EPGQSQRTHIKTEQLSPSHYSEQQHSFPQQIAYSFPNLPHYSP SYPPITRSQYDYTDHQNSSSYSHAAGQGTGLYSTFTYMNPAQ RPMYTPIDATSGVPSIPQTHSPQHWEQPVYTQLTRP	50
SOX12	MVQQRGARAKRDGGPPPPGPGPAEEGAREPGWCKTPSGHIKRP MNAFMVWSQHERRKIMDQWPDHNAEISKRLGRRWQLLDSEK IPFVREAERLRRLKHMADYPDYKYRPRKSKGAPAKARPRPPG SGGGSRLKPGPQLPGRGGRRAAGGPLGGGAAAPEDDDDDDEE LLEVRLEVETPGRELWRMPAGRAARGQAERAQGPSGEGAAAAA AASPTPSEDEPEEEEEEEEEAAAEEGEETVASGEESLGLFLSRL PPGPAGLDCSALDRDPDLQPPSGTSHFEFPDYCTPEVTEMIAG DWRPSSIADLVFTY	51
SOX18	MQRSPPGYGAQDDPPARRDCAWAPGHGAAADTRGLAAGPAALA APAAPASPPSPQRSPPRSPEPGRYGLSPAGRGERQAADSRIR RPMNAFMVWAKDERKRLAQNPDLHNAVL SKMLGKAWKELNAA EKRPFVEEAERLRVQHLDHPNYKYRPRRKKQARKARLEPGL LLPGLAPPQPPPEPFAASGSARAFRELPLGAEFDGLGLPTP ERSPLDGLPEGAFFPPPAAPEDCALRPFRAPIAPTELSRDP GGCYGAPLAEALRTAPPAAPLAGLYYGTGTGPGYPGPLSPP EAPPLESAEPLGPAADLWADVLDLTFDQYLNCSTRTPDAPGLP YHVALAKLGPRAMSCPEESSLISALSDASSAVVYSACISG	52
ETS1	MKAAVDLKPTLTIIKTEKVDLELFPSPDMECADVPLLTSSKE MMSQALKATFSGFTKEQQRLGIPKDPQWTEHVRDWMWAVN EFLKGVDFQKFCMNGAALCALGKDCFLELAPDFVGDILWEHL EILQKEDVKPYQVNGVNPAYPESRYTSDYFISYGIEHAQCVPP SEFSEPSFITESYQTLHPISSEELLSLKYENDYPSVILRDPLQ TDTLQNDYFAIKQEVVTPDNMCMGRTSRGKLGQDSFESIESY DSCDRLTQSWSSQSSFNSLQRVPSYDSFDSSEDYPAALPNHKPK	53

	GTFKDYVRDRADLNKDKPVI PAAALAGYTGSGPIQLWQFLLEL LTDKSCQSFI SWTGDGWEFKLSDPDEVARRWGKRKNKPKMNYE KLSRGLRYYYDKNIIHKTAGKRYVYRFVCDLQSLGTYTPEELH AMLDVKPDADE	
PAX3	MTTLAGAVPRMMRPGPGQNYPRSGFPLEVSTPLGQGRVNQLGG VFINGRPLPNHIRHKIVEMAHHGIRPCVISRQLRVSHGCVSKI LCRYQETGSIRPGAIGGSKPKQVTPDVEKKIEEYKRENPGMF SWEIRDKLLKDAVCDRNTVPSVSSISRILRSKFGKEEEEADL ERKEAEESEKKAKHSIDGILSERGKALVSGVSSH	54
PAX8	MPHNSIRSGHGGLNQLGGAFVNGRPLPEVVRQRIVDLAHQGVR PCDISRQLRVSHGCVSKILGRYYETGSIRPGVIGGSKPKVATP KVVEKIGDYKRQNPMTFAWEIRDRLAEGVCDNDTVPSVSSIN RIIRTKVQQPFNLPMDSCVATKSLSPGHTLIPSSAVTPPESPQ SDSLGSTYSINGLLGIAQPGSDKRKMDDSDQDSCRLSIDSQSS SSGPRKHLRTDAFSQHHLEPLECPFERQHYPEAYASPSHTKGE QGLYPLPLLNSTLDDGKATLTPSNTPLGRNLSTHQTYPVVADP HSPFAIKQETPEVSSSSSTPSSLSSSAFLDLQQVSGVPPFNA FPHAASVYGQFTGQALLSGREMGPTLPGYPPHIPTSGQGSYA SSAIAGMVAGSEYSGNAYGHTPYSSYSEAWRFPNSSLLSSPY YSSTSRSAPPTTATAFDHL	55
GLI1	MFNSMTPPPISSYGEPCCLRLPLSQGAPSVGTEGLSGPPFCHQ ANLMGPHSYGPARETNSCTEGPLFSSPRSAVKLTKKRALSI PLSDASLDLQTVIRTSPLSLVAFINSRCTSPGGSYGHLSIGTM SPSLGFPAQMNHQKGPSPSFGVQPCGPHDSARGGMI PHPQSRG PFPTCQLKSELDMLVGKCREEPLEGDMSSPNSTGIQDPLLGM DGREDLEREEKREPESVYETDCRWDGCSQEFDSQEQLVHHINS EHIHGERKEFVCHWGGCSRELRFKAQYMLVVMHRRHTGEKPH KCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCEHEGCSKAFSN ASDRAKHQNRTHSNEKPYVCKLPGCTKRYTDPSSLRKHVKTVH GPDAHVTKRHRGDGPLPRAPSI STVEPKREREGGPIREESRLT VPEGAMKPPSPGAQSSCSDHSPAGSAANTDSGVEMTGNAGG STEDLSSLDEGPCIAGTGLSTLRRLENLRDLQLHQLRPIGTRG LKLPSLSHTGTTVSRRVGPPVSLERRSSSSSSISSAYTVSRRS SLASFPFGSPENGASSLPGLMPAQHYLLRARYASARGGTS PTAASSLDRIGGLPMPWPWSRAEYPGYNPNAGVTRRASDPAQA ADRPAPARVQRFKSLGCVHTPPTVAGGGQNFDPYLPYSPQ PPSITENAAMDARGLQEEPEVGTSMVGSGLNPYMDFPPTDTLG YGGPEGAAAEPYGARGPGSLPLGPGPPTNYGNPCPQQASYPD PTQETWGEFSPSHGLYPGPKALGGTYSQCPRLEHYQVQVKPE QGCVPVSDSTGLAPCLNAHPSEGPPHPQPLFSHYQPSPQYL QSGPYTQPPPDYLPSEPRCLDFDSPHSTGQLKAQLVCNYVQ SQQELLWEGGGREDAPAQEPSYQSPKFLGGSQVSPSRAKAPVN TYGPGFGPNLPHKSGSYPTPSPCHENFVVGANRASHRAAAP RLLPLPLPTCYGPLKVGGTNPSCGHPEVGRLLGGGALYPPPEGQ VCNPLDSLDDNTQLDFVAILDEPQGLSPPPSHDQRGSSGHTP PPSGPPNMAVGNMSVLLRSLPGETEFLNSSA	56
GLI2	METSASATASEKQEAESGILEAAGFPDPGKKASPLVVAASAAA AVAAQGV PQHLLPPFHAPLPIDMRHQEGRYHYEPHSVHG VHG PALSGSPVISDISLIRLSHPAGPGESPFNAPHYVNPMEHY LRSVHSSPTLSMISAARGLSPADVAQEHLKERGLFGLPAPGTT PSDYYHQMTLVAGHPAPYGDLLMQSGGAASAPHLHDYLNPDV SRFSSPRVTPRLSRKRALSISPLSDASLDLQRMIRTSPLSLVA YINNSRSSAASGSYGHLSAGALSPAFTFPHPINPVAYQQILS QQRGLGSAFGHTPPLIQPSPTFLAQQPMALTSINATPTQLSSS SNCLSDTNQNKQSSSAVSSTVNPVAIHKRSKVKTEPEGLRPA	57

	SPLALTQGGVSGHGSCGCALPLSQEQLADLKEDLDRDDCKQEA EVVIYETNCHWEDCTKEYDTQEQLVHHINNEHIHGEKKEFVCR WQACTREQKPFKAQYMLVVHMRRTGEKPHKCTFEGCSKAYS LENLKTHTLSHTGEKPYVCEHEGCNKAFSNASDRAKHQNRTHS NEKPYICKIPGCTKRYTDPSSLRKHVKTVHGPDHAVTTKKQRND VHLRTPLLKENGDEAGTEPGGPESTEASSTSQAVEDCLHVRA IKTESSGLCQSSPGAQSSCSSEPSPPLGSAPNNDSGVEMPGTGP GSLGDLTALDDTPPGADTSALAAPSAGGLQLRKHMTTMHRFEQ LKKEKLKSLKSDSCSWAGTPHTRNTKLPLPGSGSILENFSGS GGGGPAGLLPNPRLSELSASEVTMLSQLEERRDSSTSTVSSAY TVSRRSSGISPYFSSRRSSEASPLGAGRPHNASSADSYDPIST DASRRSSEASQCSGSGLLNLTPAQQYSLRAKYAAATGGPPPT PLPGLERMSLRTRLALLDAPERTLPAGCPRPLGPRRSDGPTY GHGHAGAAPAFPHAPGGGARRASDPVRRPDALSLPRVQRFHS THNVNPGPLPPCADRRGLRLQSHPSTDGGLARGAYSPPPSIS ENVAMEAVAAGVDGAGPEADLGLPEDDLVLPDDVVQYIKAHAS GALDEGTGQVYPTESTGFSNPRLPSPGLHGQRRMVAADSNVG PSAPMLGGCQLGFGAPSSLNKNMPVQWNEVSSGTVDALASQV KPPFPQGNLAVVQKPAFGQYPGYSPQGLQASPGGLDSTQPH LQPRSGAPSQGIQPRVNYMQQLRQPVAGSQCPGMTTMTSPHACY GQVHPQLSPSTISGALNQFPQSCSNMPAKPGHLGHPQQTEVAP DPTTMGNRHRELGVPSALAGVPPHPVQSYPPQSHHLAASMS QEGYHQVPSLLPARQPGFMEPQTGPMGVATAGFGLVQPRPPL PSPTGRHRGVRAVQQQLAYARATGHAMAAMPSSQETAEAEPKG AMGNMGSVPPQPPQDAGGAPDHSMYYYYGQIHMYEQDGGLEN LGSCQVMRSQPPQACQDSIQPQLPSPGVNQVSSTVDSQLL EAPQIDFDAIMDDGDHSSLFSGALSPSLHLSQNSSRLTTPR NSLTLPSPISPAISNMAVGDMSSMLTSLAESKFLNMMT	
GLI3	MEAQSHSSTTTTEKKKVENSIVKCSTRTDVSEKAVASSTTSNED ESPGQTYHRERRNAITMQPQNVQGLSKVSEEPSTSSDERASLI KKEIHGSLPHVAEPSVPYRGTVFAMDPRNGYMEPHYHPPHLP AFHPPVPIDARHHEGRYHYDPSPIPPLHMTSALSSSPTYPDLP FIRISPHRNPTAASESPFSPHPYINPYMDYIRSLHSSPSLSM ISATRGLSPTDAPHAGVSPAEEYHQMALLTGQRSPYADIIPSA ATAGTAIHMEYLHAMDSTRFSSPRLSARPSRKRTLISPLSD HSFDLQTMIRTSPNSLVTILNNSRSSSSASGSYGHLASASISP ALSFTYSSAPVSLMHQQILSRQQSLGSAFGHSPPLIHPAPT PTQRPIPIPTVLPVQVSSGPSESSQNKPTSESAVSSTGDPM HNKRSKIKPDEDLSPSGARGQQEQPEGTTLVKEEGDKDESKQE PEVIYETNCHWEGCAREFDTQEQLVHHINNDHIHGEKKEFVCR WLDCSREQKPFKAQYMLVVHMRRTGEKPHKCTFEGCTKAYS LENLKTHTLSHTGEKPYVCEHEGCNKAFSNASDRAKHQNRTHS NEKPYVCKIPGCTKRYTDPSSLRKHVKTVHGPEAHVTTKKQRGD IHRPPPPRDSGSHSQSRSPGRPTQALGEQQDLNNTTSKREE CLQVKTVKAEKPMTSQSPGGQSSCQQSPISNYSNNGLELP LTDGGSIGDLSAIDETPIMDSTISTATTALALQARRNPAGTKW MEHVKLERLKQVNGMFPRLNPILPPKAPAVSPLIGNGTQSNNT CSLGGPMTLLPGRSDLSGVDVTMLNMLNRRDSSASTISSAYLS SRRSSGISPCFSSRRSSEASQAEGRPQNVSVADSYDPISTDAS RRSSEASQSDGLPSLLSLTPAQQYRLKAKYAAATGGPPPTPLP NMERMSLKTRLALLGDALEPGVALPPVHAPRRCSGGAHGYGR RHLQPHDAPGHGVRRASDPVRTGSEGLALPRVPRFSSLSSCNP PAMATSAEKRSVLQNYTRPEGGQSRNFHSSPCPPSITENVTL ESLTMDADANLNDEDFLPDDVVQYLNQSQAGYEQHFPSALPD DSKVPHGPGDFDAPGLPDSHAGQQFHAEQPCPEGSKTDLPIQ	58

	WNEVSSGSADLSSSKLKCGPRPAVPQTRAFCNGMVVHPQNP LRSGPAGGYQTLGENSNPYGGPEHMLHNSPGSGTSGNAFHEQ PCKAPQYGNCLNRQPVAPGALDGACGAGIQASKLKSTPMQSG GQLNFGLPVAPNESAGSMVNGMQNDPVGQGYLAHQLLGDSMQ HPGAGRPGQMLGQISATSHINIYQGPESCLPGAHGMSQPSS LAVVRGYQPCASFGGSRRQAMPRLSLALQSGQLSDTSQTCRVN GIKMEMKGQPHPLCSNLQNYSGQFYDQTVGFSQQDTKAGSFSI SDASCLLQGTSAKNSELLSPGANQVTSTVDSLSDSHDLEGVQID FDAIIDDGDHSSLSMSGALSPSIIQNLSSHSSSRLTTPRASLPFP ALSMSTTNMAIGDMSLLTSLAESKFLAVMQ	
ETV1	MDGFYDQQVPYMTNSQRGRNCNEKPTNVRKRKFINRDLAHS EELFQDLSQLQETWLAEVAFHGLPLKIKKEPHSPCSEISSACS QEPPFKFSYGEKCLYNVSAYDQKPQVGMRPSNPPTPSSTPVSP LHHASPNSTHTPKPDRAFFPAHLPPSQSIPDSSYPMDHRFRRL SEPCNSFPPLPTMPREGRPYQRMSEPNIFFPPQGFKEQYHD PVYEHNTMVGSAASQSFPPLMIKQEPDFAYDSEVPSCHSIY MRQEGFLAHPSTRTEGCMFEKGPRQFYDDTCVVPEKFDGDIKQ PGMYREGPTYQRRGSLQLWQFLVALLDDPSNSHFIAWTGRGME FKLIEPEEVARRWGIQKNRPAMNYDKLSRSLRYYYEKGIMQKV AGERYVYKFVCDPEALFMAFPDNRPLKTDMERHINEEDTV PLSHFDESMAYMEGGCCNPHPYNEGYVY	59
ETV2	MDLWNWDEASPQEVPPGNKLAGLEGAKLGFCFPDLALQGDTP ATAETCWKGPIQLWQFLLELLHDGARSSCIRWTGNSREFQLCD PKEVARLWGERKRKPGMNYEKLRLRYYYRRDIVRKSGRKY TYRFGGRVPSLAYPDCAGGGRGAETQ	60
ETV3	MKAGCSIVEKPEGGGGYQFPDWAYKTESSPGSRQIQLWHFILE LLQKEEFRHVIWQQGEYGEFVIKDPDEVARLWGRKCKPQMN YDKLSRALRYYYNKRILHKTGKRFETYKFENFKLVMPNYPFIN IRSSGKIQTLLVGN	61
RUNX1	MASDSIFESFPSYPQCFMRECILGMNPSRDVHDASTSRRFTPP STALSPGKMSEALPLGAPDAGAALAGKLRSRSDRSMVEVLADHP GELVRTDSPNFLCSVLPTHWRCNKTLPIAFKVVALGDVPDGT VTVMAGNDENYSAELRNATAAMKNQVARFNDLRFVGRSGRGKS FTLTITVFTNPPQVATYHRAIKITVDGPREPRRHRQKLDDQTK PGSLSFSERLSELEQLRRTAMRVSPHHPAPTNPASLNHSTA FNPQPQSQMDTRQIQSPSPWSYDQSYQYLGSIASPSVHPATP ISPGRASGMTLSAELSSRLSTAPDLTAFSDPRQFPALPSISD PRMHYPGAFTYSPTPVTSGIGIGMSAMGSATRYHTYLPPPYPG SSQAQGGPFQASSPSYHLYYGASAGSYQFSMVGGERSPPRILP PCTNASTGSALLNPSLPNQSDVVEAEGSHSNSPTNMAPSARLE EAVWRPY	62
RUNX2	MASNSLFSTVTPCQQNFWDPTSTRRFSPSSSLQPGKMSDVS PVVAAQQQQQQQQQQQQQQQQQQQEEAAAAAAAAAAAAA AAAVPRLRPPHDNRTMVEI IADHPAELVRTDSPNFLCSVLPSH WRCNKTLPVAFKVVALGEVPDGTVVTVMAGNDENYSAELRNAS AVMKNQVARFNDLRFVGRSGRGKSFTLTITVFTNPPQVATYHR AIKVTVDGPREPRRHRQKLDDSKPSLFSDRSLDLGRIPHPSMR VGVPQNPRLNSAPSPFNPGQSQITDPRQAQSSPPWSYDQ SYPSYLSQMTSPSIHSTPLSSTRGTGLPAITDVPRRISGASE LGPFS DPRQFPSISLTERFSNPRMHYPATFTYTPPVTSGMS LGMSATTHYHTYLPPPYPGSSQSQSGPFQTSSTPYLYGTSSG SYQFPMVPGGDRSPSRMLPPCTTTSNGSTLLNPNLPNQNDGVD ADGSHSSSPTVLNSSGRMDES VWRPY	63
RUNX3	MRIPVDPTSTRRFTPPSPAFCGGGGGKMGENSGALSAQA AVG PGGRARPEVRSMVDVLADHAGELVRTDSPNFLCSVLPSHWRCN	64

	KTLPVAFKVVVALGDVPDGTVVTVMAGNDENYSAELRNASAVMK NQVARFNDLRFVGRSGRGKSFTLTITVFTNPTQVATYHRAIKV TVDGPREPRRHRQKLEDQTKPFPDRFGDLERLRMRVTPSTPSP RGSLSSTSHFSSQPQTPIQGTSELNPFSDPRQFDRSFPTLPTL TESRFPDPRMHYPGAMSAAFPYSATPSGTSISSLSVAGMPATS RFHHTYLPPYPGAPQNNQSGPFQANPSPYHLYYGTSSGSYQFS MVAGSSSGDRSPTRMLASCTSSAASVAAGNLMNP SLGGQSDG VEADGSHSNSPTALSTPGRMDEAVWRPY	
MAFB	MAAELSMGPELPTSPLAMEYVNDFDLLKFDVKKEPLGRAERPG RPCTRLQPAGSVSSTPLSTPCSSVPSSPSFSPTQKTHLEDLY WMASNYQQMNPEALNLTPEDAVEALIGSHVPVQPLQSFDSFRG AHHHHHHHHPHPHHAYPGAGVAHDELGPHAHPHHHHHHHQASPP PSSAASPAQQLPTSHPGPGPHATASATAAGNGSVEDRFSDDQ LVMSVRELNRHLRGFTKDEVIRLKQKRRTLKNRGYAQSCRYK RVQQKHLENEKTQLIQQVEQLKQEVSRRLARERDAYKVKCEKL ANSGFREAGSTSDSPSSPEFFL	65
TFAP2C	MLWKITDNVKEYEEDCEDRHDGSSNGNPRVPHLSSAGQHLYSPA PPLSHTGVAEYQPPPYFPPPYQQLAYSQSADPYSHLGEAYAAA INPLHQAPATGSQQQAWPGRQSQEGAGLP SHHGRPAGLLPHLS GLEAGAVSARRDAYRRSDLLLPHAHALDAAGLAENLGLHDMPH QMDEVQNVDDQHLLLDQTVIRKGPISMTKNPLNLPCKELVG AVMNPTEVFCSVPGRSLSSSTSKYKVTVAEVQRRLSPPECLN ASLLGGVLRRAKSKNGRSLREKLDKIGLNLPAGRRKAAHVTL LTSLVEGEAVHLARDFAYVCEAEFPSKPVAEYLT RPHLGGRNE MAARKNMLLAAQQLCKEFTELLSDQRTPHGTSRLAPVLETNIQ NCLSHFSLITHGFGSQAICAAVSALQNYIKEALIVIDKSYMNP GDQSPADSNKTLEKMEKHRK	66
E2F1	MALAGAPAGGPCAPALEALLGAGALRLLDSSQIVIIISAAQDAS APPAPTGPAAAPAGPCDPDLLLFATPQAPRPTPSAPRPALGRP PVKRRLDLETDHQYLAESSGPARGRGRHPGKGVKSPGEKSRYE TSLNLTTRKFLELLSHSADGVVDLNWAAEVLKVQKRRIYDITN VLEGIQLIAKSKNHIQWLGSHTTVGVGGRLEGLTQDLRQLQE SEQQLDHLNICTTQLRLLEDTSQRLAYVTCQDLRSIADPA EQMVMVIKAPPETQLQAVDSENFOISLKSQGPIDVFLCPEE TVGGISPGKTPSQEVTSEENRATDSATIVSPPSSPPSSLT DPSQSLLSLEQEPLLSRMGSLRAPVDEDRLSPLVAADSLEHV REDFSGLLPEEFISLSPHEALDYHFGLEEGEGIRDLFDCDFG DLTPLDF	67

Example nucleotide sequences encoding polypeptides comprising transcription factor DNA-binding domains are:

TF	Sequence	SEQ ID NO
SOX2	ATGTACAACATGATGGAGACGGAGCTGAAGCCGCCGGGCCCGC AGCAAACCTTCGGGGGGCGGCGGCGCAACTCCACCGCGGCGGC GGCCGGCGGCAACCAAGAAAAACAGCCCGGACCGCGTCAAGCGG CCCATGAATGCCTTCATGGTGTGGTCCCGCGGCGAGCGGCGCA AGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAG CAAGCGCCTGGGCGCCGAGTGGAACTTTTGTGCGAGACGGAG AAGCGGCCGTTCATCGACGAGGCTAAGCGGCTGCGAGCGCTGC ACATGAAGGAGCACCCGATTATAAATACCGGCCCCGGCGGAA AACCAAGACGCTCATGAAGAAGGATAAGTACACGCTGCCCGGC GGGCTGCTGGCCCCCGGCGCAATAGCATGGCGAGCGGGTTCG GGGTGGGCGCCGGCCTGGGCGCGGGCGTGAACAGCGCATGGA	68

	CAGTTACGCGCACATGAACGGCTGGAGCAACGGCAGCTACAGC ATGATGCAGGACCAGCTGGGC	
MYC	ATCATCCAGGACTGTATGTGGAGCGGCTTCTCGGCCGCCGCA AGCTCGTCTCAGAGAAGCTGGCCTCCTACCAGGCTGCGCGCAA AGACAGCGGCAGCCCGAACCCGCGCGCCACAGCGTCTGC TCCACCTCCAGCTTGTACCTGCAGGATCTGAGCGCCGCCGCT CAGAGTGCATCGACCCCTCGGTGGTCTTCCCCTACCCTCTCAA CGACAGCAGCTCGCCCAAGTCCCTGCGCCTCGCAAGACTCCAGC GCCTTCTCTCCGTCCCTCGGATTCTCTGCTCTCCTCGACGGAGT CCTCCCCGAGGGCAGCCCCGAGCCCCTGGTGCTCCATGAGGA GACACCGCCACCACCAGCAGCGACTCTGAGGAGGAACAAGAA GATGAGGAAGAAATCGATGTTGTTTCTGTGGAAGAGGAGG CTCCTGGCAAAAGGTGAGAGTCTGGATCACCTTCTGCTGGAGG CCACAGCAAACCTCCTCACAGCCCACTGGTCTCAAGAGGTGC CACGTCTCCACACATCAGCACAACTACGCAGCGCCTCCCTCCA CTCGGAAGGACTATCCTGCTGCCAAGAGGGTCAAGTTGGACAG TGTCAGAGTCCCTGAGACAGATCAGCAACAACCGAAAATGCACC AGCCCCAGGTCCCTCGGACACCGAGGAGAATGTCAAGAGGCGAA CACACAACGTCTTGGAGCGCCAGAGGAGGAACGAGCTAAAACG GAGCTTTTTTGCCCTGCGTGACCAGATCCCGGAGTTGGAAAAC AATGAAAAGGCCCCCAAGGTAGTTATCCTTAAAAAAGCCACAG CATACATCCTGTCCGTCCAAGCAGAGGAGCAAAAGCTCATTTT TGAAGAGGACTTGTGCGGAAACGACGAGAACAGTTGAAACAC AAACTTGAACAGCTACGGAACCTTTGTGCG	69
MYCN	ATGCCGAGCTGCTCCACGTCCACCATGCCGGGCATGATCTGCA AGAACCCAGACCTCGAGTTTGACTCGCTACAGCCCTGCTTCTA CCCGGACGAAGATGACTTCTACTTTCGGCGGCCCGACTCGACC CCCCCGGGGAGGACATCTGGAAGAAGTTTGAGCTGCTGCCCA CGCCCCCGTGTGCGCCAGCCGTGGCTTCGCGGAGCACAGCTC CGAGCCCCGAGCTGGGTACGGAGATGCTGCTTGAGAACGAG CTGTGGGGCAGCCCGGCCGAGGAGGACGCGTTTCGGCCTGGGG GACTGGGTGGCCTCACCCCCAACCCGGTCATCCTCCAGGACTG CATGTGGAGCGGCTTCTCCGCGCGGAGAAGCTGGAGCGCGCC GTGAGCGAGAAGCTGCAGCACGGCCGCGGGCCGCCAACCGCCG GTTCCACCGCCAGTCCCCGGGAGCCGCGCGCCGAGCCCTGC GGGTGCGGGCACGGCGGGGCTGCGGGAGCCGCGCGCGGG GCCGCCCTGCCCGCCGAGCTCGCCCAACCGGCCGCGAGTGCG TGGATCCCGCCGTGGTCTTCCCCTTTCCCGTGAACAAGCGCGA GCCAGCGCCCGTGCCCGCAGCCCCGGCCAGTGCCCCGGCGGCG GGCCCTGCGGTGCGCTCGGGGGCGGGTATTGCCGCCCCAGCCG GGGCCCCGGGGTTCGCCCTCCGCGCCAGGCGGCCGCCAGAC CAGCGGCGGCGACCACAAGGCCCTCAGTACCTCCGGAGAGGAC ACCCTGAGCGATTGAGATGATGAAGATGATGAAGAGGAAGATG AAGAGGAAGAAATCGACGTGGTCACTGTGGAGAAGCGGCGTTT CTCCTCCAACACCAAGGCTGTCAACACATTACCATCACTGTG CGTCCCAAGAACGCAGCCCTGGGTCCCGGGAGGGCTCAGTCCA GCGAGCTGATCCTCAAACGATGCCTTCCCATCCACCAGCAGCA CAACTATGCCGCCCCCTCTCCCTACGTGGAGAGTGAGGATGCA CCCCACAGAAGAAGATAAAGAGCGAGGCGTCCCCACGTCCGC TCAAGAGTGTCATCCCCCAAAGGCTAAGAGCTTGAGCCCCG AAACTCTGACTCGGAGGACAGTGAGCGTCGCAGAAACCACAAC ATCCTGGAGCGCCAGCGCCGAACGACCTTCGGTCCAGCTTTC TCACGCTCAGGGACCACGTGCCGGAGTTGGTAAAGAATGAGAA GGCCGCCAAGGTGGTCATTTTGAAAAAGGCCACTGAGTATGTC CACTCCCTCCAGGCCGAGGAGCACCAGCTTTTGCTGGAAGG	70

	AAAAATTGCAGGCAAGACAGCAGCAGTTGCTAAAGAAAATTGA ACACGCTCGGACTTGC	
TEAD1	ATTGAGCCCAGCAGCTGGAGCGGCAGTGAGAGCCCTGCCGAAA ACATGGAAAGGATGAGTGACTCTGCAGATAAGCCAATTGACAA TGATGCAGAAGGGGTCTGGAGCCCCGACATCGAGCAAAGCTTT CAGGAGGCCCTGGCTATCTATCCACCATGTGGGAGGAGGAAAA TCATCTTATCAGACGAAGGCAAAATGTATGGTAGGAATGAATT GATAGCCAGATACATCAAACCTCAGGACAGGCAAGACGAGGACC AGAAAACAGGTGTCTAGTCACATTCAGGTTCTTGCCAGAAGGA AATCTCGTGATTTTCATTCCAAGCTAAAGGATCAGACTGCAAA GGATAAGGCCCTGCAGCACATGGCGGCCATGTCTCAGCCCAG ATCGTCTCGGCCACTGCCATTATAACAAGCTGGGGCTGCCTG GGATTCCACGCCCGACCTTCCCAGGGGCGCCGGGGTTCTGGCC GGGAATGATTCAAACAGGGCAGCCA	71
TEAD2	ATGGGGGAACCCCGGGCTGGGGCCGCCCTGGACGATGGCAGCG GCTGGACGGGCAGTGAGGAAGGCAGTGAGGAGGGTACCGGCGG CAGTGAGGGGGCTGGGGGTGACGGGGGCCCCGGATGCAGAGGGG GTGTGGAGCCCAGACATTGAGCAGAGCTTCCAGGAGGCCCTGG CCATCTATCCACCTGCGGCCGCCGAAAATAATTTTGTCTGA TGAAGGCAAGATGTATGGTCGGAATGAACTGATCGCCCGCTAC ATCAAGCTGAGAACGGGGAAGACCCGAACTCGAAAACAGGTTT CTAGTCACATCCAGGTTTTGGCCCGAAGGAAATCAAGGGAAT CCAGTCCAAGTTGAAGGACCAGGTTTCCAAGGACAAGGCTTTC CAGACAATGGCAACCATGTCTCTGCCCAGCTCATCTCCGCGC CTTCTCTGCAGGGCAAACCTGGGTCCCCTGGTCTCAGGCCTC TGAGCTTTTCCAGTTTTGGTCTGGAGGATCTGGGCCCCCTGG AATGTTCCAGATGTGAAGCCATTCTCACAGACACCGTTCACCT TGTCAGTACTCCCCATCTACTGACCTCCCAGGGTACGAGCC CCCCCAAGCCCTCTCACCCTGCCCCACCTACCCCATCGCCC CCAGCCTGGCAGGCTCGGGGCCTGGGCACCGCCCGTTGCAGC TGGTAGAGTTCTCAGCCTTCGTGGAACCGCCAGATGCAGTTGA TTCTTACCAGAGGCACCTGTTCTGTGCACATCAGCCAGCACTGC CCCAGCCCCGGAGCGCCGCGCTCGAGAGTGTGGACGTCCGGC AGATCTACGACAAATCCCTGAGAAAAAGGGTGGCCTCCGAGA GCTATATGATCGTGGCCCCCCCCATGCCTTCTTCTGGTCAAG TTCTGGGCGGACCTGAACTGGGGCCCCAAGTGGTGAGGAGGCAG GGGCCGGTGGCAGCATCAGCAGTGGTGGCTTCTACGGAGTGAG CAGCCAGTATGAGAGCCTGGAACACATGACCCTCACCTGTTCC TCCAAGGTCTGCTCTTTTGGCAAGCAGGTGGTGGAGAAGGTGG AGACGGAACGGGCCAGCTGGAGGACGGCAGATTTGTGTACCG CCTGCTGCGCTCGCCCATGTGCGAGTACCTGGTGAATTTCTTG CACAAGTTGCGGCAGCTGCCTGAGCGATACATGATGAACAGCG TCCTGGAAAACCTCACCATCCTCCAGGTGGTGACAAACAGAGA CACCAGGAACGTGCTGCTCTGCACCGCCTATGTCTTCGAGGTC TCCACCAGCGAGCGTGGGGCCAGCATCACATTTACCGCCTGG TCAGGGAC	72
TEAD3	ATAGCGTCCAACAGCTGGAACGCCAGCAGCAGCCCCGGGGAGG CCCGGGAGGATGGGCCCCAGGGCCTGGACAAGGGGCTGGACAA CGATGCGGAGGGCGTGTGGAGCCCCGACATCGAGCAGAGCTTC CAGGAGGCCCTGGCCATCTACCCGCCCTGCGGCCGGCGGAAGA TCATCCTGTGACAGAGGGCAAGATGTACGGCCGAAATGAGTT GATTGCACGCTATATTAACTGAGGACGGGGAAGACTCGGACG AGAAAACAGGTGTCCAGCCACATACAGGTTCTAGCTCGGAAGA AGGTGCGGGAGTACCAGGTTGGCATCAAGGCCATGAACCTGGA CCAGGTCTCCAAGGACAAAGCCCTTCAGAGCATGGCGTCCATG	73

	TCCTCTGCCCAGATCGTCTCTGCCAGTGTCTGCAGAACAAAGT TCAGCCCACCTTCCCCCTCTGCCCCAGGCCGTCTTCTCCACTTC CTCGCGGTCTTGAGCAGCCCCCTCTCTGGGACAGCAGCCT GGACCTCTCAGGACATCAAGCCCTTTGCACAGCCAGCCTACC CCATCCAGCCGCCCTGCCGCCGACGCTCAGCAGTTATGAGCC CCTGGCCCCGCTCCCCCTCAGCTGCTGCCTCTGTGCCTGTGTGG CAGGACCGTACCATTGCCTCCTCCCGGCTGCGGCTCCTGGAGT ATTACAGCCTTCATGGAGGTGCAGCGAGACCTGACACGTACAG CAAACACCTGTTTGTGCACATCGGCCAGACGAACCCCGCCTTC TCAGACCCACCCCTGGAGGCAGTAGATGTGCGCCAGATCTATG ACAAATTCCCCGAGAAAAAGGGAGGATTGAAGGAGCTCTATGA GAGGGGGCCCCCTAATGCCCTTCTTCCCTTGTCAAGTTCTGGGCC GACCTCAACAGCACCATCCAGGAGGGCCCGGGAGCCTTCTATG GGGTACGCTCTCAGTACAGCTCTGCTGATAGCATGACCATCAG CGTCTCCACCAAGGTGTGCTCCTTTGGCAAACAGGTGGTAGAG AAGGTGGAGACTGAGTATGCCAGGCTGGAGAACGGGCGCTTTG TGTACCGTATCCACCGCTCGCCCATGTGCGAGTACATGATCAA CTTCATCCACAAGCTGAAGCACCTGCCCCGAGAAGTACATGATG AACAGCGTGCTGGAGAACTTCACCATCCTGCAGGTGGTCACGA GCCGGGACTCCCAGGAGACCTGCTTGTGCTTTTGTCTT CGAAGTCTCCACAGTGAGCACGGGGCCCAGCACCATGTCTAC AAGCTCGTCAAAGAC	
TEAD4	TTGGAGGGCACGGCCGGCACCATTAACCTCCAACGAGTGGAGCT CTCCACCTCCCCCTGAGGGGAGCACCGCCTCTGGGGGAGTCA GGCACTGGACAAGCCCATCGACAATGACGCAGAGGGCGTGTGG AGCCCGGATATTGAGCAGAGTTTCCAGGAGGCCCTCGCCATCT ACCCGCCCTGTGGCAGGCGCAAAATCATCCTGTGCGACGAGGG CAAGATGTATGGTCGGAACGAGCTGATTGCCCCGTACATCAAG CTCCGGACAGGGAAGACCCGCAACAGGAAGCAGGTCTCCAGCC ACATCCAGGTGCTGGCTCGTCGCAAAGCTCGCGAGATCCAGGC CAAGCTAAAGTTTGGCAAGGAGCTTTGCCAGGCCAAGCCGGA ACGTCCCATGATGTGAAGCCTTTCTCTCAGCAAACCTATGCTG TCCAGCCTCCGCTGCCTCTGCCAGGGTTTGAGTCTCCTGCAGG GCCCCCCCCATCGCCCTCTGCGCCCCCGGCACCCCATGGCAG GGCCGAGCGTGGCCAGCTCCAAGCTCTGGATGTTGGAGTTCT CTGCCTTCCTGGAGCAGCAGCAGGACCCGGACACGTACAACAA GCACCTGTTCTGTGCACATTGGCCAGTCCAGCCCAAGCTACAGC GACCCCTACCTCGAAGCCGTGGACATCCGCCAAATCTATGACA AATTCCCGGAGAAAAAGGTGGACTCAAGGATCTCTTCGAACG GGGACCCCTCCAATGCCTTTTTTCTTGTGAAGTTCTGGGCAGAC CTCAACACCAACATCGAGGATGAAGGCAGCTCCTTCTATGGGG TCTCCAGCCAGTATGAGAGCCCCGAGAACATGATCATCACCTG CTCCACGAAGGTCTGCTCTTTCGGCAAGCAGGTGGTGGAGAAA GTTGAGACAGAGTATGCTCGCTATGAGAATGGACACTACTCTT ACCGCATCCACCGTCCCCGCTCTGTGAGTACATGATCAACTT CATCCACAAGCTCAAGCACCTCCCTGAGAAGTACATGATGAAC AGCGTGCTGGAGAACTTCACCATCCTGCAGGTGGTCACCAACA GAGACACACAGGAGACCTTGCTGTGCATTGCCTATGTCTTTGA GGTGTGAGCCAGTGAGCACGGGGCTCAGCACCATCTACAGG CTGGTGAAAGAA	74
FOXA1	ATGTTAGGAACTGTGAAGATGGAAGGGCATGAAACCAGCGACT GGAACAGCTACTACGCAGACACGCAGGAGGCCTACTCCTCCGT CCCGGTGAGCAACATGAACTCAGGCCTGGGCTCCATGAACTCC ATGAACACCTACATGACCATGAACACCATGACTACGAGCGGCA ACATGACCCCGGCGTCCCTCAACATGTCTATGCCAACCCGGG	75

	CCTAGGGGGCCGGCCTGAGTCCCGGCGCAGTAGCCGGCATGCCG GGGGGCTCGGGCGGGCGCCATGAACAGCATGACTGCGGCCGGCG TGACGGCCATGGGTACGGCGCTGAGCCCGAGCGGCATGGGCGC CATGGGTGCGCAGCAGGCGGCCTCCATGAATGGCCTGGGCCCC TACGCGGCCGCCATGAACCCGTGCATGAGCCCCATGGCGTACG CGCCGTCCAACCTGGGCCGCGCCGCGCGGGCGGGCGGGCGGA CGCCAAGACGTTCAAGCGCAGCTACCCGCGACGCCAAGCCGCC TACTCGTACATCTCGCTCATCACCATGGCCATCCAGCAGGCGC CCAGCAAGATGCTCACGCTGAGCGAGATCTACAGTGGATCAT GGACCTCTTCCCCATTACCGGCAGAACCAGCAGCGCTGGCAG AACTCCATCCGCCACTCGCTGTCCTTCAATGACTGCTTCGTCA AGGTGGCAGCTCCCCGGACAAGCCGGGCAAGGGCTCCTACTG GACGCTGCACCCGGACTCCGGCAACATGTTGAGAACGGCTGC TACTTGCGCCGCCAGAAGCGCTTCAAGTGCGAGAAGCAGCCGG GGGCCGGCGGGGGGGCGGGAGCGGAAGCGGGGGCAGCGGCGC CAAGGGCGGCCCTGAGAGCCGCAAGGACCCCTCTGGCGCCTCT AACCCAGCGCCGACTCGCCCCCTCCATCGGGGTGTGCACGGGA AGACCGGCCAGCTAGAGGGCGCGCCGGCCCCCGGGCCCGCCGC CAGCCCCCAGACTCTGGACCACAGTGGGGCGACGGCGACAGGG GGCGCCTCGGAGTTGAAGACTCCAGCCTCCTCAACTGCGCCCC CCATAAGCTCCGGGCCCGGGCGCTGGCCTCTGTGCCCGCCTC TCACCCGGCACACGGCTTGGCACCCACAGTCCCAGCTGCAC CTGAAAGGGGACCCCCACTACTCCTTCAACCACCGTTCTCCA TCAACAACCTCATGTCTCCTCGGAGCAGCAGCATAAGCTGGA CTTCAAGGCATACGAACAGGCACTGCAATACTCGCCTTACGGC TCTACGTTGCCCCGCCAGCCTGCCTCTAGGCAGCGCCTCGGTGA CCACCAGGAGCCCCATCGAGCCCTCAGCCCTGGAGCCGGCGTA CTACCAAGGTGTGTATTCCAGACCCGTCTTAAACACTTCCTAG	
FOXA2	ATGCACTCGGCTTCCAGTATGCTGGGAGCGGTGAAGATGGAAG GGCACAGAGCCGTCCGACTGGAGCAGCTACTATGCAGAGCCCGA GGGCTACTCCTCCGTGAGCAACATGAACGCCGGCCTGGGGATG AACGGCATGAACACGTACATGAGCATGTGGCGGGCCGCCATGG GCAGCGGCTCGGGCAACATGAGCGCGGGCTCCATGAACATGTC TCGTGTACGTGGGCGCTGGCATGAGCCCGTCCCTGGCGGGGATG TCCCCCGGCGCGGGCGCCATGGCGGGCATGGGCGGCTCGGCCG GGGCGGCCGGCGTGGCGGGCATGGGGCCGCACTTGAGTCCAG CCTGAGCCCGCTCGGGGGGAGGCGGCCGGGGCCATGGGCGGC CTGGCCCCCTACGCCAACATGAACTCCATGAGCCCCATGTACG GGCAGGCGGGCCTGAGCCGCGCCCGGACCCCAAGACCTACAG GCGCAGCTACACGCACGCAAAGCCGCCCTACTCGTACATCTCG CTCATCACCATGGCCATCCAGCAGAGCCCCAACAAGATGCTGA CGCTGAGCGAGATCTACCAGTGGATCATGGACCTCTTCCCCTT CTACCGGCAGAACCAGCAGCGCTGGCAGAACTCCATCCGCCAC TCGCTCTCCTTCAACGACTGTTTCTGAAGGTGCCCCGCTCGC CCGACAAGCCCGGCAAGGGCTCCTTCTGGACCCTGCACCCTGA CTCGGGCAACATGTTTCGAGAACGGCTGCTACCTGCGCCGCGAG AAGCGCTTCAAGTGCGAGAAGCAGCTGGCGCTGAAGGAGGCCG CAGGCGCCGCCGGCAGCGGCAAGAAGGCGGCCGCCGGAGCCCA GGCCTCACAGGCTCAACTCGGGGAGGCCGCCGGGCCGGCCTCC GAGACTCCGGCGGGCACCAGTGCCTCACTCGAGCGCCTCCC CGTGCCAGGAGCACAAGCGAGGGGGCCTGGGAGAGCTGAAGGG GACGCCGGCTGCGGCGCTGAGCCCCCAGAGCCGGCGCCCTCT CCCGGGCAGCAGCAGCAGGCCCGCGGCCACCTGCTGGGCCCGC CCCACCACCGGGCCTGCCGCTGAGGCCACCTGAAGCCGGA ACACCACTACGCCTTCAACCACCCGTCTCTCATCAACAACCTC	76

	ATGTCTCGGAGCAGCAGCACCACCACAGCCACCACCACCACC AACCCACAAAAATGGACCTCAAGGCCTACGAACAGGTGATGCA CTACCCCGGTACGGTTCCCCCATGCCTGGCAGCTTGGCCATG GGCCCGGTACGAACAAAACGGGCCTGGACGCCTCGCCCTGG CCGCAGATACCTCCTACTACCAGGGGTGTACTCCCGGCCCAT TATGAACCTCTCT	
ELK1	ATGGACCCATCTGTGACGCTGTGGCAGTTTCTGCTGCAGCTGC TGAGAGAGCAAGGCAATGGCCACATCATCTCCTGGACTTCACG GGATGGTGGTGAATTCAAGCTGGTGGATGCAGAGGAGGTGGCC CGGCTGTGGGGGTACGCAAGAACAAGACCAACATGAATTACG ACAAGCTCAGCCGGGCCTTGCGGTACTACTATGACAAGAACAT CATCCGCAAGGTGAGCGGCCAGAAGTTCGTCTACAAGTTTGTG TCCTACCCTGAGGTGCGAGGGTGCTCCACTGAGGACTGCCCGC CCCAGCCAGAGGTGTCTGTTACCTCCACCATGCCAAATGTGGC CCCTGCTGCTATACATGCCGCCCCAGGGGACACTGTCTCTGGA AAGCCAGGCACACCCAAGGTTGCAGGAATGGCAGGCCAGGCG GTTTGGCACGCAGCAGCCGGAACGAGTACATGCGCTCGGGCCT CTATTCCACCTTCACCATCCAGTCTCTGCAGCCGCAGCCACCC CCTCATCCTCGGCCTGCTGTGGTGTCTCCCAAGTGCAGCTCCTG CAGGGGCAGCAGCGCCCCCTCGGGGAGCAGGAGCACCAGTCC AAGCCCCCTTGGAGGCCTGTCTGGAGGCTGAAGAGGCCGGCTTG CCTCTGCAGGTCATCCTGACCCCGCCCGAGGCCCCAAACCTGA AATCGGAAGAGCTTAATGTGGAGCCGGGTTTGGGCCGGGCTTT GCCCCCAGAAGTGAAAGTAGAAGGGCCCCAAGGAAGAGTTGGAA GTTGCGGGGAGAGAGGGTTTGTGCCAGAAACCACCAAGGCCG AGCCAGAAGTCCCTCCACAGGAGGGCGTGCCAGCCCGGCTGCC CGCGGTTGTTATGGACACCGCAGGGCAGGCGGGCGGCCATGCG GCTTCCAGCCCTGAGATCTCCAGCCGCAGAAGGGCCGGAAGC CCCGGGACCTAGAGCTTCCACTCAGCCCGAGCCTGCTAGGTGG GCCGGGACCCGAACGGACCCAGGATCGGGAAGTGGCTCCGGC CTCCAGGCTCCGGGGCCGGCGCTGACCCCATCCCTGCTTCCTA CGCATACATTGACCCCGGTGCTGCTGACACCCAGCTCGCTGCC TCCTAGCATTCACCTTCTGGAGCACCTGAGTCCCATTGCGCCC CGTAGCCCGGCCAAGCTCTCCTTCCAGTTTCCATCCAGTGGCA GCGCCAGGTGCACATCCCTTCTATCAGCGTGGATGGCCTCTC GACCCCGTGGTGTCTCTCCCAAGGGCCCCAGAAGCCA	77
ELK3	ATGGAGAGTGCAATCACGCTGTGGCAGTTCTGTTGCAGTTGC TGCTGGATCAGAAACATGAGCATTTGATCTGCTGGACCTCGAA CGATGGTGAATTCAAGCTCCTCAAAGCAGAAGAAGTGGCCAAG CTGTGGGGACTCCGAAAAACAAAACAAATATGAACTATGATA AGCTGAGCAGAGCCCTGCGATACTATTATGACAAGAACATCAT CAAGAAGGTGATCGGGCAGAAGTTTGTGTACAAGTTTGTCTCT TTCCCGGAGATCCTGAAGATGGATCCTCACGCGGTGGAGATCA GCCGGGAGAGCCCTTCTGCTGCAGGACAGCGACTGCAAGGCGTC TCCGGAGGGCCGCGAGGCCACAAACACGGCCTGGCCGCCCTC AGAAGCACGAGCCGCAACGAATACATCCACTCAGGCCTGTACT CGTCCTTCACCATTAATTCCCTGCAGAACCCACCAGACGCCTT CAAGGCCATCAAGACGGAGAAGCTGGAGGAGCCGCCGAAGAC AGCCCCCCCCGTGGAAGAAGTCAGGACTGTGATCAGGTTTGTGA CCAATAAAACCGACAAGCACGTCACCAGGCCGGTGGTGTCCCT GCCTTCCACGTGAGAGGCTGCGGCGGCGTCCGCCTTCTGGCC TCGTCCGTCTCGGCCAAGATCTCCTCTTTAATGTTGCCAAACG CTGCCAGTATTTTCATCCGCCTCACCTTCTCATCTCGGTCCCC GTCCCTGTCCCCCAACTCACCCCTCCCTTCTGAACACAGAAGC CTCTTCTGGAGGCCGCTGCCATGACTCCGATTCCCTGGAGC	78

	CCTTGAACCTGTCATCGGGCTCCAAGACCAAGTCTCCATCTCT TCCCCCAAAGGCCAAAAAACCCTTGGAAATCTCAGCG CCCCCGCTGGTGCTCTCCGGCACCGACATCGGCTCCATCGCCC TCAACAGCCCAGCCCTCCCTCGGGATCCCTCACCCCAGCCTT CTTCACCGCACAGACACCAAATGGATTGCTTCTGACTCCGAGT CCACTGCTCTCCAGCATACATTTCTGGAGCAGCCTTAGTCCAG TTGCTCCGCTGAGTCCTGCCAGGCTGCAAGGGCCAAGCACGCT GTTCCAGTTCCCCACACTGCTTAATGGCCACATGCCAGTGCCA ATCCCCAGTCTGGACAGAGCTGCTTCTCCAGTACTGCTTTCTT CAAACCTCTCAGAAATCC	
ELK4	ATGGACAGTGTATCACCTGTGGCAGTTCCTTCTTCAGCTCC TGCAGAAGCCTCAGAACAAGCACATGATCTGTTGGACCTCTAA TGATGGGCAGTTTAAAGCTTTTGCAGGCAGAAGAGGTGGCTCGT CTCTGGGGGATTTCGCAAGAACAAGCCTAACATGAATTATGACA AACTCAGCCGAGCCCTCAGATACTATTATGTAAAGAATATCAT CAAAAAAGTGAATGGTCAGAAGTTTGTGTACAAGTTTGTCTCT TATCCAGAGATTTTGAACATGGATCCAATGACAGTGGGCAGGA TTGAGGGTGACTGTGAAAGTTTAAACTTCAGTGAAGTCAGCAG CAGTTCCAAAGATGTGGAGAATGGAGGGAAAGATAAACCACCT CAGCCTGGTGCCAAGACCTCTAGCCGCAATGACTACATACACT CTGGCTTATATTTCTTCACTTCTCAACTCTTTGAACTCCTC CAATGTAAAGCTTTTCAAATTGATAAAGACTGAGAATCCAGCC GAGAACTGGCAGAGAAAAATCTCCTCAGGAGCCACACCAT CTGTCATCAAATTTGTACGACACCTTCCAAAAAGCCACCGGT TGAACCTGTTGCTGCCACCATTTCATTTGGCCCAAGTATTTCT CCATCTTCAGAAGAACTATCCAAGCTTTGGAGACATTGGTTT CCCCAAACTGCCTTCCCTGGAAGCCCCAACCTCTGCCTCTAA CGTAATGACTGCTTTTGCCACCACACCACCATTTTCGTCCATA CCCCCTTTGCAGGAACCTCCCAGAACACCTTCACCACCACTGA GTTCTCACCCAGACATCGACACAGACATTGATTTCAGTGGCTTC TCAGCCAATGGAACCTCCAGAGAATTTGTCACTGGAGCCTAAA GACCAGGATTCAGTCTTGCTAGAAAAGGACAAAGTAAATAATT CATCAAGATCCAAGAAACCCAAAGGGTTAGAACTGGCACCAC CCTTGTGATCACGAGCAGTGATCCAAGCCCACTGGGAATACTG AGCCCATCTCTCCCTACAGCTTCTCTTACACCAGCATTTTTTT CACAGACACCCATCATACTGACTCCAAGCCCCCTTGCTCTCCAG TATCCACTTCTGGAGTACTCTCAGTCTGTTGCTCCCCTAAGT CCAGCCAGACTGCAAGGTGCTAACACACTTTTCCAGTTTCCTT CTGTACTGAACAGTCATGGGCCATTCACTCTGTCTGGGCTGGA TGGACCTTCCACCCCTGGGCCATTTTCCCAGACCTACAGAAG ACA	79
SRF	ATGTTACCGACCCAAGCTGGGGCCGCGCGGCTCTGGGCCGGG GCTCGGCCCTGGGGGGCAGCCTGAACCGGACCCCGACGGGGCG GCCGGGCGGCGGCGGCGGACACGCGGGGCTAACGGGGGCGCG GTCCCCGGGAATGGCGCGGGGCTCGGGCCCGCGCCTGGAGC GGGAGGCTGCGGCAGCGGCGGCAACCACCCGGCGCCACCGC GGGGGCCCTCTACAGCGGCAGCGAGGGCGACTCGGAGTCGGGC GAGGAGGAGGAGCTGGGCGCCGAGCGGCGCGGCTGAAGCGGA GCCTGAGCGAGATGGAGATCGGTATGGTGGTGGTGGGCCCGA GGCGTCGGCAGCGGCCACCGGGGGCTACGGGCGGCTGAGCGGC GCGGTGAGCGGGGCCAAGCCGGGTAAAGAAGACCCGGGGCCGCG TGAAGATCAAGATGGAGTTCATCGACAACAAGCTGCGGCGCTA CACGACCTTCAGCAAGAGGAAGACGGGCATCATGAAGAAGGCC TATGAGCTGTCCACGCTGACAGGGACACAGGTGCTGTTGCTGG TGGCCAGTGAGACAGGCCATGTGTATACCTTTGCCACCCGAAA	80

	ACTGCAGCCCATGATCACCAGTGAGACCGGCAAGGCACTGATT CAGACCTGCCTCAACTCGCCAGACTCTCCACCCCGTTGAGACC CCACAACAGACCAGAGAATGAGTGCCACTGGCTTTGAAGAGAC AGATCTCACCTACCAGGTGTCGGAGTCTGACAGCAGTGGGGAG ACCAAGGACACACTGAAGCCGGCGTTCACAGTCACCAACCTGC CGGGTACAACCTCCACCATCCAAACAGCACCTAGCACCTCTAC CACCATGCAAGTCAGCAGCGGCCCTCCTTTCCCATCACCAAC TACCTGGCACCAGTGTCTGCTAGTGTGAGCCCCAGTGTGTCA GCAGTGCCAATGGGACTGTGCTGAAGAGTACAGGCAGCGGCC TGTCTCCTCTGGGGGCCCTTATGCAGCTGCCTACCAGCTTCACC CTCATGCCTGGTGGGGCAGTGGCCCAGCAGGTCCCAGTGCAGG CCATTCAAGTGCACCAGGCCCCACAGCAAGCGTCTCCCTCCCG TGACAGCAGCACAGACCTCACGCAGACCTCCTCCAGCGGGACA GTGACGCTGCCCGCCACCATCATGACGTCATCCGTGCCACAA CTGTGGGTGGCCACATGATGTACCCTAGCCCGCATGCGGTGAT GTATGCCCCACCTCGGGCCTGGGTGATGGCAGCCTCACCGTG CTGAATGCCTTCTCCCAGGCACCATCCACCATGCAGGTGTCAC ACAGCCAGGTCCAGGAGCCAGGTGGCGTCCCCCAGGTGTTCT GACAGCATCATCTGGGACAGTGCAGATCCCTGTTTTAGCAGTT CAGCTCCACCAGATGGCTGTGATAGGGCAGCAGGCCGGGAGCA GCAGCAACCTCACCGAGCTACAGGTGGTGAACCTGGACACCGC CCACAGCACCAAGAGTGAA	
FOXMI	ATGAAAACCTAGCCCCGTCGGCCACTGATTCTCAAAGACGGA GGCTGCCCCCTTCTGTTCAAAATGCCCAAGTGAAACATCAGA GGAGGAACCTAAGAGATCCCCTGCCCAACAGGAGTCTAATCAA GCAGAGGCCCTCAAGGAAGTGGCAGAGTCCAACCTTTGCAAGT TTCCAGCTGGGATCAAGATTATTAACCACCCACCATGCCCAA CACGCAAGTAGTGGCCATCCCCAACAATGCTAATATTCACAGC ATCATCACAGCACTGACTGCCAAGGGAAAAGAGAGTGGCAGTA GTGGGCCCAACAAATTCATCTCATCAGCTGTGGGGGAGCCCC AACTCAGCCTCCAGGACTCCGGCCTCAAACCCAAACCAGCTAT GATGCCAAAAGGACAGAAGTGACCCTGGAGACCTTGGGACCAA AACCTGCAGCTAGGGATGTGAATCTTCCTAGACCACCTGGAGC CCTTTGCGAGCAGAAACGGGAGACCTGTGCAGATGGTGAGGCA GCAGGCTGCACATCAACAATAGCCTATCCAACATCCAGTGGC TTCGAAAGATGAGTTCTGATGGACTGGGCTCCCGCAGCATCAA GCAAGAGATGGAGGAAAAGGAGAATTGTCACCTGGAGCAGCGA CAGGTTAAGGTTGAGGAGCCTTCGAGACCATCAGCGTCCTGGC AGAACTCTGTGTCTGAGCGGCCACCCTACTCTTACATGGCCAT GATACAATTGCCATCAACAGCACTGAGAGGAAGCGCATGACT TTGAAAGACATCTATACGTGGATTGAGGACCACTTTCCCTACT TTAAGCACATTGCCAAGCCAGGCTGGAAGAACTCCATCCGCCA CAACCTTTCCCTGCACGACATGTTTGTCCGGGAGACGTCTGCC AATGGCAAGGTCTCCTTCTGGACCATTACCCCCAGTGCCAACC GCTACTTGACATTGGACCAGGTGTTTAAGCCACTGGACCCAGG GTCTCCACAATTGCCCGAGCACTTGAATCACAGCAGAAACGA CCGAATCCAGAGCTCCGCCGGAACATGACCATCAAACCGAAC TCCCCCTGGGCGCACGGCGGAAGATGAAGCCACTGCTACCACG GGTCAGCTCATACCTGGTACCTATCCAGTTCCCGGTGAACCAG TCACTGGTGTTCAGCCCTCGGTGAAGGTGCCATTGCCCTGG CGGCTTCCCTCATGAGCTCAGAGCTTGCCCGCCATAGCAAGCG AGTCCGCATTGCCCCCAAGGTTTGTGGGGAACAGGTGGTGT GGTTACATGAGTAAGTTCTTTAGTGGCGATCTGCGAGATTTTG GTACACCCATCACCAGCTTGTTTAATTTTATCTTTCTTTGTTT ATCAGTGCTGCTAGCTGAGGAGGGGATAGCTCCTCTTTCTTCT	81

	GCAGGACCAGGGAAAGAGGAGAAACTCCTGTTTGGAGAAGGGT TTTCTCCTTTGCTTCCAGTTCAGACTATCAAGGAGGAAGAAAT CCAGCCTGGGGAGGAAATGCCACACTTAGCGAGACCCATCAAA GTGGAGAGCCCTCCCTTGGAGAGTGGCCCTCCCCGGCCCCAT CTTTCAAAGAGGAATCATCTCACTCCTGGGAGGATTTCGTCCCA ATCTCCCACCCCAAGACCCAAGAAGTCTACAGTGGGCTTAGG TCCCCAACCCGGTGTGTCTCGGAAATGCTTGTGATTCAACACA GGGAGAGGAGGGAGAGGAGCCGGTCTCGGAGAAAACAGCATCT ACTGCCTCCCTGTGTGGATGAGCCGGAGCTGCTCTTCTCAGAG GGGCCCAGTACTTCCCGCTGGGCCGAGAGCTCCCGTTCCCAG CAGACTCCTCTGACCCTGCCTCCAGCTCAGTACTCCAGGA AGTGGGAGGACCTTTTAAGACACCCATTAAGGAAACGCTGCC ATCTCCTCCACCCCGAGCAAATCTGTCTCCCAAGAACCCCTG AATCCTGGAGGCTCACGCCCCAGCCAAAGTAGGGGACTGGA TTTCAGCCCAGTACAAACCTCCCAGGGTGCCTCTGACCCCTTG CCTGACCCCTGGGGCTGATGGATCTCAGCACCCTCCCTTGC AAAGTGCTCCCCCCTTGAATCACCGCAAAGGCTCCTCAGTTC AGAACCCTTAGACCTCATCTCCGTCCCCCTTTGGCAACTCTTCT CCCTCAGATATAGACGTCCCCAAGCCAGGCTCCCCGGAGCCAC AGGTTTCTGGCCTTGCAGCCAATCGTTCTCTGACAGAAGGCCT GGTCTTGGACACAATGAATGACAGCCTCAGCAAGATCCTGTG GACATCAGCTTTCCTGGCCTGGACGAGGACCCACTGGGCCCTG ACAACATCAACTGGTCCCAGTTTATTCCTGAGCTACAG	
FOXC1	ATGCAGGCGCGCTACTCCGTGTCCAGCCCCAACTCCCTGGGAG TGGTGCCTTACCTCGGCGGCGAGCAGAGCTACTACCGCGCGGC GGCCGCGGCGGCGGGGGCGGCTACACCGCCATGCCGGCCCCC ATGAGCGTGTACTCGCACCCCTGCGCACGCCGAGCAGTACCCGG GCGGCATGGCCCGCGCCTACGGGCCCTACACGCCGAGCCGCA GCCCCAAGGACATGGTGAAGCCGCCCTATAGCTACATCGCGCTC ATCACCATGGCCATCCAGAACGCCCGGACAAGAAGATCACCC TGAACGGCATCTACCAGTTCATCATGGACCGCTTCCCCTTCTA CCGGGACAACAAGCAGGGCTGGCAGAACAGCATCCGCCACAAC CTCTCGCTCAACGAGTGCTTCGTCAAGGTGCCGCGCGACGACA AGAAGCCGGGCAAGGGCAGCTACTGGACGCTGGACCCGGACTC CTACAACATGTTTCGAGAACGGCAGCTTCTTGCGGCGGCGGCGG CGCTTCAAGAAGAAGGACGCGGTGAAGGACAAGGAGGAGAAGG ACAGGCTGCACCTCAAGGAGCCGCCCCCGCCGGCCGAGCC CCCGCCCGCGCCGCGGAGCAGGCCGACGGCAACGCGCCCGGT CCGCAGCCGCGCCCGTGCATCCAGGACATCAAGACCGAGA ACGGTACGTGCCCTCGCCGCCCCAGCCCCGTGTCGCCGGCCGC CGCCCTGGGCAGCGGCAGCGCCGCCGCGGTGCCAAGATCGAG AGCCCCGACAGCAGCAGCAGCAGCCTGTCCAGCGGGAGCAGCC CCCCGGGACAGCTGCCGTCCGCGCGGCCGCTCAGCCTGGACGG TGCGGATTCCGCGCCGCCGCCGCCGCCCTCCGCCCGCCG CCGCACCATAGCCAGGGCTTCAGCGTGGACAACATCATGACGT CGCTGCGGGGGTTCGCCGAGAGCGCGGCCGCGGAGCTCAGCTC CGGCCTTCTGGCCTCGGCGGCGCGCTCCTCGCGCGCGGGGATC GCACCCCGCTGGCGCTCGGCGCCTACTCGCCCGGCCAGAGCT CCCTCTACAGCTCCCCCTGCAGCCAGACCTCCAGCGCGGGCAG CTCGGGCGGCGGCGGCGGCGCGGGGGCCGCGGGGGGCGCG GGCGGCGCGGGACCTACCACTGCAACCTGCAAGCCATGAGCC TGTACGCGGCCGCGAGCGCGGGGGCCACTTGCAGGGCGCGCC CGGGGGCGCGGGCGGCTCGGCCGTGGACGACCCCTGCCCGAC TACTCTCTGCCCTCCGGTCAACAGCAGCAGCTCGTCTCCCTGA GTCACGGCGGCGGCGGCGGCGGCGGCGGGGGAGGCCAGGAGGC	82

	CGGCCACCACCTGCGGCCACCAAGGCCGCTCACCTCGTGG TACCTGAACCAGGCGGGCGGAGACCTGGGCCACTTGGCGAGCG CGGCGGCGGGCGGGCGGGCGGCGGAGGCTACCCGGGCCAGCAGCA GAACTTCCACTCGGTGCGGGAGATGTTTCGAGTCACAGAGGATC GGCTTGAACAACCTCTCCAGTGAACGGGAATAGTAGCTGTCAAA TGGCCTTCCCTTCCAGCCAGTCTCTGTACCGCACGTCCGGAGC TTTCGTCTACGACTGTAGCAAGTTT	
FOXC2	ATGCAGGCGCGCTACTCCGTGTCCGACCCCAACGCCCTGGGAG TGGTGCCCTACCTGAGCGAGCAGAATTACTACCGGGCTGCGGG CAGCTACGGCGGCATGGCCAGCCCCATGGGCGTCTATTCCGGC CACCCGGAGCAGTACAGCGCGGGGATGGGCCGCTCCTACGCGC CCTACCACCACCAGCCCGCGCGCCTAAGGACCTGGTGAA GCCGCCCTACAGCTACATCGCGCTCATCACCATGGCCATCCAG AACGCGCCCGAGAAGAAGATCACCTTGAACGGCATCTACCAGT TCATCATGGACCGCTTCCCCTTCTACCGGGAGAACAAGCAGGG CTGGCAGAACAGCATCCGCCACAACCTCTCGCTCAACGAGTGC TTCGTCAAGGTGCCCCGCGACGACAAGAAGCCCGGCAAGGGCA GTTACTGGACCCCTGGACCCGGACTCCTACAACATGTTTCGAGAA CGGCAGCTTCCCTGCGGCGCCGGCGGCGCTTCAAAAAGAAGGAC GTGTCCAAGGAGAAGGAGGAGCGGGCCACCTCAAGGAGCCGC CCCCGGCGGCGTCCAAGGGCGCCCCGGCCACCCCCACCTAGC GGACGCCCCCAAGGAGGCCGAGAAGAAGGTGGTGATCAAGAGC GAGGCGGCGTCCCCGGCGCTGCCGGTCAACCAAGGTGGAGA CGCTGAGCCCCGAGAGCGCGCTGCAGGGCAGCCCGCGCAGCGC GGCCTCCACGCCCCGCGGCTCCCCCGACGGCTCGCTGCCGGAG CACCACGCCGCGGCGCCCAACGGGCTGCCTGGCTTCAGCGTGG AGAACATCATGACCCTGCGAACGTCGCCGCGGGCGGAGAGCT GAGCCCGGGGGCCGGACGCGCGGGCCTGGTGGTGCCGCGCTG GCGCTGCCCTACGCCGCGCGCCGCCCGCCGCTACGGCCAGC CGTGCGCTCAGGGCCTGGAGGCCGGGGCCGCCGGGGGCTACCA GTGCAGCATGCGAGCGATGAGCCTGTACACCGGGGCCGAGCGG CCGGCGCACATGTGCGTCCCGCCCGCCCTGGACGAGGCCCTCT CGGACCACCCGAGCGGCCCCACGTGCCCCCTGAGCGCTCTCAA CCTCGCCGCCGGCCAGGAGGGCGCGCTCGCCGCCACGGGCCAC CACCACAGCACACGCGCCACCACCACCCGCGAGGCGCGCCGC CCCCGCGGCTCCCCAGCCCCAGCCGACGCCGAGCCCGGGG CGCCGCGGCGCAGGCGGCTCCTGGTATCTCAACCACAGCGGG GACCTGAACCACCTCCCCGGCCACACGTTGCGGGCCAGCAGC AACTTTCCCCAACGTGCGGGAGATGTTCAACTCCCACCGGCT GGGGATTGAGAACTCGACCCTCGGGGAGTCCCAGGTGAGTGGC AATGCCAGCTGCCAGCTGCCCTACAGATCCACGCCGCTCTCT ATCGCCACGCAGCCCCCTACTCCTACGACTGCACGAAATAC	83
TWIST1	ATGATGCAGGACGTGTCCAGCTCGCCAGTCTCGCCGGCCGACG ACAGCCTGAGCAACAGCGAGGAAGAGCCAGACCGGCAGCAGCC GCCGAGCGGCAAGCGCGGGGACGCAAGCGGCGCAGCAGCAGG GCGAGCGCGGGCGGCGGCGGGGCCCGCGGAGCCGCGGGTG GGGGCGTTCGAGGGCGGCGACGAGCCGGGCAGCCCGGCCAGGG CAAGCGCGGCAAGAAGTCTGCGGGCTGTGGCGGCGGCGGCGG GCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAGTCCGC AGTCTTACGAGGAGCTGCAGACGCGGGTTCATGGCCAACGT GCGGGAGCGCCAGCGCACCCAGTCTGCTGAACGAGGCGTTCGCC GCGCTGCGGAAGATCATCCCCACGCTGCCCTCGGACAAGCTGA GCAAGATTGAGACCTCAAGCTGGCGGCCAGGTACATCGACTT CCTCTACCAGGTCTCCAGAGCGACGAGCTGGACTCCAAGATG GCAAGCTGCAGCTATGTGGCTCACGAGCGGCTCAGCTACGCCT	84

	TCTCGGTCTGGAGGATGGAGGGGGCCTGGTCCATGTCCGCGTC CCAC	
HIF1A	ATGGAGGGCGCCGGCGCGGAACGACAAGAAAAAGATAAGTT CTGAACGTCGAAAAGAAAAGTCTCGAGATGCAGCCAGATCTCG GCGAAGTAAAGAATCTGAAGTTTTTTATGAGCTTGCTCATCAG TTGCCACTTCCACATAATGTGAGTTCGCATCTTGATAAGGCCT CTGTGATGAGGCTTACCATCAGCTATTTGCGTGTGAGGAACT TCTGGATGCTGGTGATTGGATATTGAAGATGACATGAAAGCA CAGATGAATTGCTTTTATTTGAAAGCCTTGGATGGTTTTGTTA TGGTTCTCACAGATGATGGTGACATGATTTACATTTCTGATAA TGTGAACAAATACATGGGATTAACCTCAGTTTGAACCTAACTGGA CACAGTGTGTTTGATTTTACTCATCCATGTGACCATGAGGAAA TGAGAGAAATGCTTACACACAGAAATGGCCTTGTGAAAAAGGG TAAAGAACAAAACACACAGCGAAGCTTTTTTCTCAGAATGAAG TGTACCCTAACTAGCCGAGGAAGAACTATGAACATAAAGTCTG CAACATGGAAGGTATTGCACTGCACAGGCCACATTCACGTATA TGATACCAACAGTAACCAACCTCAGTGTGGGTATAAGAAACCA CCTATGACCTGCTTGGTGCTGATTTGTGAACCCATTCTCACC CATCAAATATTGAAATTCTTTAGATAGCAAGACTTTCCTCAG TCGACACAGCCTGGATATGAAATTTCTTATTGTGATGAAAGA ATTACCGAATTGATGGGATATGAGCCAGAAGAACTTTTAGGCC GCTCAATTTATGAATATTATCATGCTTTGGACTCTGATCATCT GACCAAACTCATCATGATATGTTTACTAAAGGACAAGTCACC ACAGGACAGTACAGGATGCTTGCCAAAAGAGGTGGATATGTCT GGGTTGAACTCAAGCAACTGTCATATATAACACCAAGAATTC TCAACCACAGTGCATTGTATGTGTGAATTACGTTGTGAGTGGT ATTATTCAGCACGACTTGATTTTCTCCCTTCAACAAACAGAAT GTGTCCTTAAACCGGTTGAATCTTCAGATATGAAAATGACTCA GCTATTCACCAAAGTTGAATCAGAAGATACAAGTAGCCTCTTT GACAAACTTAAGAAGGAACCTGATGCTTTAACTTTGCTGGCCC CAGCCGCTGGAGACACAATCATATCTTTAGATTTTGGCAGCAA CGACACAGAACTGATGACCAGCAACTTGAGGAAGTACCATTA TATAATGATGTAATGCTCCCTCACCCAACGAAAAATTACAGA ATATAAATTTGGCAATGTCTCCATTACCCACCGCTGAAACGCC AAAGCCACTTCGAAGTAGTGCTGACCCTGCACTCAATCAAGAA GTTGCATTAAAATTAGAACCAAATCCAGAGTCACTGGAACTTT CTTTTACCATGCCCCAGATTCAGGATCAGACACCTAGTCCTTC CGATGGAAGCACTAGACAAAGTTCACCTGAGCCTAATAGTCCC AGTGAATATTGTTTTTATGTGGATAGTGATATGGTCAATGAAT TCAAGTTGGAATTGGTAGAAAACTTTTGTCTGAAGACACAGA AGCAAAGAACCCATTTTCTACTCAGGACACAGATTTAGACTTG GAGATGTTAGCTCCCTATATCCCAATGGATGATGACTTCCAGT TACGTTCTTTCGATCAGTTGTCACCATTAGAAAGCAGTTCCGC AAGCCCTGAAAGCGCAAGTCTCAAAGCACAGTTACAGTATTC CAGCAGACTCAAATACAAGAACCTACTGCTAATGCCACCACTA CCACTGCCACCATTGATGAATTAAAAACAGTGACAAAAGACCG TATGGAAGACATTAAAAATATTGATTGCATCTCCATCTCCTACC CACATACATAAAGAACTACTAGTGCCACATCATCACCATATA GAGATACTCAAAGTCGGACAGCCTCACCAAACAGAGCAGGAAA AGGAGTCATAGAACAGACAGAAAAATCTCATCCAAGAAGCCCT AACGTGTTATCTGTCGCTTTGAGTCAAAGAACTACAGTTCTCTG AGGAAGAACTAAATCCAAAGATACTAGCTTTGCAGAATGCTCA GAGAAAGCGAAAAATGGAACATGATGGTTCACTTTTTCAAGCA GTAGGAATTATT	85
SALL4	ATGTCGAGGCGCAAGCAGGCGAAAACCCAGCACATCAACTCGG	86

AGGAGGACCAGGGCGAGCAGCAGCCGCAGCAGCAGACCCCGGA GTTTGCAGATGCGGCCCCAGCGGCCCGCGCGGGGGAGCTG GGTGTCTCCAGTGAACCACCCAGGGAATGACGAGGTGGCGAGTG AGGATGAAGCCACAGTAAAGCGGCTTCGTGCGGAGGAGACGCA CGTCTGTGAGAAATGCTGTGCGGAGTTCTTCAGCATCTCTGAG TTCCTGGAACATAAGAAAAATTGCACTAAAAATCCACCTGTCC TCATCATGAATGACAGCGAGGGGCCTGTGCCTTCAGAAGACTT CTCCGGAGCTGTACTGAGCCACCAGCCCACCAGTCCCAGGAGT AAGGACTGTCACAGGGAGAATGGCGGCAGCTCAGAGGACATGA AGGAGAAGCCGGATGCGGAGTCTGTGGTGTACCTAAAGACAGA GACAGCCCTGCCACCCACCCCCCAGGACATAAGCTATTTAGCC AAAGGCAAAGTGGCCAACACTAATGTGACCTTGCAGGCACTAC GGGGCACCAGGTGGCGGTGAATCAGCGGAGCGCGGATGCACT CCCTGCCCCCGTGCCTGGTGCCAACAGCATCCCGTGGGTCTC GAGCAGATCTTGTGTCTGCAGCAGCAGCAGCTACAGCAGATCC AGCTCACCGAGCAGATCCGCATCCAGGTGAACATGTGGGCCTC CCACGCCCTCCACTCAAGCGGGGCAGGGGCCGACACTCTGAAG ACCTTGGGCAGCCACATGTCTCAGCAGGTTTCTGCAGCTGTGG CTTTGCTCAGCCAGAAAGCTGGAAGCCAAGGTCTGTCTCTGGA TGCCTTGAAACAAGCCAAGCTACCTCACGCCAACATCCCTTCT GCCACCAGCTCCCTGTCCCCAGGGCTGGCACCCCTTCACTCTGA AGCCGGATGGGACCCGGGTGCTCCCGAACGTCATGTCCCGCCT CCCGAGCGCTTTGCTTCCTCAGGCCCCGGGCTCGGTGCTCTTC CAGAGCCCTTTCTCCACTGTGGCGCTAGACACATCCAAGAAAG GGAAGGGGAAGCCACCGAACATCTCCGCGGTGGATGTCAAACC CAAAGACGAGGCGGCCCTCTACAAGCACAAGTGTAAAGTACTGT AGCAAGGTTTTTGGGACTGATAGCTCCTTGCAGATCCACCTCC GCTCCCACACTGGAGAGAGACCCTTCGTGTGCTCTGTCTGTGG TCATCGCTTCACCACCAAGGGCAACCTCAAGGTGCACTTTCAC CGACATCCCCAGGTGAAGGCAAACCCCCAGCTGTTTGCCGAGT TCCAGGACAAAGTGGCGGCCGGCAATGGCATCCCTATGCACT CTCTGTACCTGACCCCATAGATGAACCGAGTCTTTCTTTAGAC AGCAAACCTGTCTTGTAAACACCTCTGTAGGGCTACCTCAGA ATCTTTCTTCGGGGACTAATCCCAAGGACCTCACGGGTGGCTC CTTGCCCGGTGACCTGCAGCCTGGGCCTTCTCCAGAAAGTGAG GGTGGACCCACACTCCCTGGGGTGGGACCAAACTATAATTCCC CAAGGGCTGGTGGCTTCCAAGGGAGTGGGACCCCTGAGCCAGG GTCAGAGACCCTGAAATTGCAGCAGTTGGTGGAGAACATTGAC AAGGCCACCACTGATCCCAACGAATGTCTCATTTGCCACCGAG TCTTAAGCTGTCAGAGCTCCCTCAAGATGCATTATCGCACCCA CACCGGGGAGAGACCGTTCCAGTGTAAGATCTGTGGCCGAGCC TTTTCTACCAAAGGTAACCTGAAGACACACCTTGGGGTTACC GAACCAACACATCCATTAAGACGCAGCATTTCGTGCCCCATCTG CCAGAAGAAGTTCACATAATGCCGTGATGCTGCAGCAACATATT CGGATGCACATGGGCGGTGAGATTCCCAACACGCCCCCTGCCAG AGAATCCCTGTGACTTTACGGGTCTTGAGCCAATGACCGTGGG TGAGAACGGCAGCACCGGCGCTATCTGCCATGATGATGTCATC GAAAGCATCGATGTAGAGGAAGTCAGCTCCCAGGAGGCTCCCA GCAGCTCCTCCAAGGTCCCCACGCCTCTTCCCAGCATCCACTC GGCATACCCACGCTAGGGTTTGCCATGATGGCTTCCTTAGAT GCCCCAGGGAAAGTGGGTCTGCCCCCTTTTAACCTGCAGCGCC AGGGCAGCAGAGAAAACGGTTCCGTGGAGAGCGATGGCTTGAC CAACGACTCATCCTCGCTGATGGGAGACCAGGAGTATCAGAGC CGAAGCCCAGATATCCTGGAAACCACATCCTTCCAGGCACTCT CCCCGGCCAATAGTCAAGCCGAAAGCATCAAGTCAAAGTCTCC	
--	--

	CGATGCTGGGAGCAAAGCAGAGAGCTCCGAGAACAGCCGCACT GAGATGGAAGGTCGGAGCAGTCTCCCTTCCACGTTTATCCGAG CCCCGCCGACCTATGTCAAGGTTGAAGTTCCTGGCACATTTGT GGGACCCTCGACATTGTCCCCAGGGATGACCCCTTTGTTAGCA GCCCAGCCACGCCGACAGGCCAAGCAACATGGCTGCACACGGT GTGGGAAGAACTTCTCGTCTGCTAGCGCTCTTCAGATCCACGA GCGGACTCACACTGGAGAGAAGCCTTTTGTGTGCAACATTTGT GGGCGAGCTTTTACCACCAAAGGCAACTTAAAGGTTCACTACA TGACACACGGGGCGAACAATAACTCAGCCCGCCGTGGAAGGAA GTTGGCCATCGAGAACACCATGGCTCTGTTAGGTACGGACGGA AAAAGAGTCTCAGAAATCTTCCCAAGGAAATCCTGGCCCCCTT CAGTGAATGTGGACCCTGTTGTGTGGAACCAAGTACACCAGCAT GCTCAATGGCGGTCTGGCCGTGAAGACCAATGAGATCTCTGTG ATCCAGAGTGGGGGGGTTCCCTACCCTCCCGGTTTCTTGGGGG CCACCTCCGTTGTGAATAACGCCACTGTCTCCAAGATGGATGG CTCCAGTCGGGTATCAGTGCAGATGTGGAAAAACCAAGTGCT ACTGACGGCGTTCCCAAACACCAGTTTCTCACTTCTCTGGAAG AAAACAAGATTGCGGTCAGC	
ELF1	ATGGCTGCTGTTGTCCAACAGAACGACCTAGTATTTGAATTTG CTAGTAACGTCATGGAGGATGAACGACAGCTTGGTGATCCAGC TATTTTTCCTGCCGTAATTGTGGAACATGTTCTGGTGCTGAT ATTCTCAATAGTTATGCCGGTCTAGCCTGTGTGGAAGAGCCCA ATGACATGATTACTGAGAGTTCAGTGGATGTTGCTGAAGAAGA AATCATAGACGATGATGATGATGACATCACCCCTACAGTTGAA GCTTCTTGTGATGACGGGGATGAAACAATTGAAACTATTGAGG CTGCTGAGGCACTCCTCAATATGGATTCCCCTGGCCCTATGCT GGATGAAAAACGAATAAATAATAATATATTTAGTTCACCTGAA GATGACATGGTTGTTGCCCCAGTCACCCATGTGTCCGTACAT TAGATGGGATTCCTGAAGTGATGGAACACAGCAGGTGCAAGA AAAATATGCAGACTCACCGGGAGCCTCATCACAGAACAGCCT AAGAGGAAAAAAGGAAGAAAACTAAACCACACGACCAGATT CCCCAGCCACTACGCCAAATATATCTGTGAAGAAGAAAAACAA AGATGGAAGGGGAAACACAATTTATCTTTGGGAGTTTTTACTG GCACTGCTCCAGGACAAGGCTACTTGTCTTAAATACATCAAGT GGACCCAGCGAGAGAAAGGCATTTTAAATTGGTGAGTTCTAA AGCAGTGTCCAGGTTGTGGGGGAAGCACAAAAACAAACCTGAT ATGAATTATGAGACCATGGGAAGAGCACTCAGGTACTATTACC AAAGGGGTATTCTGGCAAAAGTGGAAGGTCAGCGCTTGGTGTA TCAGTTTAAAGAAATGCCAAAAGATCTTATATATATAAATGAT GAGGATCCAAGTTCAGCATAGAGTCTTCAGATCCATCGCTAT CTTCATCAGCCACTTCAAATAGGAATCAAACCAGCCGGTCGAG AGTATCTTCAAGTCCAGGGGTAAAAGGAGGAGCCACTACAGTT CTAAAACCAGGGAATTCTAAAGCTGCAAAACCCAAAGATCCTG TGGAAGTTGCACAACCATCAGAAGTTTTGAGGACAGTGCAGCC CAGCAGTCTCCATATCCTACCCAGCTCTTCCGGACTGTTTCA GTAGTACAGCCAGTACAGGCTGTCCCAGAGGGAGAAGCAGTA GAACCAGTACCATGCAGGATGAAACATTAAATTCTTCCGTTCA GAGTATTAGGACTATACAGGCTCCAACCCAAGTTCCAGTGGTT GTGTCTCCTAGGAATCAGCAGTTGCATACAGTAACACTCCAAA CAGTGCCACTCACAACAGTTATAGCCAGCACAGATCCATCAGC AGGTACTGGATCTCAGAAGTTTATTTTACAAGCCATTCCATCA TCACAGCCCATGACAGTACTGAAAGAAAATGTCATGCTGCAGT CACAAAAGGCGGGCTCTCCTCCTTCAATTGTCTTGGGCCCTGC CCAGGTTCAGCAGGTCCTTACTAGCAATGTTTCAAGCATTTGC AATGGAACCGTCAGTGTGGCTTCCCTCTCCATCCTTCAAGTGCTA	87

	CTGCACCTGTGGTGACCTTTTCTCCTCGCAGTTCACAGCTGGT TGCTCACCCACCTGGCACTGTAATCACTTCAGTTATCAAACT CAAGAAACAAAACTCTTACACAGGAAGTAGAGAAAAAGGAAT CTGAAGATCATTTGAAAGAGAACTGAGAAAACGGAGCAGCA GCCACAGCCTTATGTGATGGTAGTGTCCAGTTCCAATGGATTT ACTTCTCAGGTAGCTATGAAACAAAACGAACTGCTGGAACCCA ACTCTTTT	
SOX9	ATGAATCTCCTGGACCCCTTCATGAAGATGACCGACGAGCAGG AGAAGGGCCTGTCCGGCGCCCCAGCCCCACCATGTCCGAGGA CTCCGCGGGCTCGCCCTGCCCGTCCGGCTCCGGCTCGGACACC GAGAACACGCGGCCCCAGGAGAACACGTTCCCCAAGGGCGAGC CCGATCTGAAGAAGGAGAGCGAGGAGGACAAGTTCCCGTGTG CATCCGCGAGGCGGTGAGCCAGGTGCTCAAAGGCTACGACTGG ACGCTGGTGCCCATGCCGGTGCGCGTCAACGGCTCCAGCAAGA ACAAGCCGCACGTCAAGCGGCCCATGAACGCCTTCATGGTGTG GGCGCAGGCGGCGCGCAGGAAGCTCGCGGACCAGTACCCGCAC TTGCACAACGCCGAGCTCAGCAAGACGCTGGGCAAGCTCTGGA GACTTCTGAACGAGAGCGAGAAGCGGCCCTTCGTGGAGGAGGC GGAGCGGCTGCGCGTGCAGCACAAGAAGGACCACCCGGATTAC AAGTACCAGCCGCGGCGGAGGAAGTCGGTGAAGAACGGGCAGG CGGAGGCAGAGGAGGCCACGGAGCAGACGCACATCTCCCCAA CGCCATCTTCAAGGCGCTGCAGGCCGACTCGCCACACTCCTCC TCCGGCATGAGCGAGGTGCACTCCCCCGGCGAGCACTCGGGGC AATCCCAGGGCCACCGACCCACCCACCCACCCCAAAACCGA CGTGACGCCGGGCAAGGCTGACCTGAAGCGAGAGGGGCGCCCC TTGCCAGAGGGGGGCGAGACAGCCCCCTATCGACTTCGCGGACG TGGACATCGGCGAGCTGAGCAGCGACGTATCTCCAACATCGA GACCTTCGATGTCAACGAGTTTGACCAGTACCTGCCGCCAAC GGCCACCCGGGGGTGCCGGCCACGCACGGCCAGGTCACCTACA CGGGCAGCTACGGCATCAGCAGCACCAGCGGCCACCCCGGCGAG CGCGGGCCACGTGTGGATGTCCAAGCAGCAGGCGCCGCGGCCA CCCCCGCAGCAGCCCCACAGGCCCCGCGGCCCCGCGAGGCGC CCCCGCAGCCGCAGGCGGCGCCCCACAGCAGCCGGCGGCACC CCCGCAGCAGCCACAGGCGCACAGCTGACCACGCTGAGCAGC GAGCCGGGCCAGTCCCAGCGAACGCACATCAAGACGGAGCAGC TGAGCCCCAGCCACTACAGCGAGCAGCAGCAGCACTCGCCCCA ACAGATCGCCTACAGCCCCCTCAACCTCCACACTACAGCCCC TCCTACCCGCCCATCACCCGCTCACAGTACGACTACACCGACC ACCAGAACTCCAGCTCCTACTACAGCCACGCGGCAGGCCAGGG CACCGGCCCTTACTCCACCTTACCTACATGAACCCCGCTCAG CGCCCCATGTACACCCCATCGCCGACACCTCTGGGGTCCCTT CCATCCCGCAGACCCACAGCCCCCAGCACTGGGAACAACCCGT CTACACACAGCTCACTCGACCT	88
SOX12	ATGGTGCAGCAGCGGGGCGGAGGGCCAAGCGGGACGGCGGGC CGCCGCCCCGGGACCCGGGCGGCGGAGGAGGGGGCGCGCA GCCCCGGCTGGTGCAAGACCCCGAGCGGCCACATCAAGAGGCCG ATGAACGCATTCATGGTGTGGTTCGACGACGAACGGCGGAAGA TCATGGACCAGTGGCCCGACATGCACAACGCCGAGATCTCCAA GCGCCTGGGCGCGGCTGGCAGCTGCTGCAGGACTCGGAGAAG ATCCCGTTCGTGCGGGAGGCGGAGCGGCTGCGGCTCAAGCACA TGGCGGATTACCCGACTACAAGTACCGGCCGCGCAAAAAGAG CAAGGGGGCGCCGCAAGGCGCGCCCCCGCCCCCGGTGGT AGCGGTGGCGGCAGCCGGCTCAAGCCCCGGGCGCAGCTGCCTG GCCGCGGGGGCGCCGAGCAGCGGGAGGGCCTTTGGGGGGCGG GGCGGCGGCGCCGAGGACGACGATGAAGACGACGACGAGGAG	89

	CTGCTGGAAGTGC GCCTGGTTCGAGACCCCGGGGCGGGAGCTGT GGAGGATGGTCCCGGCGGGACGGGCCGCTCGGGGACAAGCGGA GCGCGCCCAAGGGCCGTCGGGCGAGGGGGCGGCCGCCGCCGCC GCCGCTCCCCGACACCGTCGGAGGACGAGGAGCCGGAGGAAG AGGAGGAGGAGGCGGCAGCGGCTGAGGAAGGTGAAGAGGAGAC GGTGGCGTCGGGGGAGGAGTCGCTGGGCTTTCTGTCCAGGCTG CCCCCTGGCCCGGCCGGCTGGACTGCAGCGCCCTGGATCGCG ACCCGGACCTGCAGCCTCCCTCGGGCACGTGCGACTTCGAGTT CCCGGACTACTGCACCCCCGAGGTTACCGAGATGATCGCGGGG GACTGGCGCCCGTCTAGCATCGCAGACCTGGTTTTCACCTAC	
SOX18	ATGCAGAGATCGCCGCCCGGCTACGGCGCACAGGACGACCCGC CCGCCCCGCCGCACTGTGCATGGGCCCCGGGACACGGGGCCGC CGCTGACACGCGCGGCTCGCCGCCGGCCCCCGCCGCCCTCGCC GCGCCCCCGCGCCCCGCTCGCCGCCAGCCCGCAGCGCAGTC CCCCGCGCAGCCCCGAGCCGGGGCGCTATGGCCTCAGCCCGGC CGGCCGCGGGGAACGCCAGGCGGCAGACGAGTCGCGCATCCGG CGGCCCATGAACGCCTTCATGGTGTGGGCAAAGGACGAGCGCA AGCGGCTGGCTCAGCAGAACCCGACCTGCACAACGCGGTGCT CAGCAAGATGCTGGGCAAAGCGTGGAAGGAGCTGAACGCGGCG GAGAAGCGGCCCTTCGTGGAGGAAGCCGAACGGCTGCGCGTGC AGCACTTGCGCGACCACCCCACTACAAGTACCGGCCGCGCCG CAAGAAGCAGGCGCGCAAGGCCCGGCGGCTGGAGCCCCGGCCTC CTGCTCCCGGATTAGCGCCCCCGCAGCCACCGCCCGAGCCTT TCCCCGCGGCGTCTGGCTCGGCTCGCGCCTTCGCGAGCTGCC CCCGCTGGGCGCCGAGTTCGACGGCCTGGGGCTGCCACGCCC GAGCGCTCGCCTCTGGACGGCCTGGAGCCCGCGAGGCTGCCT TCTTCCCACCGCCCCGCGCGCCCGAGGACTGCGCGCTGCGGCC CTTCCGCGCGCCCTACGCGCCACCGAGTTGTGCGGGACCCC GGCGGTTGCTACGGGGCTCCCCTGGCGGAGGCGCTCAGGACCG CGCCCCCGCGGCGCCGCTCGCTGGCCTGTACTACGGCACCT GGGCACGCCCCGGCCCGTACCCCGGCCCGCTGTGCGCGCCGCC GAGGCCCGCCGCTGGAGAGCGCCGAGCCGCTGGGGCCCGCCG CCGATCTGTGGGCCGACGTGGACCTCACCGAGTTCGACCAGTA CCTCAACTGCAGCCGACTCGGCCGACGCCCCCGGGCTCCCG TACCAGTGGCACTGGCCAACTGGGCCCGCGCGCCATGTCCT GCCCAGAGGAGAGCAGCCTGATCTCCGCGCTGTGCGACGCCAG CAGCGCGGTCTATTACAGCGCGTGCATCTCCGGC	90
ETS1	ATGAAGGCGGCCGTGATCTCAAGCCGACTCTCACCATCATCA AGACGGAAAAAGTCGATCTGGAGCTTTTCCCCTCCCCGGATAT GGAATGTGCAGATGTCCCACTATTAAGTCCAAGCAGCAAAGAA ATGATGTCTCAAGCATTAAGCTACTTTTCAGTGGTTTCACTA AAGAACAGCAACGACTGGGGATCCCAAAAGACCCCCGGCAGTG GACAGAAACCCATGTTCCGGGACTGGGTGATGTGGGCTGTGAAT GAATTACAGCCTGAAAGGTGTAGACTTCCAGAAGTTCTGTATGA ATGGAGCAGCCCTCTGCGCCCTGGGTAAAGACTGCTTTCTCGA GCTGGCCCCAGACTTTGTTGGGGACATCTTATGGGAACATCTA GAGATCCTGCAGAAAGAGGATGTGAAACCATATCAAGTTAATG GAGTCAACCCAGCCTATCCAGAATCCCGCTATACCTCGGATTA CTTCATTAGCTATGGTATTGAGCATGCCCAGTGTGTTCCACCA TCGGAGTTCTCAGAGCCAGCTTCATCACAGAGTCTATCAGA CGCTCCATCCCATCAGCTCGGAAGAGCTCCTCTCCCTCAAGTA TGAGAATGACTACCCCTCGGTCACTCTCCGAGACCTCTCCAG ACAGACACCTTGCAAGATGACTACTTTGCTATCAAACAAGAAG TCGTACCCCCAGACAACATGTGCATGGGGAGGACAGTCTGTGG TAAACTCGGGGGCCAGGACTCTTTTGAAAGCATAGAGAGCTAC	91

	GATAGTTGTGATCGCCTCACCCAGTCCTGGAGCAGCCAGTCAT CTTTCAACAGCCTGCAGCGTGTTCCTCCTATGACAGCTTCGA CTCAGAGGACTATCCGGCTGCCCTGCCCAACCACAAGCCCAAG GGCACCTTCAAGGACTATGTGCGGGACCGTGCTGACCTCAATA AGGACAAGCCTGTCATTCCCTGCTGCTGCCCTAGCTGGCTACAC AGGCAGTGGACCAATCCAGCTATGGCAGTTTCTTCTGGAATTA CTCACTGATAAATCCTGTCAGTCTTTTATCAGCTGGACAGGAG ATGGCTGGGAATTCAAACCTTTCTGACCCAGATGAGGTGGCCAG GAGATGGGGAAAGAGGAAAAACAAACCTAAGATGAATTATGAG AAACTGAGCCGTGGCCTACGCTACTATTACGACAAAAACATCA TCCACAAGACAGCGGGGAAACGCTACGTGTACCGCTTTGTGTG TGACCTGCAGAGCCTGCTGGGGTACACCCCTGAGGAGCTGCAC GCCATGCTGGACGTCAAGCCAGATGCCGACGAG	
PAX3	ATGACCACGCTGGCCGGCGCTGTGCCCAGGATGATGCGGCCGG GCCCCGGGCGAGAATAACCCGCTAGCGGGTTCCCGCTGGAAGT GTCCACTCCCCTCGGCCAGGGCCGCGTCAACCAGCTCGGCGGT GTTTTTATCAACGGCAGGCCGCTGCCCAACCACATCCGCCACA AGATCGTGGAGATGGCCACCACGGCATCCGGCCCTGCGTCAT CTCGCGCCAGCTGCGCGTGTCCACGGCTGCGTCTCCAAGATC CTGTGCAGGTACCAGGAGACTGGCTCCATACGTCCTGGTGCCA TCGGCGGCAGCAAGCCCAAGCAGGTGACAACGCCTGACGTGGA GAAGAAAATTGAGGAATACAAAAGAGAGAACCCGGGCATGTTT AGCTGGGAAATCCGAGACAAATTACTCAAGGACGCGGTCTGTG ATCGAAACACCGTGCCGTCAGTGAGTTCCATCAGCCGCATCCT GAGAAGTAAATTGCGGAAAGGTGAAGAGGAGGAGGCCGACTTG GAGAGGAAGGAGGCAGAGGAAAGCGAGAAGAAGGCCAAACACA GCATCGACGGCATCCTGAGCGAGCGAGGAAAGGCCCTGGTCTC CGGAGTTTCTTCGCT	92
PAX8	ATGCCTCACAACTCCATCAGATCTGGCCATGGAGGGCTGAACC AGCTGGGAGGGGCCCTTTGTGAATGGCAGACCTCTGCCGGAAGT GGTCCGCCAGCGCATCGTAGACCTGGCCCAACAGGGTGTAAGG CCCTGCGACATCTCTCGCCAGCTCCGCGTCAGCCATGGCTGCG TCAGCAAGATCCTTGGCAGGTACTACGAGACTGGCAGCATCCG GCCTGGAGTGATAGGGGGCTCCAAGCCCAAGGTGGCCACCCCC AAGGTGGTGGAGAAGATTGGGGACTACAAACGCCAGAACCCCTA CCATGTTTGCCTGGGAGATCCGAGACCGGCTCCTGGCTGAGGG CGTCTGTGACAATGACACTGTGCCCAGTGTGAGCTCCATTAAT AGAATCATCCGGACCAAAGTGCAGCAACCATTCAACCTCCCTA TGGACAGCTGCGTGGCCACCAAGTCCCTGAGTCCCGGACACAC GCTGATCCCCAGCTCAGCTGTAACCTCCCCGGAGTCACCCAG TCGGATTCCCTGGGCTCCACCTACTCCATCAATGGGCTCCTGG GCATCGCTCAGCCTGGCAGCGACAAGAGGAAAATGGATGACAG TGATCAGGATAGCTGCCGACTAAGCATTGACTCACAGAGCAGC AGCAGCGGACCCCGAAAGCACCTTCGCACGGATGCCTTCAGCC AGCACACCTCGAGCCGCTCGAGTGCCCATTTGAGCGGCAGCA CTACCCAGAGGCCTATGCCTCCCCAGCCACACCAAGGCGAG CAGGGCCTCTACCCGCTGCCCTTGCTCAACAGCACCTGGACG ACGGGAAGGCCACCCCTGACCCCTTCCAACACGCCACTGGGGCG CAACCTCTCGACTCACCAGACCTACCCCGTGGTGGCAGATCCT CACTACCCCTTCGCCATAAAGCAGGAAACCCCGAGGTGTCCA GTTCTAGCTCCACCCCTTCCTCTTATCTAGCTCCGCCTTTTT GGATCTGCAGCAAGTCGGCTCCGGGGTCCCGCCCTTCAATGCC TTTCCCCATGCTGCCTCCGTGTACGGGCAGTTCACGGGCCAGG CCCTCCTCTCAGGGCGAGAGATGGTGGGGCCACGCTGCCCGG ATACCCACCCACATCCCCACCAGCGGACAGGGCAGCTATGCC	93

	TCCTCTGCCATCGCAGGCATGGTGGCAGGAAGTGAATACTCTG GCAATGCCTATGGCCACACCCCTACTCCTCCTACAGCGAGGC CTGGCGCTTCCCCAACTCCAGCTTGCTGAGTTCCCATATTAT TACAGTTCCACATCAAGGCCGAGTGCACCGCCCACCACTGCCA CGGCCTTTGACCATCTG	
GLI1	ATGTTCAACTCGATGACCCACCAATCAGTAGCTATGGCG AGCCCTGCTGTCTCCGGCCCTCCCCAGTCAGGGGGCCCCCAG TGTGGGGACAGAAGGACTGTCTGGCCCCGCCCTTCTGCCACCAA GCTAACCTCATGTCCGGCCCCACAGTTATGGGCCAGCCAGAG AGACCAACAGCTGCACCGAGGGCCCACTCTTTTCTTCTCCCCG GAGTGCAGTCAAGTTGACCAAGAAGCGGGCACTGTCCATCTCA CCTCTGTCTGGATGCCAGCCTGGACCTGCAGACGGTTATCCGCA CCTCACCCAGCTCCCTCGTAGCTTTCATCAACTCGCGATGCAC ATCTCCAGGAGGCTCCTACGGTCATCTCTCCATTGGCACCATG AGCCCATCTCTGGGATTCCCAGCCCAGATGAATCACCACAAAAG GGCCCTCGCCTTCCCTTGGGGTCCAGCCTTGTGGTCCCCATGA CTCTGCCCCGGGTGGGATGATCCACATCCTCAGTCCCGGGGA CCCTTCCCAACTTGCCAGCTGAAGTCTGAGCTGGACATGCTGG TTGGCAAGTGCCGGGAGGAACCTTGGAAGGTGATATGTCCAG CCCCAACTCCACAGGCATACAGGATCCCCTGTTGGGGATGCTG GATGGGCGGGAGGACCTCGAGAGAGAGGAGAAGCGTGAGCCTG AATCTGTGTATGAACTGACTGCCGTTGGGATGGCTGCAGCCA GGAATTTGACTCCCAAGAGCAGCTGGTGCACCACATCAACAGC GAGCACATCCACGGGGAGCGGAAGGAGTTCGTGTGCCACTGGG GGGGCTGCTCCAGGGAGCTGAGGCCCTTCAAAGCCAGTACAT GCTGGTGGTTACATGCGCAGACACACTGGCGAGAAGCCACAC AAGTGCACGTTTGAAGGGTGCCGGAAGTCATACTCACGCCTCG AAAACCTGAAGACGCACCTGCGGTACACACGGGTGAGAAGCC ATACATGTGTGAGCACGAGGGCTGCAGTAAAGCCTTCAGCAAT GCCAGTGACCGAGCCAAGCACCAGAATCGGACCCATTCCAATG AGAAGCCGTATGTATGTAAGCTCCCTGGCTGCACCAAACGCTA TACAGATCCTAGCTCGCTGCGAAAACATGTCAAGACAGTGCAT GGTCTTGACGCCCATGTGACCAAACGGCACCGTGGGGATGGCC CCCTGCCCTCGGGCACCATCCATTTCTACAGTGGAGCCCAAGAG GGAGCGGGAAGGAGGTCCCATCAGGGAGGAAAGCAGACTGACT GTGCCAGAGGGTGCCATGAAGCCACAGCCAAGCCCTGGGGCCC AGTCATCCTGCAGCAGTGACCACTCCCCGGCAGGGAGTGCAGC CAATACAGACAGTGGTGTGGAAATGACTGGCAATGCAGGGGGC AGCACTGAAGACCTCTCCAGCTTGGACGAGGGACCTTGCAATTG CTGGCACTGGTCTGTCCACTCTTCGCCGCCTTGAGAACCTCAG GCTGGACCAGCTACATCAACTCCGGCCAATAGGGACCCGGGGT CTCAAACCTGCCAGCTTGTCCACACCGGTACCACTGTGTCCC GCCGCTGGGCCCCCAGTCTCTCTTGAACGCCGCAGCAGCAG CTCCAGCAGCATCAGCTCTGCCTATACTGTCAGCCGCCGCTCC TCCCTGGCCTCTCCTTTCCCCCTGGCTCCCCACCAGAGAATG GAGCATCCTCCCTGCCTTGGCCTTATGCCTGCCAGCACTACCT GCTTCGGGCAAGATATGCTTCAGCCAGAGGGGGTGGTACTTCG CCCCTGCAGCATCCAGCCTGGATCGGATAGGTGGTCTTCCCA TGCCTCCTTGGAGAAGCCGAGCCGAGTATCCAGGATACAACCC CAATGCAGGGGTACCCGGAGGGCCAGTGACCCAGCCCAGGCT GCTGACCGTCTGTCCAGCTAGAGTCCAGAGGTTCAAGAGCC TGGGCTGTGTCCATACCCACCCACTGTGGCAGGGGGAGGACA GAACTTTGATCCTTACCTCCCAACCTCTGTCTACTCACCACAG CCCCCAGCATCACTGAGAATGCTGCCATGGATGCTAGAGGGC TACAGGAAGAGCCAGAAGTTGGGACCTCCATGGTGGGCAGTGG	94

	TCTGAACCCCTATATGGACTTCCCACCTACTGATACTCTGGGA TATGGGGGACCTGAAGGGGCAGCAGCTGAGCCTTATGGAGCGA GGGGTCCAGGCTCTCTGCCTCTTGGGCCTGGTCCACCCACCAA CTATGGCCCCAACCCCTGTCCCCAGCAGGCCTCATATCCTGAC CCCACCCAAGAAACATGGGGTGAGTTCCCTTCCCCTCTGGGC TGTACCCAGGCCCCAAGGCTCTAGGTGGAACCTACAGCCAGTG TCCTCGACTTGAACATTATGGACAAGTGCAAGTCAAGCCAGAA CAGGGGTGCCCAGTGGGGTCTGACTCCACAGGACTGGCACCCCT GCCTCAATGCCCCACCCAGTGAGGGGCCCCCACATCCACAGCC TCTCTTTTCCCATTACCCCCAGCCCTCTCCTCCCCAATATCTC CAGTCAGGCCCCCTATACCCAGCCACCCCTGATTATCTTCCTT CAGAACCCAGGCCTTGCTGGACTTTGATTCCCCCACCATTTC CACAGGGCAGCTCAAGGCTCAGCTTGTGTGTAATTATGTTCAA TCTCAACAGGAGCTACTGTGGGAGGGTGGGGGCAGGAAGATG CCCCCGCCAGGAACCTTCCTACCAGAGTCCCAAGTTTCTGGG GGGTTCACAGTTAGCCCAAGCCGTGCTAAAGCTCCAGTGAAC ACATATGGACCTGGCTTTGGACCCAACTTGCCCAATCACAAGT CAGGTTCCCTATCCCACCCCTTCACCATGCCATGAAAATTTTGT AGTGGGGGCAAATAGGGCTTCACATAGGGCAGCAGCACCCCT CGACTTCTGCCCCCATGCCCCACTTGCTATGGGCCTCTCAAAG TGGGAGGCACAAACCCAGCTGTGGTCATCCTGAGGTGGGCAG GCTAGGAGGGGGTCTGCTTGTACCTCCTCCCGAAGGACAG GTATGTAACCCCTGGACTCTCTTGATCTTGACAACACTCAGC TGGACTTTGTGGCTATTCTGGATGAGCCCCAGGGGCTGAGTCC TCCTCCTTCCCATGATCAGCGGGCAGCTCTGGACATACCCCA CCTCCCTCTGGGCCCCCAACATGGCTGTGGGCAACATGAGTG TCTTACTGAGATCCCTACCTGGGGAAACAGAATTCCTCAACTC TAGTGCCTAA	
GLI2	ATGGAGACGTCTGCCTCAGCCACTGCCTCCGAGAAGCAAGAAG CCAAAAGTGGGATCCTGGAGGCCGCTGGCTTCCCCGACCCGGG TAAAAAGGCCTCTCCTTTGGTGGTGGCTGCAGCGGCAGCAGCA GCGGTAGCTGCCCCAAGGAGTGCCGAGCATCTCTTGCCACCAT TCCATGCGCCCCCTACCGATTGACATGCGACACCAGGAAGGAAG GTACCATTACGAGCCTCATCTGTCCACGGTGTGCACGGGCCCC CCTGCCCTCAGCGGCAGCCCTGTCATCTCTGACATCTCCTTGA TCCGGCTTTCCCCGCACCCGGCTGGCCCTGGGGAGTCCCCCTT CAACGCCCCCACCCTACGTGAACCCCCACATGGAGCACTAC CTCCGTTCCTGTGCACAGCAGCCCCACGCTCTCCATGATCTCTG CAGCCAGGGGCTCAGCCCCGCTGATGTGGCCAGGAGCACCT TAAGGAGAGGGGACTGTTTGGCCTTCCTGCTCCAGGCACCACC CCCTCAGACTATTACCACCAGATGACCCTCGTGGCAGGCCACC CCGCGCCCTACGGGGACCTGCTGATGCAGAGCGGGGGCGCTGC CAGCGCACCCCATCTCCACGACTACCTCAACCCCGTGGACGTG TCCCGTTTCTCCAGCCCGCGGGTGACGCCCCGCCTGAGCCGCA AGCGGGCGCTGTCCATCTCCCCACTCTCAGACGCCAGCCTGGA CCTGCAGCGGATGATCCGCACCTCACCCAACTCGCTAGTGGCC TACATCAACAACCTCCCGAAGCAGCTCGGCGGCCAGCGGTTCTT ACGGGCATCTGTGACGGGTGCCCTCAGCCCAGCCTTCACCTT CCCCACCCCATCAACCCCGTGGCCTACCAGCAGATTCTGAGC CAGCAGAGGGGTCTGGGGTCAGCCTTTGGACACACACCACCCC TGATCCAGCCCTCACCCACCTTCCTGGCCCAGCAGCCCATGGC CCTCACCTCCATCAATGCCACGCCACCCAGCTCAGCAGCAGC AGCAACTGTCTGAGTGACACCAACCAGAACAAGCAGAGCAGTG AGTCGGCCGTGAGCAGCACCGTCAACCCCTGTCGCCATTACAA GCGCAGCAAGGTCAAGACCCGAGCCTGAGGGCCTGCGGCCGGCC	95

TCCCCCTCTGGCGCTGACGCAGGGCCAGGTGTCTGGACACGGCT CATGTGGGTGTGCCCTTCCCCCTCTCCCAGGAGCAGCTGGCTGA CCTCAAGGAAGATCTGGACAGGGATGACTGTAAGCAGGAGGCT GAGGTGGTCATCTATGAGACCAACTGCCACTGGGAAGACTGCA CCAAGGAGTACGACACCCAGGAGCAGCTGGTGCATCACATCAA CAACGAGCACATCCACGGGGAGAAGAAGGAGTTTGTGTGCCGC TGGCAGGCCTGCACGCGGGAGCAGAAGCCCTTCAAGGCGCAGT ACATGCTGGTGGTGCACATGCGGCGACACACGGGCGAGAAGCC CCACAAGTGCACGTTTCGAGGGCTGCTCGAAGGCCTACTCCCGC CTGGAGAACCTGAAGACACACCTGCGGTCCCACACCGGGGAGA AGCCATATGTGTGTGAGCACGAGGGCTGCAACAAAGCCTTCTC CAACGCCCTCGGACCGCGCCAAGCACCAGAATCGCACCCACTCC AACGAGAAACCCTACATCTGCAAGATCCCAGGCTGCACCAAGA GATACACAGACCCCAGCTCTCTCCGGAAGCATGTGAAAACGGT CCACGGCCCAGATGCCACGTCACCAAGAAGCAGCGCAATGAC GTGCACCTCCGCACACCGCTGCTCAAAGAGAATGGGGACAGTG AGGCCGGCACGGAGCCTGGCGGGCCAGAGAGCACCGAGGCCAG CAGCACCAGCCAGGCCGTGGAGGACTGCCTGCACGTGAGAGCC ATCAAGACCGAGAGCTCCGGGCTGTGTGAGTCCAGCCCCGGGG CCCAGTCGTCTTGCAGCAGCGAGCCCTCTCTCTGGGCAGTGC CCCCAACATGACAGTGGCGTGGAGATGCCGGGGACGGGGCCC GGGAGCCTGGGAGACCTGACGGCACTGGATGACACACCCCCAG GGGCCGACACCTCAGCCCTGGCTGCCCCCTCCGCTGGTGGCCT CCAGCTGCGCAAACACATGACCACCATGCACCGGTTTCGAGCAG CTCAAGAAGGAGAAGCTCAAGTCACTCAAGGATTCTGCTCAT GGGCCGGGCGGACTCCACACACGCGGAACACCAAGCTGCCTCC CCTCCCGGGAAGTGGCTCCATCCTGGAAAACCTTCAGTGGCAGT GGGGGGCGGGGGCCCGGGGGCTGCTGCCGAACCCGCGGCTGT CGGAGCTGTCCGCGAGCGAGGTGACCATGCTGAGCCAGCTGCA GGAGCGCCGCGACAGCTCCACCAGCACGGTCAGCTCGGCCTAC ACCGTGAGCCGCCGCTCCTCCGGCATCTCCCCCTACTTCTCCA GCCGCCGCTCCAGCGAGGCCTCGCCCCCTGGGCGCCGGCCGCC GCACAACGCGAGCTCCGCTGACTCCTACGACCCCATCTCCACG GACGCTCGCGGCGCTCGAGCGAGGCCAGCCAGTGCAGCGGCG GCTCCGGGCTGTCTAACCTCACGCCGGCGCAGCAGTACAGCCT GCGGGCCAAGTACGCGGCAGCCACTGGCGGCCCCCGCCACT CCGCTGCCGGGCTGGAGCGCATGAGCCTGCGGACCAGGCTGG CGCTGTGAGCGCGCCGAGCGCACGCTGCCCGCCGGCTGCCC ACGCCACTGGGGCCGCGGCTGGCAGCGACGGGCCGACCTAT GGCCACGGCCACGCGGGGGCTGCGCCCGCCTTCCCCACGAGG CTCCAGGCGGCGGAGCCAGCGGGCCAGCGACCCTGTGCGGCG GCCCGATGCCCTGTCCCTGCCGCGGGTGCAGCGCTTCCACAGC ACCCACAACGTGAACCCCGGCCGCTGCCGCCCTGTGCCGACA GGCGAGGCCCTCCGCTGCAGAGCCACCCGAGCACCGACGGCGG CCTGGCCCGCGGCGCCTACTCGCCCCGGCCGCTAGCATCAGC GAGAACGTGGCGATGGAGGCCGTGGCGGCAGGAGTGGACGGCG CGGGGCCGAGGCCGACCTGGGGCTGCCGGAGGACGACCTGGT GCTTCCAGACGACGTGGTGCAGTACATCAAGGCGCACGCCAGT GGCGCTCTGGACGAGGGCACCGGGCAGGTGTATCCACGGAAA GCACTGGCTTCTGTACAACCCAGACTACCCAGCCCGGGGCT GCACGGCCAGCGCAGGATGGTGGCTGCGGACTCCAACGTGGGC CCCTCCGCCCTTATGCTGGGAGGATGCCAGTTAGGCTTTGGGG CGCCCTCCAGCCTGAACAAAAATAACATGCCTGTGAGTGGAA TGAGGTGAGCTCCGGCACCGTAGACGCCCTGGCCAGCCAGGTG AAGCCTCCACCCTTTCCTCAGGGCAACCTGGCGGTGGTGCAGC	
--	--

	AGAAGCCTGCCTTTGGCCAGTACCCGGGCTACAGTCCGCAAGG CCTACAGGCTAGCCCTGGGGGCTGGACAGCACGCAGCCACAC CTGCAGCCCCGCAGCGGAGCCCCCTCCCAGGGCATCCCCAGGG TAAACTACATGCAGCAGCTGCGACAGCCAGTGGCAGGCAGCCA GTGTCTTGGCATGACTACCACTATGAGCCCCCATGCCTGCTAT GGCCAAGTCCACCCCCAGCTGAGCCCCAGCACCATCAGTGGGG CCCTCAACCAGTTCCCCCAATCCTGCAGCAACATGCCAGCCAA GCCAGGGCATCTGGGGCACCCCTCAGCAGACAGAAAGTGGCACCT GACCCACCCACGATGGGCAATCGCCACAGGGAACCTGGGGTCC CCGATTACGCCCTGGCTGGAGTGCCACCACCTCACCCAGTCCA GAGCTACCCACAGCAGAGCCATCACCTGGCAGCCTCCATGAGC CAGGAGGGCTACCACCAGGTCCCCAGCCTTCTGCCTGCCCGCC AGCCTGGCTTCATGGAGCCCCAAACAGGCCCCGATGGGGTGGC TACAGCAGGCTTTGGCCTAGTGCAGCCCCGGCCTCCCCTCGAG CCCAGCCCCACTGGCCGCCACCGTGGGGTACGTGCTGTGCAGC AGCAGCTGGCCTACGCCAGGGCCACAGGCCATGCCATGGCTGC CATGCCGTCCAGTCAGGAAACAGCAGAGGCTGTGCCCAAGGGA GCGATGGGCAACATGGGGTCGGTGCCTCCCCAGCCGCCTCCGC AGGACGCAGGTGGGGCCCCGGACCACAGCATGCTCTACTACTA CGGCCAGATCCACATGTACGAACAGGATGGAGGCCTGGAGAAC CTCGGGAGCTGCCAGGTTCATGCGGTCCCAGCCACCACAGCCAC AGGCCTGTCCAGGACAGCATCCAGCCCCAGCCCTTGCCCTCACC AGGGGTCAACCAGGTGTCCAGCACTGTGGACTCCAGCTCCTG GAGGCCCCCAGATTGACTTCGATGCCATCATGGATGATGGCG ATCACTCGAGTTTGTCTCGGGTGTCTGAGCCCCAGCCTCCT CCACAGCCTCTCCCAGAACTCCTCCCGCCTCACCACCCCCGA AACTCCTTGACCCTGCCCTCCATCCCCGCAGGCATCAGCAACA TGGCTGTGGGGACATGAGCTCCATGCTCACCAGCCTCGCCGA GGAGAGCAAGTTCCTGAACATGATGACC	
GLI3	ATGGAGGCCCAGTCCCACAGCTCCACGACCACTGAAAAGAAAA AAGTTGAGAATTCCATAGTGAAGTGCTCCACTCGAACAGATGT GAGCGAGAAAGCCGTTGCCCTCCAGCACCACCTTCTAATGAGGAT GAAAGTCCTGGACAGACTTATCACAGAGAGAGAAGAAACGCAA TCACTATGCAGCCACAGAATGTCCAGGGGCTCAGCAAAGTCAG TGAGGAACCTTCAACATCGAGTGACGAGAGGGGCTCATTGATC AAGAAAGAGATCCATGGGTCCCTGCCACACGTGGCGGAGCCCT CTGTGCCGTACCGCGGGACGGTGTGTTGCCATGGACCCAGGAA TGGTTACATGGAGCCCCACTACCACCCTCCTCATCTTTTCCCT GCCTTCCATCCTCCTGTACCAATTGATGCCAGACATCATGAGG GCCGTTACCATTACGATCCATCTCCGATTCTCCATTGCATAT GACTTCCGCCTTATCTAGTAGCCCTACGTATCCGGACCTGCCC TTCATTAGGATCTCCCCACACCGGAACCCCACTGCTGCTTCCG AGTCTCCCTTCAGCCCTCCACATCCCTACATTAATCCCTACAT GGACTATATCCGCTCCTTGACAGCAGCCCATCGCTCTCCATG ATCTCAGCAACCCGTGGGCTGAGCCCTACAGATGCGCCCCATG CAGGAGTCAGCCCAGCAGAATACTATCATCAGATGGCCCTGCT AACTGGCCAGCGCAGCCCTATGCAGACATTATTCCCTCAGCT GCCACCGCCGGCACGGGGGCCATCCACATGGAATATCTTCATG CTATGGATAGCACCAGATTCTCCAGCCCCAGGCTGTCAGCCAG GCCGAGCCGAAAACGTACACTGTCCATATCACCCTCTCCGAT CATAGCTTTGACCTTCAGACCATGATAAGGACGTCTCCCACT CCTTGGTCACGATTCTCAATAATTCCCGTAGCAGCTCTTCAGC AAGTGGCTCCTATGGTCACTTATCTGCAAGTGCAATCAGCCCT GCCTTGAGCTTACCTACTCTTCCGCGCCCGTCTCTCTCCACA TGCATCAGCAGATCCTAAGCCGACAACAGAGCTTAGGTTGAGC	96

CTTTGGACACAGCCCTCCACTCATCCACCCTGCCCCAACTTTT CCAACACAGAGGCCTATTCCAGGGATCCCTACGGTTCTGAACC CCGTCCAGGTGAGTCTCCGGCCCTTCTGAGTCTCACAGAACAA GCCCACGAGTGAGTCTGCAGTGAGCAGCACTGGTGACCCGATG CACAACAAGAGGTCCAAGATCAAACCCGATGAAGACCTCCCCA GCCCAGGGGCTCGGGGGCAGCAGGAACAGCCGAAGGAACAAC CCTTGTC AAGGAGGAAGGGGACAAAGATGAAAGCAAACAGGAG CCTGAAGTCATCTATGAGACAAACTGCCACTGGGAAGGCTGCG CGAGGGAGTTCGACACCCAAGAGCAGCTTGTGCACCATATAAA TAACGACCATATTTCATGGAGAGAAGAAGGAGTTCGTGTGCAGG TGGCTGGACTGCTCAAGAGAGCAGAAACCCCTTCAAAGCCCAGT ATATGTTGGTAGTGCATATGAGAAGACACAGGGCGAGAAGCC TCACAAATGCACTTTTGAAGGTTGCACAAAGGCCTACTCGAGA CTAGAAAACCTGAAAACACACTTGAGATCTCACACTGGAGAGA AACCATACGTCTGTGAGCACGAAGGTTGCAACAAGGCTTTCTC AAATGCCCTCTGATCGCGCCAAACACCAAAACAGAACGCATTCC AATGAGAAACCATATGTGTGCAAAATCCCAGGCTGCACTAAGC GTTACACAGACCCAAGCTCCCTCCGGAAACATGTGAAGACAGT GCATGGCCCAGAGGCTCATGTCAACCAAGAAGCAGCGAGGGGAC ATCCATCCTCGGCCGCCACCCCCGAGAGATTCCGGCAGCCATT CACAGTCCAGGTGCGCTGGCCGACCGACTCAGGGAGCCCTTGG TGAGCAGCAGGACCTCAGCAACACTACCTCAAAGCGGGAAGAA TGCCCTCCAGGTGAAAACCGTCAAGGCAGAGAAGCCAATGACAT CTCAGCCAAGCCCTGGTGGTCAGTCTTCATGCAGCAGCCAACA GTCCCCCATCAGCAACTATTCCAACAGTGGGCTCGAGCTTCCT CTGACCGATGGAGGTAGTATAGGAGACCTCAGTGCCATCGATG AAACCCCAATCATGGACTCAACCATTTCCACTGCAACCACAGC CCTTGCTTTGCAAGCCAGGAGAAACCCGGCAGGGACCAAATGG ATGGAGCACGTAAAAC TAGAAAGGCTAAAACAAGTGAATGGAA TGTTTTCCGCGACTGAACCCCATTTCTACCCCTAAAGCCCCTGC GGTCTCTCCTCTCATAGGAAATGGCACACAGTCCAACAACACC TGCAGCTTGGGTGGGCCCATGACGCTTCTCCCGGCAGAAAGCG ACCTCTCTGGGGTGGACGTCACTATGCTGAACATGCTCAACAG AAGGGACAGCAGCGCCAGCACCATCAGCTCGGCCTACCTGAGC AGCCGCGCTCCTCAGGGATCTCGCCCTGCTTCTCCAGCCGCC GCTCCAGCGAGGCGTCACAGGCCGAGGGCCGCGCCGAGAACGT GAGCGTGGCCGACTCCTACGACCCCATCTCCACCGACGCCTCG CGCCGCTCCAGCGAAGCCAGCCAGAGCGACGGCCTGCCAGCC TGCTCAGCCTCAGCCCCGCCAGCAGTACCGCCTCAAGGCCAA GTACGCGGCTGCCACAGGAGGGCCGCCGCCGACGCCCTGCCC AACATGGAGAGGATGAGCCTGAAGACGCGCCTGGCGCTGCTCG GGGATGCCCCTCGAGCCTGGCGTGGCCCTGCCTCCAGTTCATGC CCCGAGGAGGTGCAGCGACGGGGGAGCCACGGCTACGGGCGG CGCCACCTGCAGCCGACGATGCGCCGGGCCACGGCGTGAGGA GGGCCAGCGACCCGGTGGGACAGGCTCCGAGGGCCTGGCCCT GCCTCGTGTGCCGCGCTTCAGCAGCCTCAGCAGCTGCAACCCC CCGGCGATGGCCACGTCCGCGGAGAAGCGCAGTCTCGTGCTTC AGAATTACACGCGGCCCGAGGGCGGCCAGTCCCGAAACTTCCA CTCGTCCCCCTGTCTCCAGCATCACCGAGAACGTACCCCTG GAGTCCCTGACCATGGACGCTGATGCCAACCTGAACGATGAGG ATTTCTGCGGACGACGTGGTGAGTATTTAAATTCCAGAA CCAAGCAGGGTACGAGCAGCACTTCCCCAGCGCCCTCCCGAC GACAGCAAAGTGCCCCACGGGCCCCGGTGACTTTGACGCGCCG GGCTGCCAGACAGCCACGCTGGCCAGCAGTTCCATGCCCTCGA GCAGCCCTGCCCCGAGGGCAGCAAAACCGACCTGCCCATTCAG	
---	--

	TGGAACGAAGTCAGCTCCGGAAGCGCCGACCTGTCCTCCTCCA AGCTCAAGTGTGGGCCGCGCCCGCTGTGCCGCAGACTCGCGC CTTTGGGTCTGCAACGGCATGGTCGTCCACCCGAGAACCCC TTGAGGAGCGGGCCTGCTGGGGGCTATCAGACCCTCGGGGAGA ACAGCAACCCCTACGGTGGCCAGAGCACTTGATGCTCCACAA CAGCCCCGGAAGTGGCACCAGTGAAACGCCTTCCATGAACAG CCCTGTAAGGCCCCGAGTATGGGAACTGTCTCAACAGGCAGC CAGTGGCCCCCTGGTGCACTCGACGGTGCTGTGGTGCCGGGAT TCAAGCCTCAAAGCTGAAGAGCACCCCATGCAAGGGAGCGGG GGCCAGCTGAATTTCCGGCTGCCGGTAGCGCCAAATGAGTCAG CTGGCAGCATGGTGAATGGCATGCAGAACAGGACCCAGTGGG ACAGGGGTACCTGGCTCACCAGCTCCTCGGCGACAGCATGCAG CACCCGGGGGCAGGCCGCCCGGTCAGCAGATGCTTGGGCAGA TTAGTGCTACCTCACACATCAACATCTACCAAGGGCCAGAGAG CTGCCCTGCCAGGGGCTCACGGCATGGGCAGCCAGCCGTCAAGC TTGGCAGTTGTCAGGGGCTACCAGCCATGTGCCAGCTTTGGGG GCAGCAGGCGCCAGGCTATGCCGAGGGACAGCCTTGCTCTGCA GTCAGGACAGCTCAGTGACACAAGTCAGACCTGCAGGGTGAAT GGTATCAAGATGGAGATGAAAGGGCAGCCCCATCCGCTGTGCT CTAATCTGCAGAATTACTCTGGTCAGTTCTATGACCAAACCGT GGGCTTCAGTCAGCAAGACACGAAAGCTGGTTTATTCTCTATT TCAGACGCCAGCTGCCTGCTACAGGGGACCAGCGCCAAAACT CTGAGTTACTTTCCCGAGTGCTAATCAGGTGACAAGCACAGT GGACAGCCTCGACAGCCATGACCTGGAAGGGGTACAGATTGAC TTCGATGCCATCATAGACGATGGGGACCACTCCAGCCTGATGT CGGGGGCCCTGAGCCCAAGTATCATTGAGAACCTTTCCCATAG CTCCTCCCGCCTCACCACGCCTCGGGCGTCCCTCCCATTTCCA GCGCTGTCCATGAGCACCACCAACATGGCTATCGGGGACATGA GTTCTTTGCTGACCTCCCTAGCGGAAGAAAGCAAATTCCTTGC AGTTATGCAA	
ETV1	ATGGATGGATTTTATGACCAGCAAGTGCCTTACATGGTCACCA ATAGTCAGCGTGGGAGAAATTTGTAACGAGAAACCAACAAATGT CAGGAAAAGAAAATTCATTAACAGAGATCTGGCTCATGATTCA GAAGAACTCTTTCAAGATCTAAGTCAATTACAGGAAACATGGC TTGCAGAAGTGGCTTTTCATGGCCTGCCACTGAAAATCAAGAA AGAACCCACAGTCCATGTTTCAGAAATCAGCTCTGCCTGCAGT CAAGAACAGCCCTTTAAATTCAGCTATGGAGAAAAGTGCCTGT ACAATGTCAGTGCCTATGATCAGAAGCCACAAGTGGGAATGAG GCCCTCCAACCCCCCACACCATCCAGCACGCCAGTGTCCCCA CTGCATCATGCATCTCCAACTCAACTCATAACCGAAACCTG ACCGGGCCCTTCCCAGCTCACCTCCCTCCATCGCAGTCCATACC AGATAGCAGCTACCCCATGGACCACAGATTTTCGCCGCCAGCTT TCTGAACCCGTGTAACCTCTTCTCCTTTGCCGACGATGCCAA GGGAAAGGACGTCCATGTACCAACGCCAGATGTCTGAGCCAAA CATCCCCCTCCCACCACAAGGCTTTAAGCAGGAGTACCACGAC CCAGTGTATGAACACAACACCATGGTTGGCAGTGCAGGCCAGCC AAAGCTTTCCCCCTCCTCTGATGATTAAACAGGAACCCAGAGA TTTTGCATATGACTCAGAAGTGCCTAGCTGCCACTCCATTTAT ATGAGGCAAGAAGGCTTCTGGCTCATCCAGCAGAACAGAAG GCTGTATGTTTGAAGGGCCCCAGGCAGTTTATGATGACAC CTGTGTTGTCCCAGAAAAATTCGATGGAGACATCAAACAAGAG CCAGGAATGTATCGGAAGGACCCACATACCAACGGCGAGGAT CACTTCAGCTCTGGCAGTTTGTGGTAGCTCTTCTGGATGACCC TTCAAATTCATTTTATTGCCTGGACTGGTCGAGGCATGGAA TTTAACTGATTGAGCCTGAAGAGGTGGCCCCGACGTTGGGGCA	97

	TTCAGAAAAACAGGCCAGCTATGAACTATGATAAACTTAGCCG TTCACCTCCGCTATTACTATGAGAAAAGGAATTATGCAAAAGGTG GCTGGAGAGAGATATGTCTACAAGTTTGTGTGTGATCCAGAAG CCCTTTTCTCCATGGCCTTTCAGATAATCAGCGTCCACTGCT GAAGACAGACATGGAACGTCACATCAACGAGGAGGACACAGTG CCTCTTTCTCACTTTGATGAGAGCATGGCCTACATGCCGGAAG GGGGCTGCTGCAACCCCCACCCCTACAACGAAGGCTACGTGTA T	
ETV2	ATGGACCTGTGGAAGTGGGATGAGGCATCCCCACAGGAAGTGC CTCCAGGGAACAAGCTGGCAGGGCTTGAAGGAGCCAAATTAGG CTTCTGTTTCCCTGATCTGGCACTCCAAGGGGACACGCCGACA GCGACAGCAGAGACATGCTGGAAAAGGTCCCATTTCAGCTGTGGC AGTTCCCTCCTGGAGCTGCTCCACGACGGGGCGCGTAGCAGCTG CATCCGTTGGACTGGCAACAGCCGCGAGTTCCAGCTGTGCGAC CCCAAAGAGGTGGCTCGGCTGTGGGGCGAGCGCAAGAGAAAGC CGGGCATGAATTACGAGAAGCTGAGCCGGGGCCTTCGCTACTA CTATCGCCGCGACATCGTGCGCAAGAGCGGGGGGCGAAAGTAC ACGTACCGCTTCGGGGGCCGCGTGCCACGCTAGCCTATCCGG ACTGTGCGGGAGGCGGACGGGGAGCAGAGACACAA	98
ETV3	ATGAAAGCCGGCTGTAGCATCGTGGAAAAGCCAGAAGGAGGTG GAGGGTATCAGTTTCTGACTGGGCCTACAAAACAGAGTCATC CCCAGGCTCCCGGCAGATCCAGCTGTGGCACTTCATCCTGGAG CTGCTGCAGAAGGAAGAGTTCCGCCATGTCATCGCCTGGCAGC AGGGAGAGTACGGGGAATTTGTTCATCAAGGATCCAGATGAGGT GGCCCGCCTCTGGGGCCGACGAAATGCAAACCACAGATGAAT TATGACAAGCTGAGCCGGGCCCTCAGATACTATTACAACAAGA GGATCCTTCATAAAACAAAAGGGAAAAGATTTACCTATAAATT TAACTTCAACAAGCTGGTGATGCCCCAATACCCATTTCATCAAC ATTCCGTCAAGTGGTAAGATACAACTCTTTTGGTAGGGAAT	99
RUNX1	ATGGCTTCAGACAGCATATTTGAGTCATTTCTTCGTACCCAC AGTGCTTCATGAGAGAATGCATACTTGGAAATGAATCCTTCTAG AGACGTCCACGATGCCAGCACGAGCCGCCGCTTCACGCCGCCT TCCACCGCGCTGAGCCCAGGCAAGATGAGCGAGGCGTTGCCGC TGGGCGCCCCGGACGCCGCGCTGCCCTGGCCGGCAAGCTGAG GAGCGGCGACCGCAGCATGGTGGAGGTGCTGGCCGACCAACCG GGCGAGCTGGTGCGCACCGACAGCCCCAATTCCTCTGCTCCG TGCTGCCTACGCACTGGCGCTGCAACAAGACCCTGCCCATCGC TTTCAAGGTGGTGGCCCTAGGGGATGTTCCAGATGGCACTCTG GTCATGTGATGGCTGGCAATGATGAAAATACTCGGCTGAGC TGAGAAATGCTACCGCAGCCATGAAGAACCAGGTTGCAAGATT TAATGACCTCAGGTTTGTGCGTCGAAGTGGAAGAGGGAAAAGC TTCACCTTGACCATCACTGTCTTCACAAACCCACCGCAAGTCG CCACCTACCACAGAGCCATCAAAATCACAGTGATGGGCCCCG AGAACCTCGAAGACATCGGCAGAACTAGATGATCAGACCAAG CCCGGGAGCTTGTCTTTTCCGAGCGGCTCAGTGAACTGGAGC AGCTGCGGCGCACAGCCATGAGGGTCAGCCACACCAACCCAGC CCCCACGCCCCAACCTCGTGCTTCCCTGAACCACTCCACTGCC TTTAACCTCAGCCTCAGAGTCAGATGCAGGATACAAGGCAGA TCCAACCATCCCCACCGTGGTCTACGATCAGTCTACCAATA CCTGGGATCCATTGCCTCTCTTCTGTGCACCCAGCAACGCCC ATTTACCTGGACGTGCCAGCGGCATGACAACCTCTCTGCAG AACTTTCCAGTCGACTCTCAACGGCACCCGACCTGACAGCGTT CAGCGACCCGCGCCAGTTCCCCGCGCTGCCCTCCATCTCCGAC CCCCGCATGCACTATCCAGGCGCCTTCACCTACTCCCCGACGC CGGTCACCTCGGGCATCGGCATCGGCATGTGGCCATGGGCTC	100

	GGCCACGCGCTACCACACCTACCTGCCGCCGCCCTACCCCGGC TCGTCGCAAGCGCAGGGAGGCCCGTTCCAAGCCAGCTCGCCCT CCTACCACCTGTACTACGGCGCCTCGGCCGGCTCCTACCAGTT CTCCATGGTGGGCGGCGAGCGCTCGCCGCCGCGCATCCTGCCG CCCTGCACCAACGCCTCCACCGGCTCCGCGCTGCTCAACCCCA GCCTCCCGAACCAGAGCGACGTGGTGGAGGCCGAGGGCAGCCA CAGCAACTCCCCACCAACATGGCGCCCTCCGCGCGCCTGGAG GAGGCCGTGTGGAGGCCCTAC	
RUNX2	ATGGCATCAAACAGCCTCTTCAGCACAGTGACACCATGTCAGC AAAATTCTTTTGGGATCCGAGCACCAGCCGGCGCTTCAGCCC CCCCTCCAGCAGCCTGCAGCCCGGCAAAATGAGCGACGTGAGC CCGTGGTGGCTGCGCAACAGCAGCAGCAACAGCAGCAGCAGC AACAGCAGCAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGGA GGCGGCGGCGGCGGCTGCGCGGCGGCGGCGGCTGCGGCGGCG GCAGCTGCAGTGCCCCGGTTGCGGCCGCCCCACGACAACCGCA CCATGGTGGAGATCATCGCCGACCACCCGGCCGAACCTCGTCCG CACCGACAGCCCCAACTTCCTGTGCTCGGTGCTGCCCTCGCAC TGGCGCTGCAACAAGACCTGCCCCGTGGCCTTCAAGGTGGTAG CCCTCGGAGAGGTACCAGATGGGACTGTGGTTACTGTCTATGGC GGGTAACGATGAAAATTATTCTGCTGAGCTCCGGAATGCCTCT GCTGTTATGAAAAACCAAGTAGCAAGGTTCAACGATCTGAGAT TTGTGGGCCGGAGTGACGAGGCAAGAGTTTACCTTGACCAT AACCGTCTTCACAAATCCTCCCCAAGTAGCTACCTATCACAGA GCAATTAAAGTTACAGTAGATGGACCTCGGGAACCCAGAAGGC ACAGACAGAAGCTTGATGACTCTAAACCTAGTTTGTCTCTGA CCGCCTCAGTGATTTAGGGCGCATTCCTCATCCCAGTATGAGA GTAGGTGTCCCGCCTCAGAACCCACGGCCCTCCCTGAACTCTG CACCAGTCCCTTTTAATCCACAAGGACAGAGTCAGATTACAGA CCCCAGGCAGGCACAGTCTTCCCGCCGTGGTCTATGACCAG TCTTACCCCTCCTACCTGAGCCAGATGACGTCCCCGTCCATCC ACTCTACCACCCCGCTGTCTTCCACACGGGGCACTGGGCTTCC TGCCATCACCGATGTGCCTAGGCGCATTTTCAGGTGCTTCAGAA CTGGGCCCTTTTTTCAGACCCAGGCAGTTCCCAAGCATTTTCAT CCCTCACTGAGAGCCGCTTCTCCAACCCACGAATGCACTATCC AGCCACCTTTACTTACACCCGCCAGTCACCTCAGGCATGTCC CTCGGTATGTCCGCCACCACTCACTACCACACCTACCTGCCAC CACCTACCCCGGCTCTTCCCAAAGCCAGAGTGGAACCTTCCA GACCAGCAGCACTCCATATCTCTACTATGGCACTTCGTGAGGA TCCTATCAGTTTCCCATGGTGCCGGGGGAGACCGGTCTCCTT CCAGAATGCTTCCGCCATGCACCAACCTCGAATGGCAGCAC GCTATTAAATCCAAATTTGCCAACCAGAATGATGGTGTGAC GCTGATGGAAGCCACAGCAGTTCCCCAACTGTTTTGAATTCTA GTGGCAGAATGGATGAATCTGTTTGGCGACCATAT	101
RUNX3	ATGCGTATTCCCGTAGACCCAAGCACCAGCCGCCGCTTCACAC CTCCCTCCCCGGCCTTCCCCTGCGGCGGCGGCGGCGGCAAGAT GGGCGAGAACAGCGGCGCGCTGAGCGCGCAGGCGGCGGCGTGGG CCCGGAGGGCGCGCCCGGCCGAGGTGCGCTCGATGGTGGACG TGCTGGCGGACCACGCAGGCGAGCTCGTGCGCACCGACAGCCC CAACTTCCTCTGCTCCGTGCTGCCCTCGCACTGGCGCTGCAAC AAGACGCTGCCCCGTGCGCTTCAAGGTGGTGGCATTGGGGGACG TGCCGGATGGTACGGTGGTGAAGTGTGATGGCAGGCAATGACGA GAACTACTCCGTGAGCTGCGCAATGCCTCGGCCGTGATGAAG AACCAGGTGGCCAGGTTCAACGACCTTCGCTTCGTGGGCCGCA GTGGGCGAGGGAAGAGTTTCAACCTGACCATCACTGTGTTTAC CAACCCACCCAAGTGGCGACCTACCACCGAGCCATCAAGGTG	102

	ACCGTGGACGGACCCCGGGAGCCCAGACGGCACCGGCAGAAGC TGGAGGACCAGACCAAGCCGTTCCCTGACCGCTTTGGGGACCT GGAACGGCTGCGCATGCGGGTGACACCGAGCACACCCAGCCCC CGAGGCTCACTCAGCACCACAAGCCACTTCAGCAGCCAGCCCC AGACCCCAATCCAAGGCACCTCGGAACTGAACCCATTCTCCGA CCCCCGCCAGTTTGACCGCTCCTTCCCCACGCTGCCAACCCCTC ACGGAGAGCCGCTTCCCAGACCCAGGATGCATTATCCCGGGG CCATGTCAGCTGCCTTCCCCTACAGCGCCACGCCCTCGGGCAC GAGCATCAGCAGCCTCAGCGTGCGGGGCATGCCGGCCACCAGC CGCTTCCACCATACTACCTCCCGCCACCCCTACCCGGGGGGCCC CGCAGAACCAGAGCGGGCCCTTCCAGGCCAACCCGTCCCCCTA CCACCTCTACTACGGGACATCCTCTGGCTCCTACCAGTTCTCC ATGGTGGCCGGCAGCAGCAGTGGGGGCGACCGCTCACCTACCC GCATGCTGGCCTCTTGACCAGCAGCGCTGCCTCTGTGCGCCG CGGCAACCTCATGAACCCAGCCTGGGCGGCCAGAGTGATGGC GTGGAGGCCGACGGCAGCCACAGCAACTCACCCACGGCCCTGA GCACGCCAGGCCGCATGGATGAGGCCGTGTGGCGGCCCTAC	
MAFB	ATGGCCGCGGAGCTGAGCATGGGGCCAGAGCTGCCCACCAGCC CGCTGGCCATGGAGTATGTCAACGACTTCGACCTGCTCAAGTT CGACGTGAAGAAGGAGCCACTGGGGCGCGCGGAGCGTCCGGGC AGGCCCTGCACACGCCTGCAGCCAGCCGGCTCGGTGTCTCTCA CACCGCTCAGCACTCCGTGTAGCTCCGTGCCCTCGTCGCCCAG CTTTCAGCCCGACCGAACAGAAGACACACCTCGAGGATCTGTAC TGGATGGCGAGCAACTACCAGCAGATGAACCCCGAGGCGCTCA ACCTGACGCCCCGAGGACGCGGTGGAAGCGCTCATCGGCTCGCA CCCAGTGCCACAGCCGCTGCAAAGCTTCGACAGCTTTCGCGGC GCTCACCACCACCACCATCACCACCACCCCTACCCGCACCACG CGTACCCGGGCGCCGGCGTGGCCACGACGAGCTGGGCCCCGCA CGCTCACCCGCACCATCACCATCATCACCAGCGTCGCCGCCG CCGTCCAGCGCCGCTAGCCCGCGCAACAGCTGCCCCACTAGCC ACCCCGGGCCCCGGGCGCACGCGACGGCCTCGGCGACGGCGGC GGGCGGCAACGGCAGCGTGAGGACCGCTTCTCCGACGACCAG CTCGTGTCCATGTCCGTGCGCGAGCTGAACCGCCACCTGCGGG GCTTCACCAAGGACGAGGTGATCCGCCTGAAGCAGAAGCGCG GACCCCTGAAGAACCAGGGGCTACGCCCAGTCTTGACAGGTATAAA CGCGTCCAGCAGAAGCACCACTGGAGAATGAGAAGACGCAGC TCATTACGACAGGTGGAGCAGCTTAAGCAGGAGGTGTCCCGGCT GGCCCGCGAGAGAGACGCTACAAGGTCAAGTGCGAGAACTC GCCAACTCCGGCTTCAGGGAGGCGGGCTCCACCAGCGACAGCC CCTCCTCTCCCGAGTTCTTTCTG	103
TFAP2C	ATGTTGTGGAATAAACCATAATGTCAAGTACGAAGAGGACT GCGAGGATCGCCACGACGGGAGCAGCAATGGGAATCCGCGGGT CCCCCACCTCTCCTCCGCGGGCAGCACCTCTACAGCCCCGCG CCACCCCTCTCCACACTGGAGTCGCCGAATATCAGCCGCCAC CCTACTTTCCCCCTCCCTACCAGCAGCTGGCCTACTCCAGTC GGCCGACCCCTACTCGCATCTGGGGGAAGCGTACGCCGCCGCC ATCAACCCCTGCACCAGCCGGCGCCCCACAGGCAGCCAGCAGC AGGCCTGGCCCGGCCGCCAGAGCCAGGAGGGAGCGGGGCTGCC CTCGCACCACGGGCGCCCGGCCGCTACTGCCCCACCTCTCC GGGCTGGAGGCGGGCGCGGTGAGCGCCCGCAGGGATGCCTACC GCCGCTCCGACCTGCTGCTGCCCCACGCACACGCCCTGGATGC CGCGGGCCTGGCCGAGAACCTGGGGCTCCACGACATGCCTCAC CAGATGGACGAGGTGCAGAATGTCGACGACCAGCACCTGTTGC TGCACGATCAGACAGTCATTCGCAAAGGTCCCATTTCCATGAC CAAGAACCCTCTGAACCTCCCCTGTGAGAAGGAGCTGGTGGGG	104

	GCCGTAATGAACCCCACTGAGGTCTTCTGCTCAGTCCCTGGAA GATTGTCGCTCCTCAGCTCTACGTCTAAATACAAAGTGACAGT GGCTGAAGTACAGAGGCGACTGTCCCCACCTGAATGCTTAAAT GCCTCGTTACTGGGAGGTGTTCTCAGAAGAGCCAAATCGAAAA ATGGAGGCCGGTCTTGCGGGAGAAGTTGGACAAGATTGGGTT GAATCTTCCGGCCGGGAGGCGGAAAGCCGCTCATGTGACTCTC CTGACATCCTTAGTAGAAGGTGAAGCTGTTTCATTTGGCTAGGG ACTTTGCCCTATGTCTGTGAAGCCGAATTTCTTAGTAAACCAGT GGCAGAATATTTAACCAGACCTCATCTTGGAGGACGAAATGAG ATGGCAGCTAGGAAGACATGCTATTGGCGGCCAGCAACTGT GTAAAGAATTACAGAACTTCTCAGCCAAGACCGGACACCCCA TGGGACCAGCAGGCTCGCCCCAGTCTTGGAGACGAACATACAG AACTGCTTGTCTCATTTTCAGCCTGATTACCCACGGGTTTGGCA GCCAGGCCATCTGTGCCGCGGTGTCTGCCCTGCAGAACTACAT CAAAGAAGCCCTGATTGTCATAGACAAATCCTACATGAACCCT GGAGACCAGAGTCCAGCTGATTCTAACAAAACCCTGGAGAAAA TGGAGAAACACAGGAAA	
E2F1	ATGGCCTTGGCCGGGGCCCTGCGGGCGGCCCATGCGCGCCGG CGCTGGAGGCCCTGCTCGGGGCCGGCGCGCTGCGGCTGCTCGA CTCCTCGCAGATCGTCATCATCTCCGCCGCGCAGGACGCCAGC GCCCCGCCGGCTCCACCGGCCCGCGGCGCCCGCCGCCGGCC CCTGCGACCTGACCTGCTGCTCTTCGCCACACCGCAGGCGCC CCGGCCACACCCAGTGCGCCGCGGCCCGCGCTCGGCCGCCCG CCGGTGAAGCGGAGGCTGGACCTGGAAACTGACCATCAGTACC TGGCCGAGAGCAGTGGGCCAGCTCGGGGCAGAGGCCGCCATCC AGGAAAAGGTGTGAAATCCCCGGGGGAGAAGTCAAGCTATGAG ACCTCACTGAATCTGACCACCAAGCGCTTCCTGGAGCTGCTGA GCCACTCGGCTGACGGTGTGTCGACCTGAACTGGGCTGCCGA GGTGCTGAAGGTGCAGAAGCGCGCATCTATGACATCACCAAC GTCCTTGAGGGCATCCAGCTCATTGCCAAGAAGTCCAAGAACC ACATCCAGTGGCTGGGCAGCCACACCACAGTGGGCGTCGGCGG ACGGCTTGAGGGGTTGACCCAGGACCTCCGACAGCTGCAGGAG AGCGAGCAGCAGCTGGACCACCTGATGAATATCTGTACTACGC AGCTGCGCCTGCTCTCCGAGGACACTGACAGCCAGCGCCTGGC CTACGTGACGTGTGAGGACCTTCGTAGCATTGCAGACCCTGCA GAGCAGATGGTTATGGTGATCAAAGCCCCCTCTGAGACCCAGC TCCAAGCCGTGGACTCTTCGGAGAACTTTCAGATCTCCCTTAA GAGCAAACAAGGCCCGATCGATGTTTTCTGTGCCCTGAGGAG ACCGTAGGTGGGATCAGCCCTGGGAAGACCCATCCCAGGAGG TCACTTCTGAGGAGGAGAACAGGGCCACTGACTCTGCCACCAT AGTGTCACCACCACCATCATCTCCCCCTCATCCCTCACCACA GATCCCAGCCAGTCTCTACTCAGCCTGGAGCAAGAACCGCTGT TGTCCCGGATGGGCAGCCTGCGGGCTCCCGTGGACGAGGACCG CCTGTCCCCGCTGGTGGCGGCCGACTCGCTCCTGGAGCATGTG CGGGAGGACTTCTCCGGCCTCCTCCCTGAGGAGTTTCATCAGCC TTTCCCCACCCACGAGGCCCTCGACTACCACTTCGGCCTCGA GGAGGGCGAGGGCATCAGAGACCTCTTCGACTGTGACTTTGGG GACCTCACCCCCCTGGATTTC	105

The transcription factor DNA-binding domain may, for example, comprise an amino acid sequence that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identity to any one of SEQ ID NOs: 30-67, preferably wherein the amino acid sequence substantially retains the natural function of the protein represented by SEQ ID NO: 30-67, respectively.

The transcription factor DNA-binding domain may, for example, be encoded by a polynucleotide comprising a nucleic acid sequence which encodes a protein that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% amino acid identity to any one of SEQ ID NOs: 30-67, preferably wherein the amino acid sequence substantially retains the natural function of the protein represented by SEQ ID NO: 30-67, respectively.

The transcription factor DNA-binding domain may, for example, be encoded by a polynucleotide comprising a nucleic acid sequence that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% nucleotide identity to any one of SEQ ID NOs: 68-105, preferably wherein the encoded amino acid sequence substantially retains the natural function of the protein encoded by SEQ ID NO: 68-105, respectively.

EXEMPLARY ESFS

In preferred embodiments, the ESF comprises or consists of a KRAB domain, a SOX2 DNA-binding domain, a DNMT3A domain and a DNMT3L domain.

In some embodiments, the ESF comprises or consists of a CS domain and a SOX2 DNA-binding domain.

In some embodiments, the ESF comprises or consists of a SOX2 DNA-binding domain and a Y-R domain.

In some embodiments, the ESF comprises or consists of a KRAB domain, a TEAD1 DNA-binding domain, a DNMT3A domain and a DNMT3L domain.

In some embodiments, the ESF comprises or consists of a KRAB domain, a DNMT3A domain, a DNMT3L domain and a MYC DNA-binding domain.

Example sequences of ESFs of the invention are:

KRAB-hSOX2₁₋₁₇₉-DNMT3a3L

MDAKSLTAWSRITLVTFKDVVFDFTREEWKLLDTAQQILYRNVMLENYKNLVSLGYQLTKPDV
 ILRLEKGEEFWLVEREIHQETHPDSETAFEIKSSVSGSACMYNMMETELKPPGPQQTSGGGG
 GNSTAAAAGGNQKNSPDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSE
 TEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVG
 AGLGAGVNQRMDSYAHMNGWSNGSYSMQDQLGASGSGNHDQEFDPKVPYPPVPAEKRPPIR
 VLSLFDGIATGLLVKLDLGIQVDRIASEVCEDSITVGMVRHQGKIMYVGDVRSVTQKHIOE
 WGPFDLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLDARPKEGDDRPFFWLFENVV
 AMGVSDKRDISRFLSNPVMIDAKEVSAHRARYFWGNLPGMNRPLASTVNDKLELQECLEH
 GRIAKFSKVRTITTRSNSIKQKQDQHFVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRL
 ARQRLGLRSWSVPVIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMET
 YKTVSAWKRPVRVLSLFRNIDKVLKSLGFLES GSGSGGGTLKYVEDVTNVVRRDVEKWGPF
 DLVYGSTQPLGSSCDRCPGWYMFQFHRILQYALPRQESQRPFWFIFMDNLLLTEDDQETTR
 FLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEYLLQAQVRSRSLDAPKVDL
 LVKNCLLPLREYFKYFSQNSLPL

(SEQ ID NO: 106)

KRABhSOX2₁₋₁₇₉*DNMT3a3L*5 CS-hSOX2₁₋₁₇₉

MLEPEKIIIGATDSCGDLMFIMKWKDTDEADLVLAKEANVKCPQIVIAFYEEERLTWHAYPEAC
MYNMMETELKPPGPQQTSGGGGGNSTAAAAGGNQKNSPDRVKRPMNAFMVWSRGQRRKMAQE
NPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKTLMKKDK
YTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLG

10 (SEQ ID NO: 108)

CShSOX2₁₋₁₇₉SOX2₁₋₁₇₉-Y-R

15 MYNMMETELKPPGPQQTSGGGGGNSTAAAAGGNQKNSPDRVKRPMNAFMVWSRGQRRKMAQE
NPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKTLMKKDK
YTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLGASGSNMR
PRLKNVDRSTAQQQLAVTVGNVTVIIITDFKEKTRSSSTSSTVTSSAGSEQQNQ

(SEQ ID NO: 109)

hSOX2₁₋₁₇₉20 **Y-R**KRAB-hTEAD₁₋₁₆₆-DNMT3a3L

25 **MDAKSLTAWSRTLVTFKDVVFVDFTREEWKLLDTAQQILYRNVMLENYKNLVSLGYQLTKPDV**
IILRLEKGEEPWLVEREIHQETHPDSETAFEIKSSVSGSACIEPSSWSGSESPAENMERMSDS
ADKPIDNDAEGVWSPDIEQSFQEALAIYPPCGRRKIILSDEGKMYGRNELIARYIKLRTGKT
RTRKQVSSHQVLARRKSRDFHSLKLDQTAQKALQHMAAMSSAQIVSATAIHKNLGLPGIP
RPTFPGAPGFWPGMIQTGQPASGSGNHDQEFDPKVYPPVPAEKRPPIRVLSLFDGIATGLL
VLKDLGIQVDRIYIASEVCEDSITVGMVRHQKIMYVGDVRSVTQKHIQEWGPFDLVIGGSPC
NDLSIVNPARKGLYEGTGRLFFEFYRLLHDARPKEGDDRPFWWLFENVVAMGVSDKRDISRF
LESNPVMIDAKEVSAHRARYFWGNLPGMNRPLASTVNDKLELQECLEHGRIAKFSKVRTIT
30 TRSNSIKQGDQHFVFMNEKEDILWCTEMERVFGFPVHYTDVSNMSRLARQRLGRSWSVP
VIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPMIYKTVSAWKRPVR
VLSLFRNIDKVLKSLGFLESGSGGGTLKYVEDVTNVVRRDVEKWGPFDLVYGSTQPLGSS
CDRCPGWYMFQFHRILQYALPRQESQRPFFWIFMDNLLLTEDDQETTTTFLQTEAVTLQDVR
GRDYQNAMRVWSNIPGLKSKHAPLTPKEEYLLQAQVRSRSKLDAPKVDLLVKNCLLPLREYF
35 KYFSQNSLPL

(SEQ ID NO: 110)

KRAB

hTEAD1₁₋₁₆₆DNMT3a3LKRAB-DNMT3a3L-hMYC₁₄₄₋₄₅₄

5 **MDAKSLTAWSRTLVTFKDVFDFTREEWKLLDTAQQILYRNVMLENYKNLVSLGYQLTKPDV**
ILRLEKGEEPWLVEREIHQETHPDSETAFEIKSSVASGSGNHDQEFDPKVYPPVPAEKRKP
 IRVLSLFDGIATGLLVKDLGIQVDRIYASEVCEDSITVGMVRHQGKIMYVGDVRSVQKHI
 QEWGPFDFLVIGGSPCNDLSIVNPARKGLYEGTGRLFFEFYRLLDARPKEGDDRPFFWLFEN
 VVAMGVSDKRDISRFLSNPVMIDAKEVSAAHRARYFWGNLPGMNRPLASTVNDKLELQECL
 10 **EHGRIAKFSKVRTITTRSNSIKQGKDQHFVFMNEKEDILWCTEMERVFGFPVHYTDVSNMS**
RLARQRLGRSWSVPVIRHLFAPLKEYFACVSSGNSNANSRGPSFSSGLVPLSLRGSHMGPM
EIYKTVSAWKRPVRVLSLFRNIDKVLKSLGFLES GSGSGGGTLKYVEDVTNVVRDVEKWG
PFDLVYGSTQPLGSSCDRCPGWYMFQFHRI LQYALPRQESQRPFFWIFMDNLLLTEDDQETT
TRFLQTEAVTLQDVRGRDYQNAMRVWSNIPGLKSKHAPLTPKEEY LQAQVRSRSLDAPKV
 15 **DLLVKNCLLPREYFKYFSQNSLPLGAPAAAI IQDCMWSGFSAAAKLVSEKLASYQAARKDS**
GSPNPARGHSVCSTSSLYLQDL SAAASECIDPSVVFPPYPLNDSSSPKSCASQDSSAFSPSSD
SLLSSTESSPQGSPEPLVLHEETPPTSSDSEEEQEDEE EIDVVSVEKRQAPGKRSESGSPS
AGGHSKPPHSPLVLKRCHVSTHQHNYAAPSTRKDY PAAKRVKLDSVRVLRQISNNRKCTSP
RSSDTEENVKRRTHNVLERQRRNELKRSFFALRDQ IPELENNEKAPKVVLKATAYILSVQ
 20 **AEEQKLISEEDLLRKRREQLKHKLEQLRNSCA**

(SEQ ID NO: 111)

KRAB

DNMT3a3LhMYC₁₄₄₋₄₅₄

25 Example nucleotide sequences encoding ESFs of the invention are

KRAB-hSOX2₁₋₁₇₉-DNMT3a3L

ATGGATGCTAAGTCACTGACTGCCTGGTCCCGGACACTGGTGACCTTCAAGGATGTGTTTGT
 GGACTTCACCAGGGAGGAGTGGAAGCTGCTGGACACTGCTCAGCAGATCCTGTACAGAAATG
 TGATGCTGGAGAACTATAAGAACCTGGTTTCTTGGGTTATCAGCTTACTAAGCCAGATGTG
 30 ATCCTCCGGTTGGAGAAGGGAGAAGAGCCCTGGCTGGTGGAGAGAGAAATTCACCAAGAGAC
CCATCCTGATTCAGAGACTGCATTTGAAATCAAATCATCAGTTGGCAGCGGCGCATGCatgt
acaacatgatggagacggagctgaagccgccgggcccgcagcaaaacttcggggggcgggcggc
ggcaactccaccgcgggcgggcgccggcggaaccagaaaaacagcccggaccgcgtcaagcg
gcccattgaatgccttcatggtgtggtcccgcgggcagcggcgcaagatggcccaggagaacc
 35 ccaagatgcacaactcggagatcagcaagcgctgggcgccgagtggaacttttgcggag
acggagaagcgggcggttcacgcagaggctaagcggtgcgagcgctgcacatgaaggagca
cccgattataaataccggccccggcggaacaagaagcgtcatgaagaaggataagtaca
cgctgccccggcgggctgctggccccggcggaataagcatggcgagcggggtcggggtgggc
gccggcctgggcgcggcggtgaaccagcgcatggacagttacgcgcacatgaacggctggag
 40 caacggcgagctacagcatgatgcaggaccagctgggcGCTAGCGGCAGCGGC**AACCATGATC**
AAGAGTTGCATCCTCCTAAAGTGATCCCCCTGTCCCTGCCGAAAAGCGGAAACCGATCAGG

GTTCTCAGTCTCTTCGATGGAATTGCCACAGGCCTCTTGGTCCTTAAGGACCTGGGCATCCA
 AGTgGA_cCGCTACATCGCCTCAGAAGTATGTGA_gGACAGCATCACAGTCGGCATGGTGCGCC
 ACCAGGGGAAAATAATGTACGTCGGCGACGTCAGATCAGTTACTCAGAAACATATTCAGGAA
 TGGGGGCCCTTTTGATCTCGTAATTGGCGGCAGCCCCTGCAATGATCTGAGTATTGTGAACCC
 5 CGCCCGGAAAGGTTTGTACGAAGGTACGGGCAGACTGTTCTTCGAGTTTTACAGACTGTTGC
 ACGACGCACGACCAAAGGAGGGTGAcGAt_cGACCCCTCTTCTGGCTGTTTCGAGAACGTGGTC
 GCTATGGGCGTGTCTGACAAGCGGGACATTTCTAGATTCTTGGAGAGCAATCCAGTGATGAT
 TGATGCAAAGGAAGTAt_{cc}GCTGCCCCACCGCCAGATACTTCTGGGGCAATCTGCCCCGGCA
 TGAATCGACCCCTTGGAAGCACCGTGAAATGATAAGTTGGAGTTGCAGGAGTGTCT_gGAGCAT
 10 GGGCGGATCGCAAAGTTTAGCAAAGGTGCGGACCATCACGACCCGAAGCAATTCAATTAAGCA
 GGGAAAGGACCAACATTTTCCAGTGTTATGAACGAGAAAGAGGA_cATCCTGTGGTG_cACAG
 AAATGGAGCGGGTTTTCGGATTCCCCGTACATTATACTGATGTGTCCAATATGTCTAGGTTG
 GCTAGACAGAGGCTG_tTGGGACGCTCATGGAGTGTTCTGTCTATCCGCCACCTGTTTCGCTCC
 TCTGAAGGAGTATTTTCGC_cTGCGT_tTCATCCGG_gAATTCAAACGCAAACAGCAGAGGCCCAT
 15 CCTTTTCTTCCGGCCTGGT_gCCACTTAGTCTGCGCGGCTCTCACATGGGACCTATGGAAATA
 TACAAAACAGTGAGCGCCTGGAAGAGGCAGCCTGTGCGGGTCTGAGTCTGTTTCAGAAACAT
 TGACAAGGTGCTTAAGTCCCTTGGAATTTCTGGAGTCTGGCAGCGGTTCTGGGGGTG_gcACCC
 TTAAGTATGTGGAGGACGTGACAAACGTGCTGAGGAGAGATGTGGAGAAATGGGGGCCATTT
 GACCTGGTATACGGGAGCACTCAACCTCTCGGTTCTTCTGTGATCGCTGCCCAGGGTGGA
 20 TATGTTTCAGTTCCATCG_cATCCTGCAATACGCCCTTCCGCGGCAGGAGAGTCAGCGACCAT
 TCTTCTGGATATTATGGAACAATCTCCTGCTGACAGAG_gGACGACCAAGAGACTACGACTAGA
 TTTCTTCAGACCGAGGCCGTTACTCTCCAGGACGTTAGAGGTAGGGACTACCAGAATGCCAT
 GCGAGTGTGGAGTAACATCCCCGACTCAAGTCAAAGCA_cGCACCCCTGACCCCCAAGGAAG
 AGGAATACCT_cCAGGCGCAGGTGAGGAGCCGCAGTAAGCTCGATGCTCCTAAAGTCGATTTG
 25 TTGGTGAAGAATTGCTTGCTGCCCCTGAGAGAGTACTTTAAGTACTTTAGCCAGAACAGCTT
 GCCACTG

(SEQ ID NO: 112)

KRAB*hSOX2₁₋₁₇₉*30 DNMT3a3LKRAB-hSOX2₁₋₁₇₉-DNMT3a3L

ATGGATGCTAAGTCACTGACTGCCTGGTCCCGGACACTGGTGACCTTCAAGGATGTGTTTGT
 GGACTTCACCAGGGAGGAGTGGAAGCTGCTGGACACTGCTCAGCAGATCCTGTACAGAAATG
 TGATGCTGGAGAACTATAAGAACCTGGTTTCTTGGGTTATCAGCTTACTAAGCCAGATGTG
 35 ATCCTCCGTTGGAGAAGGGAGAAGAGCCCTGGCTGGTGGAGAGAGAAATTCACCAAGAGAC
 CCATCCTGATTTCAGAGACTGCATTTGAAATCAAATCATCAGTTGGCAGCGGCGCATGCatgt
 acaacatgatggagacggagctgaagccgcccgggcccgcagcaaaacttcggggggcgggcggc
 ggcaactccaccgcgggcgggcgccggcggaaccagaaaaacagcccggaccgctcaagcg
 gcccattgaatgccttcattggtgtggtcccgcgggcagcggcgaagatggcccaggagaacc
 40 ccaagatgcacaactcggagatcagcaagcgccctgggcgcccagtggaacttttgcggag
 acggagaagcgggcgttcacgcagaggttaagcggtgcgagcgctgcacatgaaggagca
 cccgattataaataaccggccccggcggaacaagaagcgcacatgaagaaggataagtaca
 cgctgcccggcgggctgctggccccggcggaatagcatggcgagcggggtcggggtgggc
 gccggcctgggcgcgggcgtgaaccagcgcattggacagttacgcgcacatgaacggctggag

caacggcagctacagcatgatgcaggaccagctgggcGCTAGCGGCAGCGGCAACCATGATC
AAGAGTTTCGATCCTCCTAAAGTGATCCCCCTGTCCCTGCCGAAAAGCGGAAACCGATCAGG
GTTCTCAGTCTCTTCGATGGAATTGCCACAGGCCTCTTGGTCCTTAAGGACCTGGGCATCCA
AGTgGAcCGCTACATCGCCTCAGAAGTATGTGA_gGACAGCATCACAGTCGGCATGGTGCGCC
 5 ACCAGGGGAAAATAATGTACGTGCGCGACGTCAGATCAGTTACTCAGAAACATATTTCAGGAA
TGGGGGCCTTTTGATCTCGTAATTGGCGGCAGCCCCTGCAATGATCTGAGTATTGTGAACCC
CGCCCGGAAAGGTTTGTACGAAGGTACGGGCAGACTGTTCTTCGAGTTTTACAGACTGTTGC
ACGACGCACGACCAAAGGAGGGTGAcGAtcGACCCTTCTTCTGGCTGTTTCGAGAACGTGGTC
 10 GCTATGGGCGTGCTGACAAGCGGGACATTTctAGATTCTTGGAGAGCAATCCAGTGATGAT
TGATGCAAAGGAAGTatccGCTGCCCCACCGCGCCAGATACTTCTGGGGCAATCTGCCCCGGCA
TGAATCGACCCCTTGGCAAGCACCGTGAAATGATAAGTTGGAGTTGCAGGAGTGTCTgGAGCAT
GGGCGGATCGCAAAGTTTAGCAAGGTGCGGACCATCACGACCCGAAGCAATTCAATTAAGCA
GGGAAAGGACCAACATTTTCCAGTGTTTATGAACGAGAAAGAGGAcATCCTGTGGTGcACAG
AAATGGAGCGGGTTTTCGGATTCCCCGTACATTATACTGATGTGTCCAATATGTCTAGGTTG
 15 GCTAGACAGAGGCTGtTGGGACGCTCATGGAGTGTTCTGTCTATCCGCCACCTGTTTCGCTCC
TCTGAAGGAGTATTTTCGCcTGCGTtTCATCCGGgAATTCAAACGCAAACAGCAGAGGCCCAT
CCTTTTCTTCCGGCCTGGTgCCACTTAGTCTGCGCGGCTCTCACATGGGACCTATGGAAATA
TACAAAACAGTGAGCGCCTGGAAGAGGCAGCCTGTGCGGGTCTGAGTCTGTTTCAGAAACAT
TGACAAGGTGCTTAAGTCCCTTGGATTTCTGGAGTCTGGCAGCGGTTCTGGGGGTGgcACCC
 20 TTAAGTATGTGGAGGACGTGACAAACGTCGTGAGGAGAGATGTGGAGAAATGGGGGCCATTT
GACCTGGTATACGGGAGCACTCAACCTCTCGGTTCTTCTTGTGATCGCTGCCAGGGTGgTA
TATGTTTTCAGTTCCATCGcATCCTGCAATACGCCCTTCCGCGGCAGGAGAGTCAGCGACCAT
TCTTCTGGATATTTCATGGACAATCTCCTGCTGACAGAgGACGACCAAGAGACTACGACTAGA
TTTCTTCAGACCGAGGCCGTTACTCTCCAGGACGTTAGAGGTAGGGACTACCAGAATGCCAT
 25 GCGAGTGTGGAGTAACATCCCCGGA_{CT}CAAGTCAAAGCAcGCACCCCTGACCCCCAAGGAAG
AGGAATACCTcCAGGCGCAGGTGAGGAGCCGAGTAAGCTCGATGCTCCTAAAGTCGATTTG
TTGGTGAAGAATTGCTTGCTGCCCCTGAGAGAGTACTTTAAGTACTTTAGCCAGAACAGCTT
GCCACTG

(SEQ ID NO: 113)

30 KRAB
hSOX2₁₋₁₇₉
DNMT3a3L

CS-hSOX2₁₋₁₇₉

35 ATGCTGGAACCAGAAAAGATCATTGGGGCAACAGATTCCTGTGGTGATTTAATGTTTCCTAAT
GAAATGGAAAGACACAGATGAAGCTGACCTGGTTCTTGCAAAAGAAGCTAATGTGAAATGTC
CACAAATTGTGATAGCATTTTATGAAGAGAGACTGACATGGCATGCATATCCTGAGGCATGC
atgtacaacatgatggagacggagctgaagccgccgggcccgcagcaaaacttcggggggcg
cggcggaactccaccgcgggcgccggcggaaccagaaaaacagcccggaccgctca
 40 *agcggcccacatgaatgccttcattggtgtggtcccgcgggcagcggcgcaagatggcccaggag*
aaccccaagatgcacaactcggagatcagcaagcgccctgggcgcccagtggaacttttgtc
ggagacggagaagcggccgttcacgcagaggctaagcgggtgcgagcgctgcacatgaagg
agcaccgcgattataaataaccggccccggcggaacccaagacgctcatgaagaaggataag
tacacgctgccggcgggctgctggccccggcggaatagcatggcgagcggggtcggggt

gggcgcgcgcctgggcgcggcgctgaaccagcgcattgacagttacgcgcacatgaacggct
ggagcaacggcagctacagcatgatgcaggaccagctgggc

(SEQ ID NO: 114)

CS

5 *hSOX2₁₋₁₇₉*

SOX2₁₋₁₇₉-Y-R

atgtacaacatgatggagacggagctgaagccgcggccgcagcaaaacttcggggggcg
cggcggaactccaccgcggcgccgcggcgcaaccagaaaaacagcccgaccgcgtca
agcggcccatgaatgccttcattggtgtggtcccgcgggcagcggcgcaagatggccaggag
10 aacccaagatgcacaactcggagatcagcaagcgcctgggcgccgagtggaacttttgtc
ggagacggagaagcggccgttcacgcagaggctaagcggctgcgagcgtgcacatgaagg
agcaccgcgattataaataccggccccgcgggaaaaaccaagcgtcatgaagaaggataag
tacacgtctcccggcgggctgctggccccgcggcgcaatagcatggcgagcgggtcggggt
gggcgcgcgcctgggcgcggcgctgaaccagcgcattgacagttacgcgcacatgaacggct
15 ggagcaacggcagctacagcatgatgcaggaccagctgggcgctagcGGATCCaacATGAGG
CCCCGGCTGAAAAACGTGGACAGGAGCACTGCACAGCAGTTGGCAGTAACGTGTGGGCAACGT
CACCGTCATTATCACAGACTTTAAGGAAAAAGACTCGCTCCTCATCGACATCCTCATCCACAG
TGACCTCCAGTGCAGGGTCAGAACAGCAGAACCAG

(SEQ ID NO: 115)

20 *hSOX2₁₋₁₇₉*

Y-R

KRAB-hTEAD₁₋₁₆₆-DNMT3a3L

ATGGATGCTAAGTCACTGACTGCCTGGTCCCGGACACTGGTGACCTTCAAGGATGTGTTTGT
GGACTTCACCAGGGAGGAGTGGAAGCTGCTGGACACTGCTCAGCAGATCCTGTACAGAAATG
25 TGATGCTGGAGAACTATAAGAACCTGGTTTCCTTGGGTTATCAGCTTACTAAGCCAGATGTG
ATCCTCCGTTTGAGAAAGGAGAGCCCTGGCTGGTGGAGAGAGAAATTCACCAAGAGAC
CCATCCTGATTGAGAGACTGCATTTGAAATCAAATCATCAGTTGGCAGCGGCGCATGCATTG
AGCCAGCAGCTGGAGCGGCAGTGAGAGCCCTGCCGAAAACATGGAAAGGATGAGTGACTCT
GCAGATAAGCCAATTGACAATGATGCAGAAGGGTCTGGAGCCCCGACATCGAGCAAAGCTT
30 TCAGGAGGCCCTGGCTATCTATCCACCATGTGGGAGGAGGAAAATCATCTTATCAGACGAAG
GCAAAATGTATGGTAGGAATGAATTGATAGCCAGATACATCAAATCAGGACAGGCAAGACG
AGGACCAGAAAACAGGTGTCTAGTCACATTCAGGTTCTTGCCAGAAGGAAATCTCGTGATTT
TCATTCCAAGCTAAAGGATCAGACTGCAAAGGATAAGGCCCTGCAGCACATGGCGGCCATGT
CCTCAGCCCAGATCGTCTCGGCCACTGCCATTATATAACAAGCTGGGGCTGCCTGGGATTCCA
35 CGCCCGACCTTCCAGGGGCGCCGGGGTTCTGGCCGGGAATGATTCAAACAGGGCAGCCAGC
TAGCGGCAGCGGC**AACCATGATCAAGAGTTTCGATCCTCCTAAAGTGTATCCCCCTGTCCCTG**
CCGAAAAGCGGAAACCGATCAGGGTTCTCAGTCTCTTCGATGGAATTGCCACAGGCCCTTGTG
GTCTTTAAGGACCTGGGCATCCAAGTgGAACGCTACATCGCCTCAGAAGTATGTGAAGACAG
CATCACAGTCGGCATGGTGCGCCACCAGGGGAAATAATGTACGTCGGCGACGTCAGATCAG
40 **TTACTCAGAAACATATTCAGGAATGGGGGCCCTTTTGATCTCGTAATTGGCGGCAGCCCTGCG**
AATGATCTGAGTATTGTGAACCCCGCCCGAAAGGTTGTACGAAGGTACGGGCAGACTGTT

CTTTCGAGTTTTACAGACTGTTGCACGACGCACGACCAAGGAGGGTGA_cGAT_cGACCCTTCT
 TCTGGCTGTTTCGAGAACGTGGTCGCTATGGGCGTGTCTGACAAGCGGGACATTTCTAGATTCT
 CTGGAGAGCAATCCAGTGATGATGATGCAAAGGAAGTAT_ccGCTGCCACCGCGCCAGATA
 CTTCTGGGGCAATCTGCCCGGCATGAATCGACCCCTGGCAAGCACCGTGAATGATAAGTTGG
 5 AGTTGCAGGAGTGTCT_gGAGCATGGGCGGATCGCAAAGTTTAGCAAGGTGCGGACCATCACG
 ACCCGAAGCAATTCAATTAAGCAGGGAAAGGACCAACATTTTCAGTGTTTATGAACGAGAA
 AGAGGA_cATCCTGTGGTG_cACAGAAATGGAGCGGGTTTTTCGGATTCCCCGTACATTATACTG
 ATGTGTCCAATATGTCTAGGTTGGCTAGACAGAGGCTGT_tGGGACGCTCATGGAGTGTTCCT
 10 GTCATCCGCCACCTGTTTCGCTCCTCTGAAGGAGTATTTTCGC_cTGCGT_tTCATCCGG_gAATTC
 AAACGCAAACAGCAGAGGCCCATCCTTTTCTTCCGGCCTGGT_gCCACTTAGTCTGCGCGGCT
 CTCACATGGGACCTATGGAAATATACAAAACAGTGAGCGCCTGGAAGAGGCAGCCTGTGCGG
 GTCCTGAGTCTGTTCAGAAACATTGACAAGGTGCTTAAGTCCCTTGGATTCTGGAGTCTGG
 CAGCGGTTCTGGGGGTG_gcACCCTTAAGTATGTGGAGGACGTGACAAACGTCGTGAGGAGAG
 ATGTGGAGAAATGGGGGCCATTTGACCTGGTATACGGGAGCACTCAACCTCTCGGTTCTTCT
 15 TGTGATCGCTGCCCAGGGTGGTATATGTTTCAGTTCCATCG_cATCCTGCAATACGCCCTTCC
 GCGGCAGGAGAGTCAGCGACCATCTCTCGGATATTCATGGACAATCTCCTGCTGACAGAG_gG
 ACGACCAAGAGACTACGACTAGATTTCTTCAGACCGAGGCCGTTACTCTCCAGGACGTTAGA
 GGTAGGGACTACCAGAATGCCATGCGAGTGTGGAGTAACATCCCCGGACTCAAGTCAAAGCA
_cGACCCCTGACCCCAAGGAAGAGGAATACCT_cCAGGCGCAGGTGAGGAGCCGCAGTAAGC
 20 TCGATGCTCCTAAAGTCGATTTGTTGGTGAAGAATTGCTTGCTGCCCCCTGAGAGAGTACTTT
 AAGTACTTTAGCCAGAACAGCTTGCCACTG

(SEQ ID NO: 116)

KRAB*hTEAD₁₋₁₆₆*

25 DNMT3a3L

KRAB-DNMT3a3L-hMYC₁₄₄₋₄₅₄

ATGGATGCTAAGTCACTGACTGCCTGGTCCCGGACACTGGTGACCTTCAAGGATGTGTTTGT
 GGA_cTTCA_cAGGGAGGAGTGGAAGCTGCTGGACACTGCTCAGCAGATCCTGTACAGAAATG
 30 TGATGCTGGAGAACTATAAGAACCTGGTTTCTTGGGTTATCAGCTTACTAAGCCAGATGTG
 ATCCTCCGGTTGGAGAAGGGAGAAGAGCCCTGGCTGGTGGAGAGAGAAATTCACCAAGAGAC
 CCATCCTGATTGAGAGACTGCATTTGAAATCAAATCATCAGTTGCTAGCGGCAGCGGC_{AACC}
 ATGATCAAGAGTTTCGATCCTCCTAAAGTGTATCCCCCTGTCCCTGCCGAAAAGCGGAA_{ACCG}
 ATCAGGGTTCTCAGTCTCTTCGATGGAAATTGCCACAGGCCTCTTGGTCTTAAAGACCTGGG
 CATCCAAGT_gGA_cCGCTACATCGCCTCAGAAGTATGTGA_gGACAGCATCACAGTCGGCATGG
 35 TCGCCACCAAGGGGAAAATAATGTACGTCGGCGACGTGAGATCAGTTACTCAGAAACATATT
 CAGGAATGGGGGCCTTTTGATCTCGTAATTGGCGGCAGCCCTGCAATGATCTGAGTATTGT
 GAACCCCGCCCGGAAAGGTTTGTACGAAGGTACGGGCAGACTGTTCTTCGAGTTTTACAGAC
 TGTTGCACGACGCACGACCAAGGAGGGTGA_cGAT_cGACCCCTTCTTCTGGCTGTTTCGAGAAC
 GTGGTCGCTATGGGCGTGTCTGACAAGCGGGACATTTCTAGATTCTGGAGAGCAATCCAGT
 40 GATGATTGATGCAAAGGAAGTAT_ccGCTGCCACCGCGCCAGATACTTCTGGGGCAATCTGC
 CCGGCATGAATCGACCCTTGGCAAGCACCGTGAATGATAAGTTGGAGTTGCAGGAGTGTCT_g
 GAGCATGGGCGGATCGCAAAGTTTAGCAAGGTGCGGACCATCACGACCCGAAGCAATTCAAT
 TAAGCAGGGAAAGGACCAACATTTTCCAGTGTTTATGAACGAGAAAGAGGA_cATCCTGTGGT
 G_cACAGAAATGGAGCGGGTTTTTCGGATTCCCCGTACATTATACTGATGTGTCCAATATGTCT

5 AGGTTGGCTAGACAGAGGCTGtTGGGACGCTCATGGAGTGTTCTGTCTATCCGCCACCTGTT
 CGCTCCTCTGAAGGAGTATTTTCGcTGCGTtTCATCCGGgAATTCAAACGCAAACAGCAGAG
 GCCCATCCTTTTCTTCCGGCCTGGTgCCACTTAGTCTGCGCGGCTCTCACATGGGACCTATG
 GAAATATACAAAACAGTGAGCGCCTGGAAGAGGCAGCCTGTGCGGGTCTGAGTCTGTTCAG
 AACATTGACAAGGTGCTTAAGTCCCTTGGATTCTTGGAGTCTGGCAGCGGTTCTGGGGGTG
 10 gcACCCTTAAGTATGTGGAGGACGTGACAAACGTCGTGAGGAGAGATGTGGAGAAATGGGGG
 CCATTGACCTGGTATACGGGAGCACTCAACCTCTCGGTTCTTCTTGTGATCGCTGCCCAGG
 GTGGTATATGTTTCAGTTCATCGcATCCTGCAATACGCCCTTCCGCGGCAGGAGAGTACAGC
 GACCATTCTTCTGGATATTTCATGGACAATCTCCTGCTGACAGAgGACGACCAAGAGACTACG
 15 ACTAGATTTCTTCAGACCGAGGCCGTTACTCTCCAGGACGTTAGAGGTAGGGACTACCAGAA
 TGCCATGCGAGTGTGGAGTAACATCCCCGGACTCAAGTCAAAGCAcGCACCCCTGACCCCCA
 AGGAAGAGGAATACCTcCAGGCGCAGGTGAGGAGCCGCAGTAAGCTCGATGCTCCTAAAGTC
 GATTTGTTGGTGAAGAATTGCTTGCTGCCCCCTGAGAGAGTACTTTAAGTACTTTAGCCAGAA
 CAGCTTGCCACTGGGCGCGCCCGCCGCTGCCATCATCCAGGACTGTATGTGGAGCGGCTTCT
 20 CGGCCCGCCGCAAGCTCGTCTCAGAGAAGCTGGCCTCCTACCAGGCTGCGCGCAAAGACAGC
 GGCAGCCCGAACCCCGCCCGGCCACAGCGTCTGCTCCACCTCCAGCTTGACCTGCAGGA
 TCTGAGCGCCGCGCCTCAGAGTGCATCGACCCCTCGGTGGTCTTCCCCTACCCTCTCAACG
 ACAGCAGCTCGCCCAAGTCTGCGCCTCGCAAGACTCCAGCGCCTTCTCTCCGTCTCGGAT
 TCTCTGCTCTCCTCGACGGAGTCTCCCCGAGGGCAGCCCCGAGCCCCTGGTGTCTCATGA
 25 GGAGACACCGCCACCACCAGCAGCGACTCTGAGGAGGAACAAGAAGATGAGGAAGAAATCG
 ATGTTGTTTCTGTGAAAAGAGGCAGGCTCCTGGCAAAAGGTGAGAGTCTGGATCACCTTCT
 GCTGGAGGCCACAGCAAACCTCCTCACAGCCCACTGGTCTCAAGAGGTGCCACGTCTCCAC
 ACATCAGCACAACTACGCAGCGCCTCCCTCCACTCGGAAGGACTATCCTGTGCAAGAGGG
 TCAAGTTGGACAGTGTGAGAGTCTGAGACAGATCAGCAACAACCGAAAATGCACCAGCCCC
 30 AGGTCTCTCGGACACCGAGGAGAAATGTCAAGAGGCGAACACACAACGTCTTGGAGCGCCAGAG
 GAGGAACGAGCTAAAACGGAGCTTTTTTGCCCTGCGTGACCAGATCCCGGAGTTGGAAAACA
 ATGAAAAGGCCCCCAAGGTAGTTATCCTTAAAAAAGCCACAGCATAATCCTGTCCGTCCAA
 GCAGAGGAGCAAAAGCTCATTTCTGAAGAGGACTTGTTGCGGAAACGACGAGAACAGTTGAA
 ACACAAACTTGAACAGCTACGGAACCTTGTGCG

(SEQ ID NO: 117)

KRAB

DNMT3a3L

hMYC₁₄₄₋₄₅₄

The ESF may, for example, comprise or consist of an amino acid sequence that has at least
 35 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% identity to any one of SEQ ID NOs:
 106-111, preferably wherein the amino acid sequence substantially retains the natural
 function of the protein represented by SEQ ID NO: 106-111, respectively.

The ESF may, for example, be encoded by a polynucleotide comprising or consisting of a
 nucleic acid sequence which encodes a protein that has at least 70%, 80%, 90%, 95%, 96%,
 40 97%, 98%, 99% or 100% amino acid identity to any one of SEQ ID NOs: 106-111, preferably

wherein the amino acid sequence substantially retains the natural function of the protein represented by SEQ ID NO: 106-111, respectively.

The ESF may, for example, be encoded by a polynucleotide comprising a nucleic acid sequence that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% nucleotide identity to any one of SEQ ID NOs: 112-117, preferably wherein the encoded amino acid sequence substantially retains the natural function of the protein encoded by SEQ ID NO: 112-117, respectively.

EXPRESSION CONTROL SEQUENCES

The polynucleotide of the invention may comprise one or more expression control sequence. Suitably, the nucleic acid sequence encoding the ESF is operably linked to one or more expression control sequence.

As used herein an “expression control sequence” is any nucleotide sequence which controls expression of a transgene, e.g. to facilitate and/or increase expression in some cell types and/or decrease expression in other cell types.

The expression control sequence and the nucleic acid sequence encoding the ESF may be in any suitable arrangement in the polynucleotide, providing that the expression control sequence is operably linked to the nucleic acid sequence encoding the ESF.

In some embodiments, the expression control sequence is a promoter.

Any suitable promoter may be used, the selection of which may be readily made by the skilled person. The promoter sequence may be constitutively active (i.e. operational in any host cell background), or alternatively may be active only in a specific host cell environment, thus allowing for targeted expression of the nucleotide of interest (e.g. the ESF) in a particular cell type (e.g. a tissue-specific promoter). The promoter may show inducible expression in response to presence of another factor, for example a factor present in a host cell. In any event, where the vector is administered for therapy, it is preferred that the promoter should be functional in the target cell (e.g. cancer cell) background.

In some embodiments, the polynucleotide further comprises a promoter operably linked to the nucleic acid sequence encoding the ESF.

In some embodiments, the promoter is a constitutive promoter. In some embodiments, the constitutive promoter is an Ef1a promoter.

An example sequence of an Ef1a promoter is:

5 CGATAAGCTTTGCAAAGATGGATAAAGTTTTAAACAGAGAGGAATCTTTGCAGCTAATGGAC
 CTTCTAGGTCTTGAAAGGAGTGGGAATTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACAT
 CGCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTGGCAATTGAACCGGTGCCCTAGAGAAGG
 TGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGG
 GGGAGAACCGTATATAAGTGACGTAGTCGCCGTGAACGTTCTTTTTTCGCAACGGGTTTGCCG
 CCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTATGGC
 CCTTGCGTGCCTTGAATTACTTCCACTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGG
 GTTGGAAGTGGGTGGGAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCCTTCGCCCTCGTGCTTG
 10 AGTTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCTGCGAATCTGGTGGCACCTTCGCGCCT
 GTCTCGCTGCTTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTT
 TTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTTCGGTTTT
 TGGGGCCGCGGGCGCGACGGGGCCCGTGCCTCCAGCGCACATGTTTCGGCGAGGCGGGGCC
 TGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTG
 CCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTGCGCAC
 15 AGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAATGGAGGA
 CGCGGCGCTCGGGAGAGCGGGCGGGTGAGTACCCACACAAAGGAAAAGGGCCTTTCCGTCC
 TCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTT
 CTCGAGCTTTTGAGTACGTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTC
 CCCACACTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCCTCTTG
 20 GAATTTGCCCTTTTTGAGTTTGGATCTTGGTTTATTCTCAAGCCTCAGACAGTGGTTCAAAG
 TTTTTTCTTCCATTTTCAGGTGTCGTGAGGAATTCTGCAGTCGATCGACGGTACCGCGGGCC
 C

(SEQ ID NO: 118)

In some embodiments, the promoter is a tissue-specific promoter, preferably a cancer cell-
 25 specific promoter.

In some embodiments, the promoter is a proliferating cell-specific promoter.

In some embodiments, the promoter is selected from the group consisting of a Mki67
 promoter, a Ccnd1 promoter, a Ccnb2 promoter, a Ccna2 promoter, a Cdc25c promoter, a
 Cdc2 promoter, a Cks1 promoter, a PCNA promoter, a CDC6 promoter, a POLD1 promoter,
 30 a CSK1B promoter, a MCM2 promoter and a PLK1 promoter.

In some embodiments, the promoter is a Mki67 promoter.

An example sequence of a Mki67 promoter is:

35 acttcaggccttgctccttgagaccttgaggagagggtagacattggttctattgccccagaag
 cggaatTTTTTaaacaccctactggaagaacactgatccaggaacaggaggaggagcatga
 aaaaagaggacaggaaaacataggtactctatatctcatctgactggcacctggcctcagata
 accccaactggtgccccaggagccctatTTTtaggatttgcaggtgcagtcttgcacaggTTA
 tTTtaatcggcatttccccctgaaacagcaaagcattgcaaacttaaccgtgccgcagcaat
 aactcTTaagttatggtttcggaaaaaccgtgccactgcctcagtgttcattaagcaggagg
 40 actgtagaatgaattataatgatgtccccagtttttaaggacagtttgagggatactgaa
 aaggcaggagaaacaaccttaggttgctgatgaggtgttgagaaacggcagaggaagggtta
 taaatTTTTaaatagaaagcttctagcgttgagcaagatgcggcggttgctggtgtttgatt
 tctggagagggagcctacaggtagcttctacaggtcgccacgctagttttgctgacactga
 aaaagctgggaccctctaccctgcctattagagcagcagtgggccgcccaccagccccacc
 tggccgccttccggagcttatctccgctcggttccgcttctccagccagcttctggtcccg

5 cccgcctaccggcctggccctccaccttcgggtccacccccccccccccccaggt
 cgcagcctggccccacctctggacttcccccaatcctctgtcgccctcgccattggtctctgg
 ctggcgcctggccccaatcgagcaacttagcgccagaatttgaaccgcggttttggttgaa
 tcggggcgggctgcggcgggagccttggcgcggaaccgcctagggcgacagtgcgggcttg
 acggggcggcgggtggcgggtggcgggcaacttgggacgcggcgggcgaggagtcgggtctg
 tcggggcggagcgaaggccgcggtggcgtggtcggtcctccgcggttaaggagccgagg
 ctccgacgcgggctgcgcccgtgagcggcgccagagctaacttgcgctgactggaccagc
 tgaggagcggcccgcgggcgactgcgagcttcaccgagaggcttctccgacctggtccgc
 agtcccgacggcggcgagacc

10 (SEQ ID NO: 119)

The promoter may, for example, comprise or consist of a nucleic acid sequence that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% nucleotide identity to any one of SEQ ID NOs: 118 or 119, preferably wherein the promoter substantially retains the natural function of the promoter of SEQ ID NO: 118 or 119, respectively.

15 In some embodiments, the polynucleotide further comprises one or more miRNA target sequence. Suitably, the nucleic acid sequence encoding the ESF is operably linked to the one or more miRNA target sequence.

MicroRNA (miRNA) genes are scattered across all human chromosomes, except for the Y chromosome. They can be either located in non-coding regions of the genome or within
 20 introns of protein-coding genes. Around 50% of miRNAs appear in clusters which are transcribed as polycistronic primary transcripts. Similar to protein-coding genes, miRNAs are usually transcribed from polymerase-II promoters, generating a so-called primary miRNA transcript (pri-miRNA). This pri-miRNA is then processed through a series of endonucleolytic cleavage steps, performed by two enzymes belonging to the RNase Type III family, Drosha
 25 and Dicer. From the pri-miRNA, a stem loop of about 60 nucleotides in length, called miRNA precursor (pre-miRNA), is excised by a specific nuclear complex, composed of Drosha and DiGeorge syndrome critical region gene (DGCR8), which crops both strands near the base of the primary stem loop and leaves a 5' phosphate and a 2 bp long, 3' overhang. The pre-miRNA is then actively transported from the nucleus to the cytoplasm by RAN-GTP and
 30 Exportin. Then, Dicer performs a double strand cut at the end of the stem loop not defined by the Drosha cut, generating a 19-24 bp duplex, which is composed of the mature miRNA and the opposite strand of the duplex, called miRNA*. In agreement with the thermodynamic asymmetry rule, only one strand of the duplex is selectively loaded into the RNA-induced silencing complex (RISC), and accumulates as the mature microRNA. This strand is usually
 35 the one whose 5' end is less tightly paired to its complement, as was demonstrated by single-nucleotide mismatches introduced into the 5' end of each strand of siRNA duplexes. However, there are some miRNAs that support accumulation of both duplex strands to similar extent.

MicroRNAs trigger RNAi, very much like small interfering RNAs (siRNA) which are extensively used for experimental gene knockdown. The main difference between miRNA and siRNA is their biogenesis. Once loaded into RISC, the guide strand of the small RNA molecule interacts with mRNA target sequences preferentially found in the 3' untranslated region (3'UTR) of protein-coding genes. It has been shown that nucleotides 2-8 counted from the 5' end of the miRNA, the so-called seed sequence, are essential for triggering RNAi. If the whole guide strand sequence is perfectly complementary to the mRNA target, as is usually the case for siRNAs and plant miRNAs, the mRNA is endonucleolytically cleaved by involvement of the Argonaute (Ago) protein, also called "slicer" of the small RNA duplex into the RNA-induced silencing complex (RISC). DGRC (DiGeorge syndrome critical region gene 8) and TRBP (TAR (HIV) RNA binding protein 2) are double-stranded RNA-binding proteins that facilitate mature miRNA biogenesis by Drosha and Dicer RNase III enzymes, respectively. The guide strand of the miRNA duplex gets incorporated into the effector complex RISC, which recognises specific targets through imperfect base-pairing and induces post-transcriptional gene silencing. Several mechanisms have been proposed for this mode of regulation: miRNAs can induce the repression of translation initiation, mark target mRNAs for degradation by deadenylation, or sequester targets into the cytoplasmic P-body.

On the other hand, if only the seed is perfectly complementary to the target mRNA but the remaining bases show incomplete pairing, RNAi acts through multiple mechanisms leading to translational repression. Eukaryotic mRNA degradation mainly occurs through the shortening of the polyA tail at the 3' end of the mRNA, and de-capping at the 5' end, followed by 5'-3' exonuclease digestion and accumulation of the miRNA in discrete cytoplasmic areas, the so called P-bodies, enriched in components of the mRNA decay pathway.

Expression of the nucleic acid sequence encoding the ESF may be regulated by one or more endogenous miRNAs using one or more corresponding miRNA target sequence. Using this method, one or more miRNAs endogenously expressed in a cell prevent or reduce transgene expression in that cell by binding to its corresponding miRNA target sequence positioned in the polynucleotide or vector.

The target sequence may be fully or partially complementary to the miRNA. The term "fully complementary", as used herein, may mean that the target sequence has a nucleic acid sequence which is 100% complementary to the sequence of the miRNA which recognises it. The term "partially complementary", as used herein, may mean that the target sequence is only in part complementary to the sequence of the miRNA which recognises it, whereby the partially complementary sequence is still recognised by the miRNA. In other words, a

partially complementary target sequence in the context of the present invention is effective in recognising the corresponding miRNA and effecting prevention or reduction of transgene expression in cells expressing that miRNA.

5 Copies of miRNA target sequences may be separated by a spacer sequence. The spacer sequence may comprise, for example, at least one, at least two, at least three, at least four or at least five nucleotide bases.

10 The one or more miRNA target sequence may, for example, suppress expression of the nucleic acid sequence encoding the ESF in non-cancer cells. This may, for example, increase safety of a therapy using the ESF. Expression of the nucleic acid sequence encoding the ESF in cancer cells may, for example, not be suppressed by the one or more miRNA target sequence.

15 The term "suppress expression" as used herein may refer to a reduction of expression in the relevant cell type(s) of a transgene to which the one or more miRNA target sequence is operably linked as compared to transgene expression in the absence of the one or more miRNA target sequence, but under otherwise substantially identical conditions. In some embodiments, transgene expression is suppressed by at least 50%. In some embodiments, transgene expression is suppressed by at least 60%, 70%, 80%, 90% or 95%. In some embodiments, transgene expression is substantially prevented.

EXEMPLARY CONSTRUCTS

20 The polynucleotide may, for example, comprise or consist of a nucleic acid sequence that has at least 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% nucleotide identity to any one of SEQ ID NOs: 120-125, preferably wherein the polynucleotide substantially retains the natural function of the polynucleotide of SEQ ID NO: 120-125, respectively.

TARGET GENE TRANSCRIPTION AND EXPRESSION

25 The ESF of the invention may be used in a method that represses transcription and/or expression of at least one target gene. Suitably, the target gene is an endogenous gene.

The at least one target gene transcription and/or expression may be repressed by epigenetic editing.

30 The level of transcription or expression of a target gene may be decreased by, for example, at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99% or 100% compared to the level of transcription or expression in the absence of the ESF.

The at least one target gene may be silenced. By “silencing a target gene”, it is to be understood that transcription and/or expression of the target gene is reduced to an extent sufficient to achieve a desired effect. The reduced expression may be sufficient to achieve a therapeutically relevant effect, such as the prevention or treatment of a disease, such as cancer. For example, a target gene is preferably repressed to an extent that there is either no transcription and/or expression of the target gene, or the residual level of transcription and/or expression of the target gene is sufficiently low to ameliorate or prevent the disease state.

Methods of analysing transcription or expression of a gene are well known in the art. Methods for determining the transcription of a gene are known in the art and include reverse transcription PCR and Northern blot-based approaches. Methods for determining the expression of a gene are known in the art and include Western blot-based or flow cytometry approaches.

Preferably, the repression of the target gene occurs following transient delivery or expression of the ESF of the invention to or in a cell.

By “transient expression”, it is to be understood that the expression of the ESF is not stable over a prolonged period of time. Preferably, a polynucleotide encoding the ESF does not integrate into the host genome. More specifically, transient expression may be expression which is substantially lost within 20 weeks following introduction of the polynucleotide encoding the ESF into the cell. Preferably, expression is substantially lost within 12, 6, 4 or 2 weeks following introduction of the polynucleotide encoding the ESF into the cell.

Similarly, by “transient delivery”, it is to be understood that the ESF substantially does not remain in the cell (i.e. is substantially lost by the cell) over a prolonged period of time. More specifically, transient delivery may result in the ESF being substantially lost by the cell within 20 weeks following introduction of the ESF into the cell. Preferably, the ESF is substantially lost within 12, 6, 4 or 2 weeks following introduction of the ESF into the cell.

In some embodiments, the ESF is delivered transiently. Transient delivery may result in permanent change

Preferably, the at least one target gene is repressed or silenced permanently. By “permanent repression” or “permanent silencing” of a target gene, it is to be understood that transcription or expression of the target gene is reduced (e.g. reduced by at least 60%, at least 70%, at least 80%, at least 90% or 100%) compared to the level of transcription or expression in the absence of the ESF for at least 2 months, 6 months, 1 year, 2 year or the entire lifetime of

the cell/organism. Preferably, a permanently repressed or silenced target gene remains repressed or silenced for the remainder of the cell's life.

In some embodiments, the ESF is stably expressed.

PROTEINS

- 5 The term "protein" as used herein includes single-chain polypeptide molecules as well as multiple-polypeptide complexes where individual constituent polypeptides are linked by covalent or non-covalent means. The terms "polypeptide" and "peptide" as used herein refer to a polymer in which the monomers are amino acids and are joined together through peptide or disulfide bonds.

10 PROTEIN TRANSDUCTION

As an alternative to the delivery of polynucleotides to cells, the ESF of the invention may be delivered to cells by protein transduction.

- Protein transduction may be via vector delivery (Cai et al. (2014) *Elife* 3: e01911; Maetzig et al. (2012) *Curr. Gene Ther.* 12: 389-409). Vector delivery involves the engineering of viral particles (e.g. lentiviral particles) to comprise the proteins to be delivered to a cell. Accordingly, when the engineered viral particles enter a cell as part of their natural life cycle, the proteins comprised in the particles are carried into the cell.

- Protein transduction may be via protein delivery (Gaj et al. (2012) *Nat. Methods* 9: 805-7). Protein delivery may be achieved, for example, by utilising a vehicle (e.g. nanoparticles, such as liposomes) or even by administering the protein itself directly to a cell.

In some embodiments, the ESF is comprised in a nanoparticle. In some embodiments, the nanoparticle is a polymeric nanoparticle, inorganic nanoparticle or lipid nanoparticle. In some embodiments, the nanoparticle is a liposome.

- The nanoparticle may be targeted to a specific cell type(s) (e.g. cancer cells) using one or more ligand displayed on its surface.

POLYNUCLEOTIDES

- Polynucleotides of the invention may comprise DNA or RNA. They may be single-stranded or double-stranded. It will be understood by a skilled person that numerous different polynucleotides can encode the same polypeptide as a result of the degeneracy of the genetic code. In addition, it is to be understood that the skilled person may, using routine

techniques, make nucleotide substitutions that do not affect the polypeptide sequence encoded by the polynucleotides of the invention to reflect the codon usage of any particular host organism in which the polypeptides of the invention are to be expressed.

5 The polynucleotides may be modified by any method available in the art. Such modifications may be carried out in order to enhance the in vivo activity or lifespan of the polynucleotides of the invention.

Polynucleotides such as DNA polynucleotides may be produced recombinantly, synthetically or by any means available to the skilled person. They may also be cloned by standard techniques.

10 Longer polynucleotides will generally be produced using recombinant means, for example using polymerase chain reaction (PCR) cloning techniques. This may involve making a pair of primers (e.g. of about 15 to 30 nucleotides) flanking the target sequence which it is desired to clone, bringing the primers into contact with mRNA or cDNA obtained from an animal or human cell, performing a polymerase chain reaction under conditions which bring
15 about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture with an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable vector.

VECTORS

20 A vector is a tool that allows or facilitates the transfer of an entity from one environment to another. In accordance with the invention, and by way of example, some vectors used in recombinant nucleic acid techniques allow entities, such as a segment of nucleic acid (e.g. a heterologous DNA segment, such as a heterologous cDNA segment), to be transferred into a target cell. The vector may serve the purpose of maintaining the heterologous nucleic acid
25 (DNA or RNA) within the cell, facilitating the replication of the vector comprising a segment of nucleic acid or facilitating the expression of the protein encoded by a segment of nucleic acid. Vectors may be non-viral or viral. Examples of vectors used in recombinant nucleic acid techniques include, but are not limited to, plasmids, mRNA molecules (e.g. in vitro transcribed mRNAs), chromosomes, artificial chromosomes and viruses. The vector may
30 also be, for example, a naked nucleic acid (e.g. DNA). In its simplest form, the vector may itself be a nucleotide of interest.

The vectors used in the invention may be, for example, plasmid, mRNA or virus vectors and may include a promoter for the expression of a polynucleotide and optionally a regulator of the promoter.

5 Vectors comprising polynucleotides used in the invention may be introduced into cells using a variety of techniques known in the art, such as transfection, transformation and transduction. Several such techniques are known in the art, for example infection with recombinant viral vectors, such as retroviral, lentiviral (e.g. integration-defective lentiviral), adenoviral, adeno-associated viral, baculoviral and herpes simplex viral vectors; direct injection of nucleic acids and biolistic transformation.

10 Non-viral delivery systems include but are not limited to DNA transfection methods. Here, transfection includes a process using a non-viral vector to deliver a gene to a target cell. Typical transfection methods include electroporation, DNA biolistics, lipid-mediated transfection, compacted DNA-mediated transfection, liposomes, immunoliposomes, lipofectin, cationic agent-mediated transfection, cationic facial amphiphiles (CFAs) (Nat. Biotechnol. (1996) 14: 556) and combinations thereof.

Transfection of cells with mRNA vectors can be achieved, for example, using nanoparticles, such as liposomes.

In some embodiments, the vector (e.g. mRNA vector) is comprised in a nanoparticle. In some embodiments, the nanoparticle is a polymeric nanoparticle, inorganic nanoparticle or
20 lipid nanoparticle. In some embodiments, the nanoparticle is a liposome.

The nanoparticle may be targeted to a specific cell type(s) (e.g. cancer cells) using one or more ligand displayed on its surface.

Viral vectors

In preferred embodiments, the vector is a viral vector. The viral vector may be in the form of
25 a viral vector particle.

The viral vector may be, for example, a retroviral, lentiviral, adeno-associated viral (AAV) or adenoviral vector.

In some embodiments, the vector is a lentiviral vector. In some embodiments, the vector is an AAV vector.

30 *Retroviral and lentiviral vectors*

A retroviral vector may be derived from or may be derivable from any suitable retrovirus. A large number of different retroviruses have been identified. Examples include murine leukaemia virus (MLV), human T-cell leukaemia virus (HTLV), mouse mammary tumour virus (MMTV), Rous sarcoma virus (RSV), Fujinami sarcoma virus (FuSV), Moloney murine
5 leukaemia virus (Mo-MLV), FBR murine osteosarcoma virus (FBR MSV), Moloney murine sarcoma virus (Mo-MSV), Abelson murine leukaemia virus (A-MLV), avian myelocytomatosis virus-29 (MC29) and avian erythroblastosis virus (AEV). A detailed list of retroviruses may be found in Coffin et al. (1997) *Retroviruses*, Cold Spring Harbour Laboratory Press, 758-63.

Retroviruses may be broadly divided into two categories, "simple" and "complex".
10 Retroviruses may be even further divided into seven groups. Five of these groups represent retroviruses with oncogenic potential. The remaining two groups are the lentiviruses and the spumaviruses. A review of these retroviruses is presented in Coffin et al. (1997) *Retroviruses*, Cold Spring Harbour Laboratory Press, 758-63.

The basic structure of retrovirus and lentivirus genomes share many common features such
15 as a 5' LTR and a 3' LTR. Between or within these are located a packaging signal to enable the genome to be packaged, a primer binding site, integration sites to enable integration into a host cell genome, and gag, pol and env genes encoding the packaging components – these are polypeptides required for the assembly of viral particles. Lentiviruses have additional features, such as rev and RRE sequences in HIV, which enable the efficient
20 export of RNA transcripts of the integrated provirus from the nucleus to the cytoplasm of an infected target cell.

In the provirus, these genes are flanked at both ends by regions called long terminal repeats (LTRs). The LTRs are responsible for proviral integration and transcription. LTRs also serve as enhancer-promoter sequences and can control the expression of the viral genes.

25 The LTRs themselves are identical sequences that can be divided into three elements: U3, R and U5. U3 is derived from the sequence unique to the 3' end of the RNA. R is derived from a sequence repeated at both ends of the RNA. U5 is derived from the sequence unique to the 5' end of the RNA. The sizes of the three elements can vary considerably among different retroviruses.

30 In a defective retroviral vector genome gag, pol and env may be absent or not functional.

In a typical retroviral vector, at least part of one or more protein coding regions essential for replication may be removed from the virus. This makes the viral vector replication-defective.

Lentivirus vectors are part of the larger group of retroviral vectors. A detailed list of lentiviruses may be found in Coffin et al. (1997) *Retroviruses*, Cold Spring Harbour Laboratory Press, 758-63. Lentiviruses can be divided into primate and non-primate groups. Examples of primate lentiviruses include but are not limited to human immunodeficiency virus (HIV), the causative agent of human acquired immunodeficiency syndrome (AIDS); and simian immunodeficiency virus (SIV). Examples of non-primate lentiviruses include the prototype "slow virus" visna/maedi virus (VMV), as well as the related caprine arthritis-encephalitis virus (CAEV), equine infectious anaemia virus (EIAV), and the more recently described feline immunodeficiency virus (FIV) and bovine immunodeficiency virus (BIV).

10 The lentivirus family differs from retroviruses in that lentiviruses have the capability to infect both dividing and non-dividing cells (Lewis et al. (1992) *EMBO J.* 11: 3053-8; Lewis et al. (1994) *J. Virol.* 68: 510-6). In contrast, other retroviruses, such as MLV, are unable to infect non-dividing or slowly dividing cells such as those that make up, for example, muscle, brain, lung and liver tissue.

15 A lentiviral vector, as used herein, is a vector which comprises at least one component part derivable from a lentivirus. Preferably, that component part is involved in the biological mechanisms by which the vector infects cells, expresses genes or is replicated.

The lentiviral vector may be a "primate" vector. The lentiviral vector may be a "non-primate" vector (i.e. derived from a virus which does not primarily infect primates, especially humans).

20 Examples of non-primate lentiviruses may be any member of the family of lentiviridae which does not naturally infect a primate.

Preferably, the viral vector used in the present invention has a minimal viral genome.

By "minimal viral genome" it is to be understood that the viral vector has been manipulated so as to remove the non-essential elements and to retain the essential elements in order to provide the required functionality to infect, transduce and deliver a nucleotide sequence of interest to a target host cell. Further details of this strategy can be found in WO 1998/017815.

Preferably, the plasmid vector used to produce the viral genome within a host cell/packaging cell will have sufficient lentiviral genetic information to allow packaging of an RNA genome, in the presence of packaging components, into a viral particle which is capable of infecting a target cell, but is incapable of independent replication to produce infectious viral particles within the final target cell. Preferably, the vector lacks a functional gag-pol and/or env gene and/or other genes essential for replication.

However, the plasmid vector used to produce the viral genome within a host cell/packaging cell will also include transcriptional regulatory control sequences operably linked to the lentiviral genome to direct transcription of the genome in a host cell/packaging cell. These regulatory sequences may be the natural sequences associated with the transcribed viral sequence (i.e. the 5' U3 region), or they may be a heterologous promoter, such as another viral promoter (e.g. the CMV promoter).

The vectors may be self-inactivating (SIN) vectors in which the viral enhancer and promoter sequences have been deleted. SIN vectors can be generated and transduce non-dividing cells in vivo with an efficacy similar to that of wild-type vectors. The transcriptional inactivation of the long terminal repeat (LTR) in the SIN provirus should prevent mobilisation by replication-competent virus. This should also enable the regulated expression of genes from internal promoters by eliminating any cis-acting effects of the LTR.

The vectors may be integration-defective. Integration defective lentiviral vectors (IDLVs) can be produced, for example, either by packaging the vector with catalytically inactive integrase (such as an HIV integrase bearing the D64V mutation in the catalytic site) or by modifying or deleting essential att sequences from the vector LTR, or by a combination of the above.

Adeno-associated viral (AAV) vectors

Adeno-associated virus (AAV) is an attractive vector system for use in the invention as it has a high frequency of integration.

AAV has a broad host range for infectivity. Details concerning the generation and use of AAV vectors are described in US Patent No. 5139941 and US Patent No. 4797368.

Recombinant AAV vectors have been used successfully for in vitro and in vivo transduction of marker genes and genes involved in human diseases.

VARIANTS, DERIVATIVES, ANALOGUES, HOMOLOGUES AND FRAGMENTS

In addition to the specific proteins and polynucleotides mentioned herein, the invention also encompasses the use of variants, derivatives, analogues, homologues and fragments thereof.

In the context of the invention, a variant of any given sequence is a sequence in which the specific sequence of residues (whether amino acid or nucleic acid residues) has been modified in such a manner that the polypeptide or polynucleotide in question substantially retains at least one of its endogenous functions. A variant sequence can be obtained by

addition, deletion, substitution, modification, replacement and/or variation of at least one residue present in the naturally-occurring protein.

The term “derivative” as used herein in relation to proteins or polypeptides of the invention includes any substitution of, variation of, modification of, replacement of, deletion of and/or
5 addition of one (or more) amino acid residues from or to the sequence providing that the resultant protein or polypeptide substantially retains at least one of its endogenous functions.

The term “analogue” as used herein in relation to polypeptides or polynucleotides includes any mimetic, that is, a chemical compound that possesses at least one of the endogenous functions of the polypeptides or polynucleotides which it mimics.

10 Typically, amino acid substitutions may be made, for example from 1, 2 or 3 to 10 or 20 substitutions provided that the modified sequence substantially retains the required activity or ability. Amino acid substitutions may include the use of non-naturally occurring analogues.

Proteins used in the invention may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent protein.

15 Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues as long as the endogenous function is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar
20 hydrophilicity values include asparagine, glutamine, serine, threonine and tyrosine.

Conservative substitutions may be made, for example according to the table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R H
AROMATIC		F W Y

25 The term “homologue” as used herein means an entity having a certain homology with the wild type amino acid sequence or the wild type nucleotide sequence. The term “homology” can be equated with “identity”.

A homologous sequence may include an amino acid sequence which may be at least 50%, 55%, 65%, 75%, 85% or 90% identical, preferably at least 95% or 97% or 99% identical to the subject sequence. Typically, the homologues will comprise the same active sites etc. as the subject amino acid sequence. Although homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity.

A homologous sequence may include a nucleotide sequence which may be at least 50%, 55%, 65%, 75%, 85% or 90% identical, preferably at least 95% or 97% or 99% identical to the subject sequence. Although homology can also be considered in terms of similarity, in the context of the present invention it is preferred to express homology in terms of sequence identity.

Preferably, reference to a sequence which has a percent identity to any one of the SEQ ID NOs disclosed herein refers to a sequence which has the stated percent identity over the entire length of the SEQ ID NO referred to.

Homology comparisons can be conducted by eye or, more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate percentage homology or identity between two or more sequences.

Percentage homology may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid in one sequence is directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues.

Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion in the nucleotide sequence may cause the following codons to be put out of alignment, thus potentially resulting in a large reduction in percent homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local homology.

However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment

- with as few gaps as possible, reflecting higher relatedness between the two compared sequences, will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example when using the GCG Wisconsin Bestfit package the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.
- Calculation of maximum percentage homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A.; Devereux et al. (1984) *Nucleic Acids Res.* 12: 387). Examples of other software that can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al. (1999) *ibid* – Ch. 18), FASTA (Atschul et al. (1990) *J. Mol. Biol.* 403-410) and the GENWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel et al. (1999) *ibid*, pages 7-58 to 7-60). However, for some applications, it is preferred to use the GCG Bestfit program. Another tool, called BLAST 2 Sequences is also available for comparing protein and nucleotide sequences (see *FEMS Microbiol. Lett.* (1999) 174: 247-50; *FEMS Microbiol. Lett.* (1999) 177: 187-8).

- Although the final percentage homology can be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix – the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table if supplied (see the user manual for further details). For some applications, it is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

Once the software has produced an optimal alignment, it is possible to calculate percentage homology, preferably percentage sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

“Fragments” are also variants and the term typically refers to a selected region of the polypeptide or polynucleotide that is of interest either functionally or, for example, in an assay. “Fragment” thus refers to an amino acid or nucleic acid sequence that is a portion of a full-length polypeptide or polynucleotide.

- 5 Such variants may be prepared using standard recombinant DNA techniques such as site-directed mutagenesis. Where insertions are to be made, synthetic DNA encoding the insertion together with 5' and 3' flanking regions corresponding to the naturally-occurring sequence either side of the insertion site may be made. The flanking regions will contain convenient restriction sites corresponding to sites in the naturally-occurring sequence so that
- 10 the sequence may be cut with the appropriate enzyme(s) and the synthetic DNA ligated into the cut. The DNA is then expressed in accordance with the invention to make the encoded protein. These methods are only illustrative of the numerous standard techniques known in the art for manipulation of DNA sequences and other known techniques may also be used.

Codon optimisation

- 15 The polynucleotides used in the invention may be codon-optimised. Codon optimisation has previously been described in WO 1999/41397 and WO 2001/79518. Different cells differ in their usage of particular codons. This codon bias corresponds to a bias in the relative abundance of particular tRNAs in the cell type. By altering the codons in the sequence so that they are tailored to match with the relative abundance of corresponding tRNAs, it is
- 20 possible to increase expression. By the same token, it is possible to decrease expression by deliberately choosing codons for which the corresponding tRNAs are known to be rare in the particular cell type. Thus, an additional degree of translational control is available.

COMPOSITIONS

- The ESFs, polynucleotides, vectors and cells of the invention may be formulated for
- 25 administration to subjects with a pharmaceutically-acceptable carrier, diluent or excipient. Suitable carriers and diluents include isotonic saline solutions, for example phosphate-buffered saline, and potentially contain human serum albumin.

- Materials used to formulate a pharmaceutical composition should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or
- 30 other material may be determined by the skilled person according to the route of administration.

The pharmaceutical composition is typically in liquid form. Liquid pharmaceutical compositions generally include a liquid carrier such as water, petroleum, animal or vegetable oils, mineral oil or synthetic oil. Physiological saline solution, magnesium chloride, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol may be included. In some cases, a surfactant, such as pluronic acid (PF68) 0.001% may be used. In some cases, serum albumin may be used in the composition.

For injection, the active ingredient may be in the form of an aqueous solution which is pyrogen-free, and has suitable pH, isotonicity and stability. The skilled person is well able to prepare suitable solutions using, for example, isotonic vehicles such as Sodium Chloride Injection, Ringer's Injection or Lactated Ringer's Injection. Preservatives, stabilisers, buffers, antioxidants and/or other additives may be included as required.

For delayed release, the medicament may be included in a pharmaceutical composition which is formulated for slow release, such as in microcapsules formed from biocompatible polymers or in liposomal carrier systems according to methods known in the art.

Handling of the cell therapy products is preferably performed in compliance with FACT-JACIE International Standards for cellular therapy.

METHOD OF TREATMENT

In another aspect, the invention provides the ESF, polynucleotide, vector, cell or composition of the invention for use in therapy.

In another aspect, the invention provides the ESF, polynucleotide, vector, cell or composition of the invention for use in the treatment of cancer.

In some embodiments, the cancer is glioma, glioblastoma, medulloblastoma, astrocytoma, neuroblastomas, ependymoma, meningioma, retinoblastoma, rhabdomyosarcoma, lung cancer, prostate cancer, breast cancer, liver cancer, pancreatic cancer, bladder cancer, oropharyngeal cancer or kidney cancer. In some embodiments, the cancer is a brain tumour. In some embodiments, the cancer is glioblastoma multiforme.

In some embodiments, the treatment reduces tumour size.

In some embodiments, the treatment is as an adjuvant therapy, optionally in combination with surgery.

All references herein to treatment include curative, palliative and prophylactic treatment. The treatment of mammals, particularly humans, is preferred. Both human and veterinary treatments are within the scope of the invention.

5 In some embodiments, the method of treatment provides the ESF, polynucleotide, vector or cell of the invention to a tumor.

In some embodiments, the method of treatment provides the ESF, polynucleotide, vector or cell of the invention to the brain of a subject.

ADMINISTRATION

10 In some embodiments, the ESF, polynucleotide, vector or cell is administered to a subject locally.

In some embodiments, the ESF, polynucleotide, vector or cell is administered to a subject's brain.

In preferred embodiments, the ESF, polynucleotide, vector or cell is administered to a tumor.

15 The term "systemic delivery" or "systemic administration" as used herein means that the agent of the invention is administered into the circulatory system, for example to achieve broad distribution of the agent. In contrast, topical or local administration restricts the delivery of the agent to a localised area, e.g. a tumor.

DOSAGE

20 The skilled person can readily determine an appropriate dose of an agent of the invention to administer to a subject. Typically, a physician will determine the actual dosage which will be most suitable for an individual patient and it will depend on a variety of factors including the activity of the specific compound employed, the metabolic stability and length of action of that compound, the age, body weight, general health, sex, diet, mode and time of administration, rate of excretion, drug combination, the severity of the particular condition,
25 and the individual undergoing therapy. There can of course be individual instances where higher or lower dosage ranges are merited, and such are within the scope of the invention.

SUBJECT

The term "subject" as used herein refers to either a human or non-human animal.

Examples of non-human animals include vertebrates, for example mammals, such as non-human primates (particularly higher primates), dogs, rodents (e.g. mice, rats or guinea pigs), pigs and cats. The non-human animal may be a companion animal.

Preferably, the subject is a human.

5

The skilled person will understand that they can combine all features of the invention disclosed herein without departing from the scope of the invention as disclosed.

Preferred features and embodiments of the invention will now be described by way of non-limiting examples.

- 10 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, biochemistry, molecular biology, microbiology and immunology, which are within the capabilities of a person of ordinary skill in the art. Such techniques are explained in the literature. See, for example, Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press; Ausubel, F.M. et al. (1995 and periodic supplements) Current Protocols in Molecular Biology, Ch. 9, 13 and 16, John Wiley & Sons; Roe, B., Crabtree, J. and Kahn, A. (1996) DNA Isolation and Sequencing: Essential Techniques, John Wiley & Sons; Polak, J.M. and McGee, J.O'D. (1990) In Situ Hybridization: Principles and Practice, Oxford University Press; Gait, M.J. (1984) Oligonucleotide Synthesis: A Practical Approach, IRL Press; and Lilley, D.M. and Dahlberg, J.E. (1992) Methods in Enzymology: DNA Structures Part A: Synthesis and Physical Analysis of DNA, Academic Press. Each of these general texts is herein incorporated by reference.
- 15
- 20

EXAMPLES

EXAMPLE 1

25 RESULTS

We sought to build a set of epigenetic silencer factors by adding the Kruppel-associated box (KRAB) domain and/or the catalytic domain of DNA methyltransferase DNMT3A together with its cofactor DNMT3L to the full length Sox2 or Sox2 lacking its C-terminal transcriptional activation domain (Fig. 1a).

- 30 The KRAB domain (from the zinc finger protein ZNF10) recruits different epigenetic complexes that are able to both induce repressive chromatin modification (e.g. H3K9me3)

and remove active marks (e.g. H3K4ac), while the DNMT domains coordinate *de novo* DNA methylation, thus preventing altogether genic transcription (Amabile et al. (2016) Cell 167: 219-232.e14). Only the factor composed of the N-terminal and DNA binding regions of SOX2 plus the KRAB and the DNMT domains, respectively at the 5' and 3' terminals (referred to as

5 SES), rapidly killed glioma cell lines (e.g. U87, U251 and SNB19) *in vitro* (Fig. 1b-f). Accordingly, genes associated with cell proliferation and *bona fide* SOX2 targets (*CDK1*, *CDC6*, *CCND1*, etc.) were strongly repressed in SES-treated cancer cells (Fig. 1g). Thus, removal of the SOX2 transcriptional activation domain was critical to obtain a synthetic factor capable of repressing cancer cell proliferation. Interestingly, proliferation of cell lines from

10 prostatic, liver and pancreatic cancers were largely unaffected by SES treatment suggesting a cancer-specific SES efficacy (Fig. 1h). We extended the *in vitro* analysis to patient-derived glioblastoma multiforme cancer stem cells (GBM CSCs) of both classical and mesenchymal subtypes that better preserve primary tumor features. Both lines of GBM CSCs displayed a strong proliferative loss and high cell death after SES lentiviral (LV) transduction (Fig. 2a).

15 To determine SES specificity, we mutated two residues in the HMG-box domain (arginine at position 74 and leucine at position 97 were replaced by two prolines) that have been described as important for Sox2 binding the DNA (Fig. 1e). Expression of SES(R74P/L97P) failed to arrest CSC growth, indicating that SES activity relies on its DNA binding activity (Fig. 2a). Additionally, clonogenic analysis showed that SES-treated CSCs exhibited a

20 diminished self-renewal capacity with impairment in forming tumor spheres and sustaining their growth (Fig. 2b).

Next, we set out to assess global SES transcriptional output and its genome-wide occupancy. SES-treated GBM CSCs exhibited massive transcriptional changes with general upregulation of apoptosis-related genes and silencing of genes encoding for proliferative and

25 cancer-promoting factors in at least two different glioma cell lines (Fig. 3a-c). Interestingly, using a list of putative SOX2 targets inferred with computational methods (Janky (2014) et al. PLoS Comput Biol. 10: e1003731), we discovered that they were largely downregulated in SES-expressing cells (Fig. 3d). Comparative ChIP-seq experiments to determine SOX2 and SES genome-wide binding on CSCs revealed that SES maintained the ability to bind to

30 the majority of the SOX2 genomic sites (Fig. 3e). Given that SES includes the catalytic domains of the DNMT3A/L *de novo* DNA methyltransferases, we used MeDIP-seq to profile the methylated DNA. Notably, SOX2-bound regions presented a significant increase in the methylation state in SES transduced cells (Fig. 2f).

We then asked whether SES expression *in vivo* could exert any anti-tumoral activity. Initially,

35 we performed subcutaneous xenografts in NSG immunodeficient mice using CSCs

previously transduced with lentiviruses expressing either GFP (mock) or SES. Importantly, heterotopic transplants grew only from grafted mock cells, while SES transduced cells were unable to sustain tumor growth (Fig. 4a-c), suggesting that this factor is toxic for glioma cells and prevents tumor formation. No SES⁺ cells were ever identified even when some scar-like
5 epithelial mass could be isolated within the injection site (Fig. 4c). Similar results were obtained inducing subcutaneous tumoral growth with glioma cell lines (U87, SNB19, U251) (Fig. 4d). Next, orthotopic intracranial xenografts were carried out with GFP or SES expressing U87 cancer cells. 4 weeks after brain transplantation, mock GFP⁺ grafted cells generated large tumoral masses extending throughout the striatum up to the cerebral cortex
10 that were rich in proliferating cells (Fig. 5a-e). Accordingly, all the mice transplanted with mock U87 cells succumbed within 30 days from the grafting (Fig. 5f). Conversely, grafted SES⁺ U87 cells completely lost their tumor-initiating capacity and no tumoral masses could be retrieved from the transplanted mice that exhibited an unaltered survival curve (Fig. 5b-g). Similar results were obtained with SES-treated GBM CSCs that were able to produce very
15 small tumoral masses 6 weeks after brain grafting (Fig. 5h). To test GBM CSC tumorigenicity in a full humanized model system, we developed brain organoids by 3D differentiation of human iPSCs (Lancaster et al. (2013) Nature 501: 373-9) and assembled them together with GBM CSC-derived spheroids using Matrigel embedding (Linkous et al. (2019) Cell Rep. 26: 3203-3211.e5; Goranci-Buzhala et al. (2020) Cell Rep. 31: 107738). In these assembloids
20 (GBM-cortex organoids), mock GFP⁺ CSCs succeed over time to infiltrate and diffuse within the brain cortical-like tissue (Fig. 6a). By contrast, in SES-treated organoids the GBM part remained very small failing to grow over time and delaminating cells in the organoid territory (Fig. 6b). These results strongly imply that SES expression in GBM cancer cells restrains their growth and survival erasing their *in vivo* tumor-initiating capacity. Following these
25 findings that validate the SES anti-tumor activity based on *in vitro* transduction of SES in cancer cells before grafting, we moved to an approach of SES viral transduction *in vivo* to curb the development of an already growing GBM mass in the brain parenchyma. Intracranial transplanted U87 cancer cells were left to proliferate and form a tumoral mass for 4 days before injecting a mock (GFP) or SES-expressing lentivirus (Fig. 7a,b). 4 weeks
30 after lentiviral gene transfer mock transduced tumors developed large masses diffused throughout the striatum (Fig. 7b,c). By contrast, *in situ* SES expression was sufficient to strongly reduce the tumor mass and significantly extend the survival rate of the affected animals (Fig. 7b-d). Notably, V5 immunohistochemistry in 4 weeks transplanted brains showed that SES⁺ cells were not present in the tumor mass but only in the surrounding
35 brain parenchymal tissue (Fig. 7e). This finding indicates that cancer cells transduced with the SES lentivirus were lost over time and that the remaining tumor tissue at 4 weeks from transplantation was composed exclusively by SES⁻ cells.

Then, we challenged our approach by treating tumors generated from patient-derived CSCs. Either a mock (GFP) or SES expressing lentivirus were injected *in situ* (7 days after the transplant of the CSCs) and the tumor growth was followed by weekly MRI T1 scanning up to the histological analysis after 6 weeks (Fig. 7f). SES-treated tumors are smaller compared to the mock treated group both at the MRI screening and at the endpoint (Fig. 7g-j). These data suggested that *in situ* viral treatment with SES is able to reduce the mass of patient derived GBM in mice.

SOX2 is a pivotal factor in stem cells and neural progenitors, but it is strongly downregulated during neuronal differentiation. Thus, SES should have a minor impact on mature brain neurons in adulthood. To determine the effect of SES on brain cells, primary mouse cortical neuronal cultures were treated with the SES LV and survival, morphology and gene expression were assessed two weeks later. Mock and SES expressing neurons displayed similar morphology with no sign of cell sufferance and with comparable low numbers of PI⁺ dying cells (Fig. 8a). Notably, transcriptomes of mock and SES transduced cultures were substantially comparable with only Sox2 and 15 other genes differentially expressed between the two neuronal populations (Fig. 8b). Similar results were collected using human iPSC-derived neuronal cultures where SES treatment did not alter survival and morphology of MAP2⁺ neurons (Fig. 8c,d). This analysis could last no more than 2 weeks given the short survival time of primary neurons *in vitro*. Thus, we moved *in vivo* and transduced SES in the hippocampus of C57BL/6 adult mice and assessed behavioral performance at long term (Fig. 9a,b). 4 weeks after stereotactic injections of mock (GFP) and SES LVs both mock and SES viral vectors were equally present in the hippocampi post-mortem (Fig. 9c). Similar results were obtained with the evaluation of transgenes' mRNA (Fig. 9c). Of note, the same level of cell death was observed followed both GFP and SES injection (Fig. 9d). Before sacrifice, mice were evaluated in the spontaneous alternation, Radial maze and Morris water maze tests to assess their exploratory behavior and cognitive function related to spatial learning and memory (Fig. 9e-g). Both groups of animals performed equally well in these tasks suggesting that SES expression did not elicit significant functional alterations in hippocampal neurons.

We then conceived a strategy to restrict SES expression to cancer cells after brain viral inoculation. We isolated a KI67 promoter (Zamboni (2010) Cytometry A 77: 564-70) that was inserted in the LV cassette (SES v1.1) to drive SES expression exclusively in proliferative cells (Fig. 10a). Firstly, we confirmed that our strategy did not affect either GFP (when it is placed downstream of the KI67 promoter) or SES expression levels in cancer cells and, in the case of SES, triggered a significant cell loss in the transduced cancer cells *in vitro* (Fig.

10b). Next, mouse primary neuronal cultures constituted by both neurons and glial cells, were transfected with either EF1a-GFP (constitutive) or pKI67-GFP LVs and investigated for the presence of fluorescent protein 1 week later. Interestingly, the great majority of neurons (MAP2⁺) infected with pKI67-GFP were found to be GFP⁻ indicating that its transcription was strongly reduced to undetectable levels, while constitutive GFP was expressed in all the cell of the dish (Fig. 10c). Curiously, the few GFP⁺ cells in the pKI67-GFP transduced cultures were also Ki67⁺ and, thus, in active cell-cycle that likely correspond to young proliferative astrocytes (red arrows in Fig. 10c). Thus, this strategy resulted in effective silencing of SES expression in brain post mitotic cells without compromising the SES transgene activation in cancer cells.

Then we generated additional variants of SES using alternative repressor domains such as the chromo shadow domain from CBX5 protein (inserted at the 5') (SESv2) or the YAF2-RYBP domain from RYBP protein (inserted at the 3') (SESv3) that replaced the KRAB and DNMT3A/L catalytic domains (Fig. 11a). Cancer cells transduced with either SESv2 or v3 showed a rapid proliferative loss and diffuse cell death which caused a premature termination of the cultures within two weeks from the initial treatment (Fig. 11b-c). Patient-derived GBM CSCs of classical subtype, that better preserve primary tumor features, also displayed a strong proliferative loss after SESv3 lentiviral (LV) transduction (Fig. 12a-b).

These results obtained with SES indicate that reconfiguration of activatory TFs into epigenetic silencer factors (ESFs) can be generalized to build more epigenetic regulators by engineering other activatory cancer associated TFs. With this goal, we applied the same rational design to TEAD1 and c-MYC, two other transcriptional activators with crucial oncogenic activity, thereby, generating the TES and MES factors, respectively (Fig. 11d). As for SES, ESFs were generated by removing the transcriptional activation domain and adding the KRAB and DNMT3A/L catalytic domain to the remaining parts of the two TFs (Fig. 11d). In case of c-MYC the repressor domains were inserted at the 5' to avoid spatial hindrance at the 3'-end which directly interacts with MAX for the formation of heterodimers (Fig. 11d). Cancer cells transduced with either TES or MES showed a rapid proliferative loss and diffuse cell death which caused a premature termination of the cultures within two weeks from the initial treatment (Fig. 11e-f and Fig. 13).

We then assessed TES/MES repression of tumor development in vivo. Both TES and MES pre-infected CSCs displayed decreased tumorigenic potential when xenotransplanted subcutaneously in NSG mice (Fig. 14). TES/MES pre-infected CSCs also showed limited tumor growth in NSG brain (orthotopic transplantation) compared to mock infected CSCs (Fig. 15). Notably, in situ TES expression was sufficient to strongly reduce the growth of a

tumor mass formed by transplantation of naïve CSCs a week before the injection of the virus (Fig. 16).

Thus, the ESF design can be applied to different oncogenic TFs generating a family of synthetic factors with strong anti-tumor activity. ESFs represent a new class of rationally designed factors with pervasive and long-lasting epigenetic functions that can remodel entire transcriptional pathways with precision and efficacy. Targeted ESF expression can suppress cancer development and represent new gene-based therapeutics for glioblastoma and other cancers.

DISCUSSION

Activity of developmental TFs is mainly restricted during morphogenesis executing a prominent role in stem cell identity, cell lineage commitment and differentiation. However, these TFs can be re-activated or hijacked by the cancer genetic program to propel tumor development and progression. It is estimated that about 20% of all the known oncogenic proteins are represented by TFs with critical importance for acquiring malignant cell dedifferentiation, proliferation, and migration. Despite their pervasive role in tumors, interfering with their functions in a translational perspective has proven challenging. In fact, stable and complete gene silencing by various genetic tools or small molecules have been difficult to achieve in cancer cells. Moreover, the cancer genetic program has been repeatedly shown to overcome the inactivation of single genes by reconfiguring the transcriptional network to promote cancer resistance and recurrence. Herein, we designed an epigenetic repressor (SES) by reconfiguring SOX2 TF by rational assembling of transcriptional and epigenetic negative regulators of gene transcription. This design is modular and versatile and can be in principle applied to other activatory oncogenic TFs as we showed for TEAD1 and c-MYC. Importantly, SES-dependent *de novo* DNA methylation in SOX2 target genes triggered by DNMTA/L catalytic domains promotes the widespread and stable silencing of the SOX2 downstream network. These wide transcriptional changes inhibited cell proliferation in cancer cells that failed to cope with these alterations, eventually dying. Here, we show that these domains can reconfigure the transcriptional activity of endogenous TFs while preserving their chromatin occupancy and target selectivity. We also showed that different configuration of SES, by means of using different epigenetically active protein domains (e.g. Chromo Shadow domain from CBX5 and YAF2-RYBP domain from RYBP), can elicit the same functional activity.

SOX2 expression is fundamental to control self-renewal and a malignant phenotype in GBM cancer cells as well as in many stem cells including pluripotent and neural types.

Importantly, SOX2 has a primary role in promoting tumor development in many other malignancies other than GBM including medulloblastoma and lung, prostate, breast cancers.

Thus, the use of SES or other ESFs can be expanded for the treatment of other cancers. Tumor targeting of ESFs by viral-mediated delivery can be in principle effective on cancers confined into those solid tissues which can be efficiently targeted by viral transduction *in vivo*. On this view, cancers in liver, lung, breast and kidney represent plausible targets for this approach since delivery routes and viral strains are known to obtain wide and high tissue transduction efficiency. Likewise, the same approach can be proposed to treat metastatic masses in the same organs.

10 Herein, SES was directly injected in the tumoral mass repressing its development. Similar approaches can be useful in a clinical setting to treat glioblastoma whose surgical resection is impractical due to unattainable locations within the brain or the close proximity to vital brain regions. Moreover, SES can be delivered in the brain parenchyma surrounding the resected primary tumor as adjuvant therapy in order to target the remaining cancer cells and
15 restrain subsequent tumor recurrence.

Herein, glioblastoma treatment with ESFs was carried out through lentiviral transduction by local injections into the affected tissue. However, alternative therapeutic viruses could be similarly employed as, in particular, strains of adeno-associated viruses (AAVs) that can spread through the brain tissue due to their small size and reduced binding to cell
20 membranes. Maximizing viral spreading in the brain parenchyma will increase the targeting efficiency of cancer cells scattered in the tissue providing a better protection from tumor recurrence. Moreover, non-viral vehicles such as nanoparticles or liposomes can be used to deliver SES mRNA or protein to obtain an acute transgene expression which can be still sufficient to inhibit cancer cells while strongly enhancing the overall safety profile of the
25 procedure.

We showed that SES expression is not harmful to neuronal cultures and in murine brain, and we further elaborated a strategy to restrict its activation only to proliferative cells, strongly enriched in cancers and rarely present in brain parenchyma. The system described here proved effective in expressing the viral transgene mainly in cancer cells, but not in brain
30 post-mitotic cells.

Altogether, we assembled an epigenetic repressor which operates as a dominant negative version of the oncogenic SOX2 TF and is able to bind and stably repress the SOX2 transcriptional network. Targeted viral delivery of SES in glioblastoma is sufficient to inhibit tumor development by blocking cell proliferation and inducing cell death. Given its wide

applicability to other oncogenic TFs and the high efficiency of targeting cancer cells by viral transduction, this approach provides the opportunity to repress glioblastoma and other deadly cancers.

MATERIAL AND METHODS

5 Constructs

KRAB-hSOX2: the full length human SOX2 gene was fused with the KRAB repressor domain (from the gene *ZNF10* encoding for a Zinc finger protein, aa 1-97) at its N-terminus while a V5 tag was fused at the C-terminus of the SOX2. The transgene was used in a lentiviral construct with Ef1a as a promoter.

- 10 KRAB-hSOX2-D3A&L: the full length human SOX2 gene was fused with the KRAB repressor domain (from the gene *ZNF10* encoding for a Zinc finger protein, aa 1-97) at its N-terminus while the functional domains of DNMT3A (aa 388-689) and 3L (aa 206-421) were fused at the C-terminus of the SOX2, a V5 tag was fused at the end of the last domain, at the C-terminus of the new chimeric transgene. The transgene was used in a lentiviral
15 construct with Ef1a as a promoter.

- KRAB-hSOX2₁₋₁₇₉: the initial part of SOX2 gene, coding for aa 1-179 (thus excluding the SOX2 activator domain) was fused with the KRAB repressor domain (from the gene *ZNF10* encoding for a Zinc finger protein, aa 1-97) at its N-terminus while a V5 tag was fused at the C-terminus of the SOX2. The transgene was used in a lentiviral construct with Ef1a as a
20 promoter.

hSOX2₁₋₁₇₉-D3A&L: the initial part of SOX2 gene, coding for the aa 1-179 (thus excluding the SOX2 activator domain) was fused with the functional domains of DNMT3A (aa 388-689) and 3L (aa 206-421), with a V5 tag fused at the C-terminus of the SOX2. The transgene was used in a lentiviral construct with Ef1a as promoter.

- 25 SES v1: the initial part of SOX2 gene, coding for aa 1-179 (thus excluding the SOX2 activator domain) was fused with the KRAB repressor domain (from the gene *ZNF10* encoding for a Zinc finger protein, aa 1-97) at its N-terminus while the functional domains of DNMT3A (aa 388-689) and 3L (aa 206-421) were fused at the C-terminus of the SOX2 portion, a V5 tag was fused at the end of the last domain, at the C-terminus of the new
30 chimeric transgene. The transgene was used in a lentiviral construct with Ef1a as a promoter.

SES (R74P/L97P): the SESv1 was mutagenized at residues 74 and 97 of the initial part of SOX2 gene (arginine at position 74 and leucine at position 97 to prolines).

SES v1.1: The transgene was the same as the SES version 1, however the Ef1a promoter was replaced with the proximal promoter of the murine *Mki67* gene (-1263 to -1 related to
5 *Mki67* atg).

SES v2: the initial part of SOX2 gene, coding for aa 1-179 (thus excluding the SOX2 activator domain) was fused with the Chromo Shadow (CS) repressor domain (from the gene *CBX5*, aa 121-179) at its N-terminus while a V5 tag was fused at the C-terminus of the SOX2.

10 SES v3: the initial part of SOX2 gene, coding for aa 1-179 (thus excluding the SOX2 activator domain) was fused with the YAF2-RYBP (Y-R) repressor domain (from the gene *RYBP*, aa 145-189) at its C-terminus, and a V5 tag was fused at the C-terminus of the Y-R.

TES: the initial part of *TEAD1* gene, coding for the aa 1-166 (thus excluding the TEAD1 activator domain) was fused with the KRAB repressor domain (from the gene *ZNF10*
15 encoding for a Zinc finger protein, aa 1-97) at its N-terminus while the functional domains of DNMT3A (aa 388-689) and 3L (aa 206-421) were fused at the C-terminus of the TEAD1 portion, a V5 tag was fused at the end of the last domain, at the C-terminus of the new chimeric transgene. The transgene was used in a lentiviral construct with Ef1a as a promoter.

20 MES: the C-terminal part of *MYC* gene, coding for the aa 144-454 (thus excluding the MYC activator domain) was fused with the KRAB repressor domain (from the gene *ZNF10* encoding for a Zinc finger protein, aa 1-97) and DNMT3A (aa 388-689) and 3L (aa 206-421) in tandem at the N-terminus of the MYC portion while a V5 tag was fused at the C-terminus. The transgene was used in a lentiviral construct with Ef1a as a promoter.

25 Construct sequences were:

Ef1a::KRAB-hSOX2₁₋₁₇₉-DNMT3a3L-V5 [SES]

30 CGATAAGCTTTGCAAAGATGGATAAAGTTTTAAACAGAGAGGAATCTTTGCAGCTAATGGAC
CTTCTAGGTCTTGAAAGGAGTGGGAATTGGCTCCGGTGCCCGTCAGTGCGCAGAGCGCACAT
CGCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCCGCAATTGAACCGGTGCCTAGAGAAGG
TGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTCCCGAGGGTGG
GGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTCGCAACGGGTTTGCCG
CCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC
CCTTGCGTGCCTTGAATTACTTCCACTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGG
GTTGGAAGTGGGTGGGAGAGTTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTG

AGTTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCT
 GTCTCGCTGCTTTTCGATAAGTCTCTAGCCATTTAAATTTTGTATGACCTGCTGCGACGCTT
 TTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTTCGGTTTTT
 TGGGGCCGCGGGCGGCGACGGGGCCCGTGCCTCCAGCGCACATGTTTCGGCGAGGCGGGGCC
 5 TGCAGCGCGGCCACCGAGAAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTG
 CCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTTCGGCACC
 AGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGA
 CGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCCGTCC
 TCAGCCGTGCTTCATGTGACTCCACGGAGTACCGGGCGCGTCCAGGCACCTCGATTAGTT
 10 CTCGAGCTTTTGGAGTACGTCGCTTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTC
 CCCACACTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCCTCTTG
 GAATTTGCCCTTTTTGAGTTTGGATCTTGGTTTATTCTCAAGCCTCAGACAGTGGTTCAAAG
 TTTTTTCTTCCATTTTCAGGTGTCGTGAGGAATTCTGCAGTCGATCGACGGTACCGCGGGCC
 CAACGGATCCACCATGGATGCTAAGTCACTGACTGCCTGGTCCCGGACACTGGTGACCTTCA
 15 AGGATGTGTTTGTGGACTTCACCAGGGAGGAGTGAAGCTGCTGGACACTGCTCAGCAGATC
 CTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTCCTTGGGTTATCAGCTTAC
 TAAGCCAGATGTGATCCTCCGTTGGAGAAGGGAGAAGAGCCCTGGCTGGTGGAGAGAGAAA
 TTCACCAAGAGACCCATCCTGATTCAGAGACTGCATTTGAAATCAAATCATCAGTTGGCAGC
 GGCGCATGCatgtacaacatgatggagacggagctgaagccgcccgggccccgcagcaaaacttc
 20 gggggggcgggcgggcgaactccaccgcgggcgggcgccggcggaaccagaaaaacagcccgg
 accgctgaagcgggcccatgaatgcctcatggtgtggtcccggggcagcgggcgcaagatg
 gccagggagaacccaagatgcacaactcggagatcagcaagcgccctgggcgccgagtgga
 acttttgcggagacggagaagcgggcgttcacgcagaggctaagcggtgagcgctgc
 acatgaaggagcaccggattataaataccggccccggggaacccaagacgctcatgaag
 25 aaggataagtacacgctgcccggcggtgctggccccggcggaatagcatggcgagcgg
 ggtcggggtgggcgcggcctgggcgcggcggtgaaccagcgcatggacagttacgcgcaca
 tgaacggctggagcaacggcagctacagcatgatgcaggaccagctgggcGCTAGCGGCAGC
 GGC**AACCATGATCAAGAGTTCGATCCTCCTAAAGTGTATCCCCCTGTCCCTGCCGAAAAGCG**
GAAACCGATCAGGGTTCTCAGTCTCTTCGATGGAATTGCCACAGGCCTCTTGGTCTT**TAAGG**
 30 **ACCTGGGCATCCAAGTgGac**CGCTACATCGCCTCAGAAAGTATGTGA**gGACAGCATCACAGTC**
GGCATGGTGCGCCACCAGGGGAAAATAAATGTACGTCGGCGACGTCAGATCAGTTACTCAGAA
ACATATTCAGGAATGGGGGCCTTTTGATCTCGTAATTGGCGGCAGCCCCCTGCAATGATCTGA
GTATTGTGAACCCCGCCCGGAAAGGTTTGTACGAAGGTACGGGCAGACTGTTCTTCGAGTTT
TACAGACTGTTGCACGACGCACGACCAAGGAGGGTGAcGAtcGACCCTTCTTCTGGCTGTT
 35 **CGAGAACGTGGTCGCTATGGGCGTGCTGACAAGCGGGACATTTCTAGATTCTTGAGAGCA**
ATCCAGTGATGATTGATGCAAAGGAAGTatccGCTGCCCACCGGCCAGATACTTCTGGGGC
AATCTGCCCGGCATGAATCGACCCTTGGCAAGCACCGTGAATGATAAGTTGGAGTTGCAGGA
GTGTCTgGAGCATGGGCGGATCGCAAAGTTTAGCAAGGTGCGGACCATCACGACCCGAAGCA
ATTCAATTAAGCAGGGAAAGGACCAACATTTTCCAGTGTTTATGAACGAGAAAGAGGAcATC
 40 **CTGTGGTGcACAGAAATGGAGCGGGTTTTCGGATTCCCCGTACATTATACTGATGTGTCCAA**
TATGTCTAGGTTGGCTAGACAGAGGCTGtTGGGACGCTCATGGAGTGTTCCTGTCTATCCGCC
ACCTGTTTCGCTCCTCTGAAGGAGTATTTTCGcTGCGTtTCATCCGGgAATTCAAACGCAAAC
AGCAGAGGCCCATCCTTTTCTCCGGCCTGGTgCCACTTAGTCTGCGCGGCTCTCACATGGG
ACCTATGGAATATACAAAACAGTGAGCGCCTGGAAGAGGCAGCCTGTGCGGGTCTTGAGTC
 45 **TGTT**CAGAAACATTGACAAGGTGCTTAAGTCCCTTGGATTTCTGGAGTCTGGCAGCGGTTCT
GGGGGTGgcACCTTAAGTATGTGGAGGACGTGACAAACGTCGTGAGGAGAGATGTGGAGAA
ATGGGGGCCATTTGACCTGGTATACGGGAGCACTCAACCTCTCGGTTCTTCTGTGATCGCT
GCCCAGGGTGGTATATGTTTCAGTTCCATCGcATCCTGCAATACGCCCTTCCGCGGCAGGAG

5 AGTCAGCGACCATTCCTTCTGGATATTCATGGACAATCTCCTGCTGACAGAgGACGACCAAGA
GACTACGACTAGATTTCTTCAGACCGAGGCCGTTACTCTCCAGGACGTTAGAGGTAGGGACT
ACCAGAATGCCATGCGAGTGTGGAGTAACATCCCCGGAAGTCAAGTCAAAGCAcGCACCCCTG
ACCCCAAGGAAGAGGAATACCTcCAGGCGCAGGTGAGGAGCCGCAGTAAGCTCGATGCTCC
TAAAGTCGATTTGTTGGTGAAGAATTGCTTGCTGCCCCTGAGAGAGTACTTTAAGTACTTTA
GCCAGAACAGCTTGCCACTGGGCGCGCCCGGTAAGCCTATCCCTAACCCCTCTCCTCGGTCTC
GATTCTACGTGA

(SEQ ID NO: 120)

Ef1a promoter

10 KRAB

hSOX2₁₋₁₇₉

DNMT3a3L

V5

pMki67::KRAB-hSOX2₁₋₁₇₉- DNMT3a3L-V5 [SES v1.1]

15 acttcaggccttgctccttgagaccttgaggagagggtagacattgggttctattgccccagaag
 cggaatTTTTtacaacaccctactggaagaacactgatccaggaacaggaggaggagcatga
 aaaaagaggacaggaaaacataggtactctatatctcatctgactggcacctggcctcagata
 accccaactgggtgccccaggagccctatttttaggatttgcaggtgcagtcttgcacagggtta
 ttttaatcggcatttccccctgaaacagcaaagcattgcaaacttaaccgtgccgcagcaat
 20 aactcttaagttatggtttcggaaaaaccgtgccactgcctcagtgttcattaagcaggaggag
 actgtagaatgaattataatgatgtccccagtttttaaaggacagtttgaggggatactgaa
 aaggcaggagaaacaaccttaggttgcgatgaggtgttcagaaaacggcagaggaagggta
 taaatTTTTtaaaatagaaagcttctagcgttgagcaagatgcggcggttgctcggtgtttgatt
 tctggagagggagcctacaggtagcttccctacaggctgccacgctagttttgctgacactga
 25 aaaagctgggaccctctaccctgcctatttagagcagcagtgggccgcccaccagccccacc
 tggccgccttccggagccttatctccgctcgggttccgcttctccagccagcttctggcccg
 ccccgccctaccggcctggccctcccaacctccggtccacccccccccccccccaggtc
 cgcagcctggccccacctctggacttccccaatcctctgtgcacctcgccattgggtctctgg
 ctggccgcctggcccaatcgcagcacttagcgccagaatttgaaccgcccgttttggtttgaa
 30 tggggcgggctgcggcgggcgagccttgggcgggaaaccgcctaggcgggacagtcggggctgg
 acggggcgggcggtggcggtggcgggcacttgggacggggcgggcgggagtcgggtctgg
 tggggcgggagcgaaggccgcgggtggcgtgggtcggtccctccgcggttaaggagccgaggg
 ctccgacgcgggctgcgccgggtgagcgggcgccagagctaaacttgcgctgactggaccagc
 tgaggagcgggcccgggcgggcgactgcgagcttccacgagaggttctccgcccctggtccgc
 35 agtcccgcagcgccggggcggaaccCGCGGATCCACCATTGGATGCTAAGTCACTGACTGCCTGGT
CCCGGACACTGGTGACCTTCAAGGATGTGTTTGTGGACTTCACCAGGGAGGAGTGGAAGCTG
CTGGACACTGCTCAGCAGATCCTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGT
TTCCTTGGGTTATCAGCTTACTAAGCCAGATGTGATCCTCCGTTGGAGAAGGGAGAAGAGC
CCTGGCTGGTGGAGAGAGAAATCACCAAGAGACCCATCCTGATTAGAGACTGCATTTGAA
 40 ATCAATCATCAGTTGGCAGCGGCATGC*atgtacaacatgatggagacggagctgaagcc*
gccccggccgcagcaaaacttcggggggcgggcgggcggaactccaccgcgggcgggcgggcg
gcaaccagaaaaacagcccgaccgcgtcaagcgggcccatgaatgccttcatgggtgtgtcc
cgcgggcgagcgggcgcaagatggcccaggagaaacccaagatgcacaactcgagatcagcaa

ggcgcctgggagccgagtggaacttttgtcggagacgggagaagcgccgttcacgcagcagg
 ctaagcgggctgcgagcgtgcacatgaaggagcaccggttataaataccggccccggcg
 aaaaccaagacgctcatgaagaaggataagtaacgctgcccggcggtgctggccccgg
 5 cggcaatagcatggcgagcgggtcgggtgggcgcggcctgggcgcggcgctgaaccagc
 gcatggacagttacgcgcacatgaacggctggagcaacggcagctacagcatgatgcaggac
 cagctgggcGCTAGCGGCAGCGGCAACCATGATCAAGAGTTCGATCCTCCTAAAGTGATCC
CCCTGTCCCTGCCGAAAAGCGGAAACCGATCAGGGTTCTCAGTCTCTTCGATGGAATTGCCA
CAGGCCTCTTGGTCCCTTAAGGACCTGGGCATCCAAGTgGAcCGCTACATCGCCTCAGAAGTA
 TGTGAgGACAGCATCACAGTCGGCATGGTGCGCCACCAGGGGAAAATAATGTACGTCGGCGA
 10 CGTCAGATCAGTTACTCAGAAACATATTCAGGAATGGGGGCCTTTTGATCTCGTAATTGGCG
 GCAGCCCCGTGCAATGATCTGAGTATTGTGAACCCCGCCCGGAAAGGTTTGTACGAAGGTACG
 GGCAGACTGTTCTTCGAGTTTACAGACTGTTGCACGACGCACGACCAAAGGAGGGTGAcGA
 tcGACCCCTTCTTCTGGCTGTTTCGAGAACGTGGTTCGCTATGGGCGTGTCTGACAAGCGGGACA
 TTTCTaGATTCTTGGAGAGCAATCCAGTGATGATTGATGCAAAGGAAGTAtccGCTGCCAC
 15 CGCGCCAGATACTTCTGGGGCAATCGCCCGGCATGAATCGACCCTTGGCAAGCACCGTGAA
 TGATAAGTTGGAGTTGCAGGAGTGTCTgGAGCATGGGCGGATCGCAAAGTTTAGCAAGGTGC
 GGACCATCACGACCCGAAGCAATTCAATTAAGCAGGGAAAGGACCAACATTTTCCAGTGTTT
 ATGAACGAGAAAGAGGAcATCCTGTGGTGcACAGAAATGGAGCGGGTTTTTCGGATTCCCCGT
 ACATTATACTGATGTGTCCAATATGTCTAGGTTGGCTAGACAGAGGCTGtTGGGACGCTCAT
 20 GGAGTGTTCTGTATCCGCCACCTGTTTCGCTCCTCTGAAGGAGTATTTTCGcTGCGTtTCA
 TCCGGgAATTCAAACGCAAACAGCAGAGGCCCATTCCTTTTCTTCCGGCCTGGTgCCACTTAG
 TCTGCGCGGCTCTCACATGGGACCTATGGAAATATACAAACAGTGAGCGCCTGGAAGAGGC
 AGCCTGTGCGGGTCCCTGAGTCTGTTTCAGAAACATTGACAAGGTGCTTAAGTCCCTTGATT
 CTGGAGTCTGGCAGCGGTTCTGGGGGTGgcACCCCTTAAGTATGTGGAGGACGTGACAAACGT
 25 CGTGAGGAGAGATGTGGAGAAATGGGGGCCATTTGACCTGGTATACGGGAGCACTCAACCTC
 TCGGTTCTTCTTGATCGCTGCCAGGGTGGTATATGTTTCAGTTCCATCGcATCCTGCAA
 TACGCCCTTCCGCGGCAGGAGAGTCAGCGACCATTCCTTCTGGATATTCATGGACAATCTCCT
 GCTGACAGAgGACGACCAAGAGACTACGACTAGATTTCTTCAGACCGAGGCCGTTACTCTCC
 AGGACGTTAGAGGTAGGGACTACCAGAATGCCATGCGAGTGTGGAGTAACATCCCCGGACTC
 30 AAGTCAAAGCAcGCACCCCTGACCCCAAGGAAGAGGAATACCTcCAGGCGCAGGTGAGGAG
 CCGCAGTAAGCTCGATGCTCCTAAAGTCGATTTGTTGGTGAAGAATTGCTTGCTGCCCTTGA
 GAGAGTACTTTAAGTACTTTAGCCAGAACAGCTTGCCACTGGGGCGCGCCCGGTAAGCCTATC
CCTAACCCCTCTCCTCGGTCTCGATTCTACGTGA

(SEQ ID NO: 121)

35 pMki67 promoter

KRABhSOX2₁₋₁₇₉DNMT3a3LV540 Ef1a::CS-hSOX2₁₋₁₇₉ [SESv2]

CGATAAGCTTTGCAAAGATGGATAAAGTTTTAAACAGAGAGGAATCTTTGCAGCTAATGGAC
 CTTCTAGGTCTTGAAAGGAGTGGGAATTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACAT
 CGCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGG

TGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGG
 GGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTTCGCAACGGGTTTGCCG
 CCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC
 CCTTGCGTGCCCTTGAATTACTTCCACTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGG
 5 GTTGGAAGTGGGTGGGAGAGTTCGAGGCCCTTGCCTTAAGGAGCCCTTCGCCTCGTGCTTG
 AGTTGAGGCCCTGGCCGGGCGCTGGGGCCGCCGCTGCGAATCTGGTGGCACCTTCGCGCCT
 GTCTCGCTGCTTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTT
 TTTTTCTGGCAAGATAGTCTTGTAATGCGGGCCAAGATCTGCACACTGGTATTTTCGGTTTTT
 TGGGGCCCGCGGGCGCGACGGGGCCCGTGCCTCCAGCGCACATGTTCCGGCAGGCGGGGCC
 10 TGCAGCGCGGGCCACCGAGAAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTG
 CCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTTCGGCACC
 AGTTGCGTGAGCGGAAAGATGCCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGA
 CGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCCGTCC
 TCAGCCGTGCTTCATGTGACTCCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTT
 15 CTCGAGCTTTTGGAGTACGTGCTCTTTAGGTTGGGGGAGGGGTTTTATGCGATGGAGTTTC
 CCCACACTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACCTGATGTAATTCCTCTTG
 GAATTTGCCCTTTTTGAGTTTGGATCTTGGTTTATTCTCAAGCCTCAGACAGTGGTTCAAAG
 TTTTTTTCTTCCATTTTCAGGTGTCGTGAGGAATTCTGCAGTCGATCGACGGTACCGCGGGCC
 CAACGGATCCACCATGCTGGAACCAGAAAAGATCATTGGGGCAACAGATTCTGTGGTGATT
 20 TAATGTTCTTAATGAAATGGAAAAGACACAGATGAAGCTGACCTGGTTCTTGCAAAAGAAGCT
 AATGTGAAATGTCCACAAATTGTGATAGCATTATGAAGAGAGACTGACATGGCATGCATA
 TCCTGAGGCATGCatgtacaacatgatggagacggagctgaagccgcccgggcccgcagcaaa
 cttcggggggcgggcgccggaactccaccgcgccggcgccggcgccggaaccagaaaaacagc
 ccggaccgctcaagcgcccatgaatgccttcatggtgtggtcccgcgggcagcgccgcaa
 25 gatggcccaggagaaccccaagatgcacaaactcggagatcagcaagcgcctgggcgcccagtg
 ggaaacttttgcggagacggagaagcgccgcttcatcgacgaggctaagcggtgcgagcg
 ctgcacatgaaggagcaccggattataaataccggccccggcggaacccaagacgctcat
 gaagaaggataagtacacgctgcccggcggtgctggcccccgcggaatagcatggcga
 gcggggtcgggtgggcgcccgcctgggcgcgccgctgaaccagcgcatggacagttacgcg
 30 cacatgaacggctggagcaacggcagctacagcatgatgcaggaccagctgggcGCTAGCGG
 CAGCGCGGCGCGCCCGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGTGA

(SEQ ID NO: 122)

Efla promoter

CS35 *hSOX2₁₋₁₇₉*V5Efla::SOX2₁₋₁₇₉-Y-R [SESv3]

CGATAAGCTTTGCAAAGATGGATAAAAGTTTTAAACAGAGAGGAATCTTTGCAGCTAATGGAC
 CTTCTAGGTCTTGAAAGGAGTGGGAATTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACAT
 40 CGCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTTCGGCAATTGAACCGGTGCCCTAGAGAAGG
 TGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGG
 GGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTTCGCAACGGGTTTGCCG
 CCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC

CTTGCGTGCCTTGAATTACTTCCACTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGG
 GTTGGAAGTGGGTGGGAGAGTTTCGAGGCCCTTGCCTTAAGGAGCCCCCTTCGCCCTCGTGCCTG
 AGTTGAGGCCCTGGCCCTGGGCGCTGGGGCCGCCGCTGCGAATCTGGTGGCACCTTCGCGCCT
 GTCTCGCTGCTTTTCGATAAGTCTCTAGCCATTTAAAATTTTGGATGACCTGCTGCGACGCTT
 5 TTTTCTGGCAAGATAGTCTTGTAATGCGGGCCAAGATCTGCACACTGGTATTTTCGGTTTT
 TGGGGCCGCGGGCGGCGACGGGGCCCGTGCCTCCAGCGCACATGTTTCGGCGAGGCGGGGCC
 TGCAGCGCGGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTG
 CCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTTCGGCACC
 AGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGA
 10 CGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCCGTCC
 TCAGCCGCTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTT
 CTCGAGCTTTTGGAGTACGTCTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTC
 CCCACACTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACCTTGATGTAATTCCTCCTTG
 GAATTTGCCCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAG
 15 TTTTTTCTTCCATTTTCAGGTGTCGTGAGGAATTCGTCAGTCGATCGACGGTACCGCGGGGCC
 CAACGGATCCACCatgtacaacatgatggagacggagctgaagccgccccgccccgcagcaaa
 cttcggggggcgggcgggcgaactccaccgcgcgcgcgcgcgcggaaccagaaaaacagc
 cgggaccgctgaagcgggcccatgaatgccttcatggtgtggtcccgcgggcagcgcgcaa
 gatggcccaggagaaccccaagatgcacaactcggagatcagcaagcgctggcgccgagt
 20 ggaaacttttgtcggagacggagaagcgggcgttcatcgacgaggctaagcggtgagagc
 ctgcacatgaaggagcaccggtattataaataccggccccggcggaaccagacgctcat
 gaagaaggataagtacacgctgcccggcggtgctggccccggcggaatagcatggcga
 gcggggtcggggtggcgcgcggtggtgagcggcggtgaaccagcgcatggacagttacgag
 cacatgaacggctggagcaacggcagctacagcatgatgcaggaccagctggcgctagcGG
 25 ATCCaacATGAGGCCCGGCTGAAAAACGTGGACAGGAGCACTGCACAGCAGTTGGCAGTAA
 CTGTGGGCAACGTCACCGTCATTATCACAGACTTTAAGGAAAAGACTCGCTCCTCATCGACA
 TCCTCATCCACAGTGACCTCCAGTGCAGGGTCAGAACAGCAGAACAGGGCGCGCCCGGTAA
GCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGTGA

(SEQ ID NO: 123)

30 **Ef1a promoter**

*hSOX2₁₋₁₇₉*Y-RV5Ef1a::KRAB-hTEAD₁₋₁₆₆- DNMT3a3L-V5 [TES]

35 CGATAAGCTTTGCAAAGATGGATAAAGTTTAAACAGAGAGGAATCTTTGCAGCTAATGGAC
 CTTCTAGGTCTTGAAAGGAGTGGGAATTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACAT
 CGCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTGGCAATTGAACCGGTGCCTAGAGAAGG
 TGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCTTTTTCCCGAGGGTGG
 GGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTTCGCAACGGGTTTGCCG
 40 CCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTATGGC
 CTTGCGTGCCTTGAATTACTTCCACTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGG
 GTTGGAAGTGGGTGGGAGAGTTTCGAGGCCCTTGCCTTAAGGAGCCCCCTTCGCCCTCGTGCCTG
 AGTTGAGGCCCTGGCCCTGGGCGCTGGGGCCGCCGCTGCGAATCTGGTGGCACCTTCGCGCCT

GTCTCGCTGCTTTTCGATAAGTCTCTAGCCATTTAAAATTTTGGATGACCTGCTGCGACGCTT
TTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTTCGGTTTTT
TGGGGCCGCGGGCGGCGACGGGGCCCGTGCCTCCCAGCGCACATGTTTCGGCGAGGCGGGGCC
TGCAGCGCGGGCCACCGAGAAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTG
5 CCTGGCCTCGCGCCCGCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTGCGCACC
AGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGA
CGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCCGTCC
TCAGCCGTGCTTTCATGTGACTCCACGGAGTACCGGGCGCGTCCAGGCACCTCGATTAGTT
CTCGAGCTTTTGGAGTACGTCTCTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTC
10 CCCACACTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCCTCTTG
GAATTTGCCCTTTTTGAGTTTGGATCTTGGTTTCACTTCAAGCCTCAGACAGTGGTTCAAAG
TTTTTTTCTTCCATTTTCAGGTGTCGTGAGGAATTCTGCAGTCGATCGACGGTACCGCGGGCC
CAACGGATCCACCATGGATGCTAAGTCACTGACTGCCTGGTCCCGGACACTGGTGACCTTCA
AGGATGTGTTTTGTGGACTTCACCAGGGAGGAGTGAAGCTGCTGGACACTGCTCAGCAGATC
15 CTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTCCTTGGGTTATCAGCTTAC
TAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCCTGGCTGGTGGAGAGAGAAA
TTCACCAAGAGACCCATCCTGATTGAGAGACTGCATTTGAAATCAAATCATCAGTTGGCAGC
GGCGCATGCATTGAGCCCAGCAGCTGGAGCGGCAGTGAGAGCCCTGCCGAAAACATGGAAAG
GATGAGTGAATCTGCAGATAAGCCAATTGACAATGATGCAGAAGGGGTCTGGAGCCCCGACA
20 TCGAGCAAAGCTTTCAGGAGGCCCTGGCTATCTATCCACCATGTGGGAGGAGGAAAATCATC
TTATCAGACGAAGGCAAAATGTATGGTAGGAATGAATTGATAGCCAGATACATCAAATCAG
GACAGGCAAGACGAGGACCAGAAAACAGGTGTCTAGTCACATTCAGGTTCTTGCCAGAAGGA
AATCTCGTGATTTTCATTCCAAGCTAAAGGATCAGACTGCAAAGGATAAGGCCCTGCAGCAC
ATGGCGGCCATGTCTCAGCCAGATCGTCTCGGCCACTGCCATTATAACAAGCTGGGGCT
25 GCCTGGGATTCCACGCCCGACCTTCCAGGGGCGCGGGGTTCTGGCCGGGAATGATTCAAA
CAGGGCAGCCAGCTAGCGGCAGCGGCAACCATGATCAAGAGTTCGATCCTCCTAAAGTGAT
CCCCCTGTCCCTGCCGAAAAGCGGAAACCGATCAGGGTCTCAGTCTCTTCGATGGAATTGC
CACAGGCCTCTTGGTCCCTAAGGACCTGGGCATCCAAGTgGAcCGCTACATCGCCTCAGAAG
TATGTGAgGACAGCATCACAGTCGGCATGGTGCGCCACCAGGGGAAAATAATGTACGTGGC
30 GACGTGAGATCAGTTACTCAGAAAATATTCAGGAATGGGGGCCTTTTGATCTCGTAATTGG
CGGCAGCCCCTGCAATGATCTGAGTATTGTGAACCCCGCCCGGAAAGGTTTGTACGAAGGTA
CGGGCAGACTGTTCTTCGAGTTTACAGACTGTTGCACGACGCACGACCAAAGGAGGGTGAc
GATcGACCCCTTCTTCTGGCTGTTTCGAGAACGTGGTTCGCTATGGGCGTGTCTGACAAGCGGGA
CATTTCTAGATTCTTGGAGAGCAATCCAGTGATGATTGATGCAAAGGAAGTATccGCTGCCC
35 ACCGCGCCAGATACTTCTGGGGCAATCTGCCCGGCATGAATCGACCCTTGGCAAGCACCGTG
AATGATAAGTTGGAGTTGCAGGAGTGCTgGAGCATGGGCGGATCGCAAAGTTTAGCAAGGT
GCGGACCATCACGACCCGAAGCAATTCAATTAAGCAGGGAAAGGACCAACATTTTCCAGTGT
TTATGAACGAGAAAGAGGAcATCCTGTGGTGcACAGAAATGGAGCGGGTTTTTCGGATTCCCC
GTACATTATACTGATGTGTCCAATATGTCTAGGTTGGCTAGACAGAGGCTGtTGGGACGCTC
40 ATGGAGTGTTCCTGTCTATCCGCCACCTGTTTCGCTCCTCTGAAGGAGTATTTCGCcTGCgtT
CATCCGGgAATTCAAACGCAAACAGCAGAGGCCCATCCTTTTCTTCCGGCCTGGTgCCACTT
AGTCTGCGCGGCTCTCACATGGGACCTATGGAAATATACAAAACAGTGAGCGCCTGGAAGAG
GCAGCCTGTGCGGGTCTGAGTCTGTTTCAGAAACATTGACAAGGTGCTTAAGTCCCTTGGAT
TTCTGGAGTCTGGCAGCGGTTCTGGGGGTGgcACCCCTAAGTATGTGGAGGACGTGACAAAC
45 GTCTGAGGAGAGATGTGGAGAAATGGGGGCCATTTGACCTGGTATACGGGAGCACTCAACC
TCTCGGTTCTTCTTGTGATCGCTGCCAGGGTGGTATATGTTTCAGTTCCATCGcATCCTGC
AATACGCCCTTCCGCGGCAGGAGAGTCAGCGACCATTTCTTGGATATTATGGACAATCTC
CTGCTGACAGAgGACGACCAAGAGACTACGACTAGATTTCTTTCAGACCGAGGCCGTTACTCT

5 CCAGGACGTTAGAGGTAGGGACTACCAGAATGCCATGCGAGTGTGGAGTAACATCCCCGGAC
TCAAGTCAAAGCAcGCACCCCTGACCCCCAAGGAAGAGGAATACCTcCAGGCGCAGGTGAGG
AGCCGCAGTAAGCTCGATGCTCCTAAAGTCGATTTGTTGGTGAAGAATTGCTTGCTGCCCT
GAGAGAGTACTTTAAGTACTTTAGCCAGAACAGCTTGCCACTGGGCGCGCCCGGTAAGCCTA
TCCCTAACCCCTCTCCTCGGTCTCGATTCTACGTGA

(SEQ ID NO: 124)

Ef1a promoter

KRAB

hTEAD₁₋₁₆₆

10 DNMT3a3L

V5

Ef1a::KRAB-DNMT3a3L-hMYC₁₄₄₋₄₅₄-V5 [MES]

CGATAAGCTTTGCAAAGATGGATAAAAGTTTTTAAACAGAGAGGAATCTTTGCAGCTAATGGAC
 CTTCTAGGTCTTGAAAGGAGTGGGAATTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACAT
 15 CGCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCCTAGAGAAGG
 TGGCGCGGGGTAAAC'TGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGG
 GGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTTCGCAACGGGTTTGCCG
 CCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGC
 CCTTGCGTGCCCTGAATTACTTCCACTGGCTGCAGTACGTGATTCTTGATCCCAGCTTCGG
 20 GTTGGAAGTGGGTGGGAGAGTTCGAGGCCCTTGCGCTTAAGGAGCCCCCTTCGCCCTCGTGCTTG
 AGTTGAGGCCCTGGCC'TGGGCGCTGGGGCCGCCGCTGCGAATCTGGTGGCACCTTCGCGCCT
 GTCTCGCTGCTTTTCGATAAGTCTCTAGCCATTTAAATTTTTGATGACCTGCTGCGACGCTT
 TTTTTCTGGCAAGATAGTCTTGTAATGCGGGCCAAGATCTGCACACTGGTATTTTCGGTTTTT
 TGGGGCCCGGGCGGCGACGGGGCCCGTGGCTCCAGCGCACATGTTTCGGCGAGGCGGGGCC
 25 TGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTG
 CCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTTCGGCACC
 AGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGA
 CGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCAACACACAAAGGAAAAGGGCCTTTCCGTCC
 TCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTT
 30 CTCGAGCTTTTGGAGTACGTGCTCTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTC
 CCCACACTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCCTCTTG
 GAATTTGCCCTTTTTGAGTTTGGATCTTGGTTTCAATTCTCAAGCCTCAGACAGTGGTTCAAAG
 TTTTTTCTTCCATTTTCAGGTGTCGTGAGGAATTCTGCAGTCGATCGACGGTACCGCGGGGCC
 CAACGGATCCACCATGGATGCTAAGTCACTGACTGCCTGGTCCCGGACACTGGTGACCTTCA
 35 AGGATGTGTTTGTGGACTTCACCAGGGAGGAGTGGAAGCTGCTGGACACTGCTCAGCAGATC
 CTGTACAGAAATGTGATGCTGGAGAACTATAAGAACCTGGTTTTCTTGGGTTATCAGCTTAC
 TAAGCCAGATGTGATCCTCCGGTTGGAGAAGGGAGAAGAGCCCTGGCTGGTGGAGAGAGAAA
 TTCACCAAGAGACCCATCCTGATTCAGAGACTGCATTTGAAATCAAATCATCAGTTGCTAGC
 GGCAGCGGCAACCATGATCAAGAGTTCGATCCTCCTAAAGTGTATCCCCCTGTCCCTGCCGA
 40 AAAGCGGAAACCGATCAGGGTTCAGTCTCTTCGATGGAATTGCCACAGGCCTCTTGGTCC
TTAAGGACCTGGGCATCCAAGTgGAcCGCTACATCGCCTCAGAAGTATGTGAqGACAGCATC
ACAGTCGGCATGGTGCGCCACCAGGGGAAAATAATGTACGTCGGCGACGTCAGATCAGTTAC
TCAGAAACATATTCAGGAATGGGGGCCTTTTGATCTCGTAATTGGCGGCAGCCCCTGCAATG

ATCTGAGTATTGTGAACCCCGCCCGGAAAGGTTTGTACGAAGGTACGGGCAGACTGTTCTTC
 GAGTTTTACAGACTGTTGCACGACGCACGACCAAGGAGGGTGAcGATcGACCCTTCTTCTG
 GCTGTTTCGAGAACGTGGTCGCTATGGGCGTGTCTGACAAGCGGGACATTTCTAGATTCCCTGG
 AGAGCAATCCAGTGATGATTGATGCAAAGGAAGTAtccGCTGCCACCGCGCCAGATACTTC
 5 TGGGGCAATCTGCCCGGCATGAATCGACCCCTGGCAAGCACCGTGAATGATAAGTTGGAGTT
 GCAGGAGTGTCTgGAGCATGGGCGGATCGCAAAGTTTAGCAAGGTGCGGACCATCACGACCC
 GAAGCAATTCAATTAAGCAGGGAAAGGACCAACATTTTCCAGTGTTTATGAACGAGAAAGAG
 GAcATCCTGTGGTGcACAGAAATGGAGCGGGTTTTCCGATTCCCCGTACATTATACTGATGT
 GTCCAATATGTCTAGGTTGGCTAGACAGAGGCTGtTGGGACGCTCATGGAGTGTTCTGTCA
 10 TCCGCCACCTGTTTCGCTCCTCTGAAGGAGTATTTTCGcTGCGTtTCATCCGGgAATTCAAAC
 GCAAACAGCAGAGGCCCATCCTTTTCTTCCGGCCTGGTgCCACTTAGTCTGCGCGGCTCTCA
 CATGGGACCTATGGAAATATACAAAACAGTGAGCGCCTGGAAGAGGCAGCCTGTGCGGGTCC
 TGAGTCTGTTTCAGAAACATTGACAAGGTGCTTAAGTCCCTTGGATTTCTGGAGTCTGGCAGC
 GGTTCCTGGGGGTGgcACCCTTAAGTATGTGGAGGACGTGACAAACGTCGTGAGGAGAGATGT
 15 GGAGAAATGGGGGCCATTTGACCTGGTATACGGGAGCACTCAACCTCTCGGTTCTTCTTGTG
 ATCGCTGCCCAGGGTGGTATATGTTTCAGTTCCATCGcATCCTGCAATACGCCCTTCCGCGG
 CAGGAGAGTCAGCGACCATTCTTCTGGATATTATGGAACAATCTCCTGCTGACAGAgGACGA
 CCAAGAGACTACGACTAGATTCTTTCAGACCGAGGCCGTTACTCTCCAGGACGTTAGAGGTA
 GGGACTACCAGAATGCCATGCGAGTGTGGAGTAACATCCCCGGACTCAAGTCAAAGCAcGCA
 20 CCCCTGACCCCCAAGGAAGAGGAATACCTcCAGGCGCAGGTGAGGAGCCGCAGTAAGCTCGA
 TGCTCCTAAAGTCGATTTTGTGGTGAAGAATTGCTTGCTGCCCCCTGAGAGAGTACTTTAAGT
 ACTTTAGCCAGAACAGCTTGCCACTGGGCGCGCCCGCCGCTGCCATCATCCAGGACTGTATG
 TGGAGCGGCTTCTCGGCCGCCGCCAAGCTCGTCTCAGAGAAGCTGGCCTCCTACCAGGCTGC
 GCGCAAAGACAGCGGCAGCCCCGAACCCCGCCCGGCCACAGCGTCTGCTCCACCTCCAGCT
 25 TGTACCTGCAGGATCTGAGCGCCGCCGCTCAGAGTGcATCGACCCCTCGGTGGTCTTCCCC
 TACCTCTCAACGACAGCAGCTCGCCCCAAGTCTCGCGCTCGCAAGACTCCAGCGCCTTCTC
 TCCGTCTCGGATTCTCTGCTCTCCTCGACGGAGTCTCTCCCGCAGGGCAGCCCCGAGCCCC
 TGGTGCTCCATGAGGAGACACCGCCCACCACCAGCAGCGACTCTGAGGAGGAACAAGAAGAT
 GAGGAAGAAATCGATGTTGTTTCTGTGGAAAAAGAGGCAGGCTCCTGGCAAAAGGTGAGATC
 30 TGGATCACCTTCTGCTGGAGGCCACAGCAAACCTCCTCACAGCCCACTGGTCTCAAGAGGT
 GCCACGTCTCCACACATCAGCACAACTACGCAGCGCCTCCCTCCACTCGGAAGGACTATCCT
 GCTGCCAAGAGGGTCAAGTTGGACAGTGTCAGAGTCTTGAGACAGATCAGCAACAACCGAAA
 ATGCACCAGCCCCAGGTCTCGGACACCGAGGAGAATGTCAAGAGGCGAACACACAACGTCT
 TGGAGCGCCAGAGGAGGAACGAGCTAAAACGGAGCTTTTTTGCCCTGCGTGACCAGATCCCCG
 35 GAGTTGGAAAACAATGAAAAGGCCCCCAAGGTAGTTATCCTTAAAAAAGCCACAGCATACAT
 CCTGTCCGTCCAAGCAGAGGAGCAAAAGCTCATTTCTGAAGAGGACTTGTTGCGGAAACGAC
 GAGAACAGTTGAAACACAACTTGAACAGCTACGGAACCTTGTGCGGCTAGCGGCGCGCCC
 GGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGTGA

(SEQ ID NO: 125)

40 **Efla promoter**KRABDNMT3a3L*hMYC₁₄₄₋₄₅₄*V5

Lentivirus production

Replication-incompetent, VSVg-coated lentiviral particles were packaged in 293 T cells. Cells were transfected with 30 µg of vector and packaging constructs, according to a conventional CaCl₂ transfection protocol. After 30 hrs, medium was collected, filtered
5 through 0.44 µm cellulose acetate and centrifuged at 20000 rpm for 2 hrs at 20°C in order to concentrate the virus.

Cell culture

U-87, U-251 and SNB-19 (human glioblastoma cell lines), HeLa (human cervix cancer cell line), DU145 (human prostatic cancer cell line) and HepG2 (human liver carcinoma cell line)
10 were cultured in plastic-adherence conditions in DMEM medium (Dulbecco's Modified Eagle's Medium - high glucose, Sigma-Aldrich) containing 10% fetal bovine serum (FBS, Sigma-Aldrich), 1% Pen/Strept (Sigma-Aldrich), 2 mM Glutamine (Sigma-Aldrich), 1% non-essential amino acids (MEM NEAA, ThermoFisher Scientific), 1% sodium pyruvate solution (Sigma-Aldrich) and passaged twice a week using Trypsin-EDTA solution (Sigma-Aldrich).

15 BxPC3 (human pancreatic cancer cell line) were cultured in plastic-adherence conditions in RPMI-1640 (Sigma-Aldrich) containing 10% FBS, 1% Pen/Strept, 2 mM Glutamine. All the cell lines were passaged twice a week using Trypsin-EDTA solution (Sigma-Aldrich).

Cancer stem cells (CSCs) from classical (L0627) and mesenchymal (1312) glioblastoma tumors were maintained in spheres in suspension cultures in DMEM/F12 (Sigma-Aldrich)
20 supplemented with Hormon Mix (DMEM/F12, 0.6% Glucose (Sigma-Aldrich) (30% in phosphate buffer (PBS) (Euroclone)), Insulin (Sigma-Aldrich) 250 µg/ml, putrescine powder (Sigma-Aldrich) 97 µg/ml, apotransferrin powder (Sigma Aldrich), sodium selenite 0.3 µM, progesterone 0.2 µM), 1% Pen/Strept, 2 mM Glutamine, 0.66% Glucose (30% in phosphate buffer salt (PBS) (Euroclone)), and heparin (4 mg/ml, Sigma-Aldrich); bFGF (20 ng/ml,
25 ThermoFisher Scientific) and EGF (20 ng/ml, ThermoFisher Scientific) were added freshly to culture medium. Sphere cultures were passaged once a week by mechanical dissociation of the sphere to single cell suspension.

All the cultures were kept in humidified atmosphere of 5% CO₂ at 37°C under atmospheric oxygen conditions.

30 Growth curve analysis

5x10⁵ of cancer cell lines were seeded in adherent condition in a 6 multi-well plate at day 0; at day 1 cultures were infected with lentiviral vectors and at day 3, cells were detached, live

cells were stained with Trypan Blue Solution (0.4%, ThermoFisher Scientific) and counted using Countess™ II Automated Cell Counter (ThermoFisher Scientific); after this passage, 3×10^5 cells were seeded again. This was repeated for 3 time points every 3-4 days; the experiment was repeated 3 times for each time point. Brightfield representative pictures were taken at each time-point.

25×10^4 CSCs were seeded in single cell suspension condition in a 24 multi-well plate at day 0; at day 1 cultures were infected with lentiviral vectors (expressing either SES or GFP). At day 4, day 7, day 10, and day 13 CSC spheres were dissociated to single cell suspension and live cells were stained with Trypan Blue Solution and counted as previously described. Live cell number and the % of dead cells were reported on graphs for each time point; the experiment was repeated 3 times for each time-point. Bright-field representative pictures were taken at each time-point.

Western blot analysis

Cells were homogenized in RIPA buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, SDS (0.1% for cells, 1% for 3D cultures), 1% Triton X-100, Roche Complete EDTA-free Protease Inhibitor Cocktail, Roche PhosSTOP EASYpack) and Western blot analysis was performed incubating primary antibodies overnight at 4°C in blocking solution composed of 5% BSA (Sigma-Aldrich) or 5% Non-Fat Dry Milk in PBS-TWEEN 0.1% (Sigma-Aldrich) according to antibody datasheet. The primary antibodies utilized were as follows: anti-V5 (mouse, 1:1000, ThermoFisher Scientific, R96025), anti-SOX2 (Clone # 245610, mouse, 1:500, R&D system, MAB2018), anti-Histone H3 (rabbit, 1:2000, Abcam, ab1791). Band densitometry relative to control was calculated using Fiji software (NIH, USA), normalized on housekeeping (H3).

Clonogenic assay

25×10^4 CSCs were seeded in single cell suspension condition in a 24 multi-well plate at day 0; at day 1 cultures were or were not infected with lentiviral vectors (expressing either SES or GFP). At day 6 spheres were dissociated to single cell suspension and live cells were counted as previously described and 25×10^4 CSCs were seeded again, letting cells grow and form spheres until day 10. Bright-field images were taken at day 6 and day 10 and the resulting number of sphere was counted for each condition (not infected, GFP-infected or SES-infected); sphere diameter was measured and the % of sphere having a diameter < 100 μ m was reported on the graph for each condition and time-point. The experiment was repeated 3 times for each time point.

RNA isolation and real-time RT-qPCR

RNA was extracted using the TRI Reagent isolation system (Sigma-Aldrich) according to the manufacturer's instructions. For quantitative RT-PCR (qRT-PCR), one microgram of RNA was reverse transcribed using the ImProm-II Reverse Transcription System (Promega).
5 Thereafter qRT-PCR was performed in triplicate with custom designed oligos using the CFX96 Real-Time PCR Detection System (Bio-Rad, USA) with the Titan HotTaq EvaGreen qPCR Mix (BIOATLAS). Obtained cDNA was diluted 1:10 and was amplified in a 16 µl reaction mixture containing 2 µl of diluted cDNA, 1× Titan Hot Taq EvaGreen qPCR Mix (Bioatlas, Estonia) and 0.4 mM of each primer. Analysis of relative expression was
10 performed using the $\Delta\Delta C_t$ method, using 18S rRNA as a housekeeping gene and CFX Manager software (Bio-Rad, USA).

RNA-sequencing

RNA libraries were generated starting from 1 mg of total RNA (deriving from U87, SNB19 and murine hippocampi) the quality of which was assessed by using a TapeStation
15 instrument (Agilent). To avoid over-representation of 30 ends, only high-quality RNA with a RNA Integrity Number (RIN) R 8 was used. RNA was processed according to the TruSeq Stranded mRNA Library Prep Kit protocol. The libraries were sequenced on an Illumina HiSeq 3000 with 76 bp stranded reads using Illumina TruSeq technology. Image processing and base call were performed using the Illumina Real Time Analysis Software. Fastq files
20 were aligned to hg19 or mm10 human or mouse reference genomes by using the splice junction map- per TopHat. Differential gene expression and Functional enrichment analyses were performed with DESeq2 and GSEA, respectively.

ChIP-sequencing

Chromatin was isolated from SNB19. Cells were plated in adherent condition using Matrigel
25 coated 15 mm plates at a density of 6×10^6 per plate. When the plates reached 90% of confluence, cells were fixed by adding formaldehyde directly to the cell culture media to reach a final concentration of 1%, and then were incubated for 10 min at RT. The reaction was quenched adding glycine to a final concentration of 125 mM and incubated 5 min at RT. Medium was then removed and cells were washed 3 times with cold, sterile PBS+protease
30 inhibitor, then cells were gently scraped and collected for centrifugation at 4°C for 50 at 1200 rpm. For ChIP experiments, collected cell pellets were lysed in lysis buffer (50 mM Tris-HCl pH 8, 0.1% SDS, 10 mM EDTA pH 8, 1 mM phenylmethylsulfonyl fluoride (PMSF, Sigma # P7626), protease inhibitor cocktail (Roche #04693159001)) and chromatin was sonicated with a Branson D250 sonifier (4 cycles of 30 s, 20% amplitude), to reach an average

fragment size of 0.1–0.5 kb. Following quantification, 100 mg of sonicated chromatin was used in each immunoprecipitation and incubated overnight at 4° C with 4 mg of V5 antibody (mouse, 1:5, ThermoFisher Scientific, R96025).

ChIP-seq libraries were produced using 5 ng of each immunoprecipitated and purified DNA.

5 End repair of DNA fragments was achieved by sequential 15 min incubations at 12°C and 25°C with 0.15 U/ml T4 PNK (NEB #M0201L), 0.04 U/ml T4 POL (NEB #M0203L) and 0.1 mM dNTPs (NEB #N0446S). A-base addition was performed by an incubation with 0.25 U/ml of Klenow fragment (NEB # M0212L) and 167 mM dATP (NEB N0440S) for 30 min at 30°C. Adaptor ligation was achieved by using the Quick ligation kit (NEB #M2200L) and performing

10 an incubation of 15 min at 25°C. DNA fragments were finally amplified for 14 cycles, by using the PfuUltra II Fusion HS DNA Pol kit (Agilent #600674). DNA purification steps after each enzymatic reaction were performed using Agencourt AMPure XP SPRI beads (Beckman #A63882). The obtained libraries were quality controlled at an Agilent Bioanalyzer (Agilent Technologies #G2943CA) before sequencing with Illumina HiSeq 2000. Sequencing

15 read quality was assessed by using fastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and total reads were aligned to the human genome (hg19) using Bowtie2 version 2.2.3 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>). Only uniquely mapped reads were used in the subsequent analyses, with an average mappability > 96% of initial total reads. Normalized BigWig tracks of ChIP-seq

20 seq experiments were generated with bedtools 2.24.0 (<https://bedtools.readthedocs.io/en/latest/>) and the bedGraphTo-BigWig program (<https://www.encodeproject.org/software/bedgraphtobigwig/>) and visualized in the UCSC Genome Browser (<http://genome.ucsc.edu/>). In order to find the regions of ChIP-seq enrichment over background, we used SICER V1.1 ([https://](https://home.gwu.edu/~wpeng/Software.html)

25 home.gwu.edu/~wpeng/Software.html) (window size = 200; gap size = 200; FDR < 0.01 parameters for all ChIP-seq data, FDR 0.01 parameters were used). Density plots (±10 kb) were generated with the ngsplot 2.47 ([https:// github.com/shenlab-sinai/ngsplot](https://github.com/shenlab-sinai/ngsplot)) command ngs.plot.r and eventually re-plotted with GraphPad Prism.

MeDIP sequencing

30 1 µg of purified genomic DNA (gDNA) was used with qiAMP DNA mini kit (Qiagen, cat. 51304). In brief, for methylated DNA immune precipitation and purification MagMeDIPseq kit was used (Diagenode, cod. C02010040). First, gDNA was sonicated to obtain a fragment size between 150-300 bp, then, it was de-natured to ssDNA and immune-precipitated using α-methyl-cytosine antibody provided by the kit. The next day, immunoprecipitated DNA and

35 Input were purified and eluted. Library preparation was performed using NEBNext Ultra II kit

for Illumina (cod.E7645), following the manufacturer's instructions. Each library was dual indexed using NEBNext Multiplex Oligos for Illumina (cod.E6440) and sequenced at 30 million pair-end depth with Illumina HiSeq 2000. First Adaptor trimming was performed using Trimmomatic (<http://www.usadellab.org/cms/?page=trimmomatic>). Trimmed reads were then
5 aligned to reference Hg19 genome using Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>). To obtain Coverage tracks bamCoverage was used (<https://deeptools.readthedocs.io/en/develop/content/tools/bamCoverage.html>), BAM files were converted to BigWigs using CPM normalization and effective genome size parameters with a bin size of 10. Peaks were called using Macs2 (<https://github.com/macs3-project/MACS>). Normal peak calling mode was used, in paired end mode with a q-value set
10 at 0.05. Differential peaks were called with Macs2, using the BedGraph obtained comparing BAM files of treated against control condition. Afterwards, differential peaks were intersected with normal peaks for each condition in order to filter differential peaks, actually present in at least one of the two conditions. Density plots (± 10 kb) were generated with the ngsplot 2.47
15 (<https://github.com/shenlab-sinai/ngsplot>) command ngs.plot.r and eventually re-plotted with GraphPad Prism.

Xenografts

GBM lines or Cancer stem cells L0627 were seeded in 6-well dishes and infected with 10 μ l LV-EIF1 α -SES/TES/MES or 5 μ l LV-EIF1 α -GFP each well for 48 hours, as described in the
20 infection section.

Heterotopic xenografts. The infected cells were counted and resuspended 3×10^6 cells in 100 μ l of Matrigel (Matrigel growth factor reduced, Corning). By using 1 ml syringes previously cooled at 20°C, GFP-infected cells were subcutaneously injected into the left flank of NOD-SCID mice (NOD.cg-Prkdc scid Il2rg tm1Wjl/SzJ), whereas SES/TES/MES-
25 infected cells were subcutaneously injected into the right flank of the same animal. Mice were sacrificed at 1-3 months after injections (according with growth rate) and subcutaneous growing tumors were extracted and fixed in 4% PFA for at least 24 hours. Tumor samples were sized and kept overnight in Sucrose 30% in PBS and then embedded in O.C.T. for cryopreservation. Histological slides were cut in 50- μ m sections on a cryostat (CM1850 UV,
30 Leica). Subsequently, the sections were processed for immunofluorescence or mounted on gelatin-coated glass slides and processed for Nissl staining.

Intracranial xenograft. The infected cells were counted and resuspended 3×10^5 cells in 3 μ l PBS 1X and then were unilaterally injected in the striatum of NOD-SCID mice (AP +0.5; ML ± 1.8 ; DV -3.3 from skull). Mice were sacrificed upon observation of general condition or after

40 days from the U87 injection or after 3-5 weeks of CSC injection; following anesthesia, mice were transcardially perfused with 4% PFA in PBS, then brains were removed from the skull and kept in the same solution for overnight fixation. After fixation, brains were kept overnight in Sucrose 30% in PBS and then embedded in O.C.T. for cryopreservation. The samples were cut coronally in 50- μ m sections on a cryostat (CM1850 UV, Leica). Subsequently, the sections were processed for immunofluorescence or mounted on gelatin-coated glass slides and processed for Nissl staining.

In vivo treatment (with U87). U87 orthotopic xenograft were induced as described before but with 75000 naïve cells. After 4 days the mice, randomly divided in two groups were injected at the same topological coordinates with LV carrying either GF or SES. A cohort of animals was sacrificed after 26 days from the LV injection and another cohort was left alive for performing survival rate and sacrificed upon observation of general condition or at 90 days after the first surgery; following anesthesia, mice were transcardially perfused with 4% PFA in PBS, then brains were removed from the skull and kept in the same solution for overnight fixation. After fixation, brains were kept overnight in Sucrose 30% in PBS and then embedded in O.C.T. for cryopreservation. The samples were cut coronally in 50- μ m sections on a cryostat (CM1850 UV, Leica). Subsequently, the sections were processed for immunofluorescence or mounted on gelatin-coated glass slides and processed for Nissl staining.

In vivo treatment (with CSCs). Classical naïve CSCs orthotopic xenograft were induced as described previously (3×10^5 cells). After 7 days the mice, randomly divided in two groups, were injected at the same topological coordinates with LV carrying either GFP or TES and sacrificed after 3 weeks from the LV injection; following anesthesia, mice were transcardially perfused with 4% PFA in PBS, then brains were removed from the skull and kept in the same solution for O/N fixation. After fixation, brains were kept O/N in Sucrose 30% in PBS and then embedded in O.C.T. for cryopreservation. The samples were cut coronally in 50- μ m sections on a cryostat (CM1850 UV, Leica). Subsequently, the sections were processed for immunofluorescence or mounted on gelatin-coated glass slides and processed for Nissl staining.

***In vivo* SES delivery in WT animals**

Hippocampi of Wt C57BL/6 animals were injected with LV carrying either GFP or SES (2 injections per hippocampus AP -2.8, ML +/- 3, DV -3.5; -2.5., 0.8 μ l each). After 1 month from the surgery the animals were tested for behavioral tasks and sacrificed for molecular and histological analyses.

Immunostaining

Cells were seeded on glass coverslips (for CSCs previously coated with Matrigel to allow cell adhesion) and they were fixed for 20 minutes on ice in 4% paraformaldehyde (PFA, Sigma), solution in phosphate-buffered saline (PBS, Euroclone). Then they were washed twice with
5 PBS and were permeabilized for 30' in blocking solution, containing 0.2% Triton X-100 (SigmaAldrich) and 5% donkey serum (Euroclone), and incubated overnight at 4°C with the primary antibodies diluted in blocking solution. The primary antibodies utilized were as follows: anti-V5 (mouse, 1:500, ThermoFisher Scientific, R96025), anti-GFP (chicken, 1:1000, Thermo Fisher Scientific, A10262), anti-MAP2 (chicken, 1:1000, Abcam, ab92434),
10 anti-phospho-Histone H3 (Ser10, rabbit, 1:200, Sigma-Aldrich, 06-570), anti-Cleaved Caspase-3 (Asp175, rabbit, 1:200, Cell Signaling Technology, 9661), anti-Ki-67 (Clone SP6, rabbit, 1:500, Immunological Sciences, MAB-90948), anti-Human Nuclei (mouse, 1:500, Millipore, MAB1281). The next day, cells were washed 3 times with PBS for 5 minutes and incubated for 1 hour at room temperature with Hoechst 33342 (ThermoFischer Scientific)
15 and with secondary antibodies (ThermoFisher Scientific) in blocking solution. Brain sections were blocked in 10% donkey serum and 0.2% Triton X-100 for 1 hr at RT. Incubation with primary antibodies was performed at 4°C overnight. Secondary antibodies were applied to sections for 2 hrs at RT in blocking solution containing Hoechst 33342. Finally, slices were washed and mounted in Fluorescent Mounting Medium (Dako Cytomation). Images were
20 acquired with epifluorescence microscope Nikon DS-Qi2 and analyzed with Fiji software.

Nissl staining

Brain sections were rinsed in distilled H₂O for 1 min, then stained in 0.1% Cresyl Violet solution boiled at 50°C for 7 min. After, they were first rinsed in distilled H₂O for 3 min and then washed in 70% to 100% ethanol serial dilutions for 1 min. Finally, they were cleared in
25 xylene for 2 hours and mounted with mounting solution (Eukitt, Sigma Aldrich).

MRI acquisition

MRI was performed on a small animal-dedicated 7 T scanner (30/70 BioSpec; Bruker, Ettlingen, Germany). The animal protocol included high resolution T2 sequence. Analysis of the tumor volume was performed using MIPAV software (<https://mipav.cit.nih.gov>).

30 GBM-cortex organoids

For cerebral organoid generation WT iPSCs at 70-80% confluence were detached by Accutase solution incubation at 37°C for 10 minutes in order to obtain a single cell

suspension. Cells were centrifuged, counted and a total of 9000 cells were then plated into each well of an ultra-low-attachment 96-well plate (Corning), in medium containing DMEM/F12, 20% KnockOut™ Serum Replacement (KSR, Thermo Fisher Scientific), 2 mM Glutamine, 1% Pen/Strept, 1% non-essential amino acids, 50 nM β -mercaptoethanol (ThermoFisher Scientific) and 4 ng/ml bFGF. After seeding, plates were briefly centrifuged to allow single EB formation inside each well; ROCK inhibitor Y27632 (50 μ M) was included in the first 24 hours. EBs were maintained in 96-well plates for 6 days, then transferred by firmly pipetting (with a cut end of a P200 tip) medium in the well up and down to ultra-low-attachment 24-well plates (Corning), in neural induction medium containing DMEM/F12, 1X N-2 supplement, 1% non-essential amino acids, 2 mM Glutamine and 1 μ g/ml heparin (Sigma-Aldrich). On day 10, EBs were embedded in Matrigel (Matrigel growth factor reduced, Corning) together with CSC spheres (previously infected with either RFP only or SES+RFP) in the same droplet of Matrigel to allow fusion, and then gel at 37°C for 30-60 minutes. Embedded EBs-CSCs were subsequently cultured in neural maturation medium containing 50% DMEM/F12, 50% Neurobasal A, 0.5X N-2 supplement, 0.5X B-27 supplement without vitamin A, 2mM Glutamine, 2.5 ng/ml human insulin, 1% non-essential amino acids, and 25 nM β -mercaptoethanol. Droplets were cultured in stationary condition in 6 cm suspension dishes for 4 days, followed by transfer to an orbital shaker (Orbit™ LS Low Speed Orbital Shaker) rotating continuously at 60 rpm; here 0.5X B-27 supplement with vitamin A was substituted in the neural maturation medium.

Behavioral testing

Animals were housed at a constant 23°C in a 12 h light/dark cycle (lights off at 19:00), with food and water available *ad libitum*. We analyzed WT C57BL/6 mice, both males and females, at adult stage (ranging from 2 to 4 months of age) (all tests) infected 4 weeks before in their hippocampi with either GFP (mock) or SES. The sessions were recorded with the video tracking software Ethovision XT (Noldus).

Spontaneous alternation test. To test exploratory behavior and cognitive function related to spatial learning and memory the mice were inserted in a 4-arm maze and video recorded for 10 minutes to evaluate: total number of the entries in all arms, percentage of entries in each arm, and the consecution of the arm entries. This latter allows to identify pattern of behavior as (see also Fig 9e): Spontaneous Alternation Performance (SAP), a score index in which the visit of the 4 different arms without repetition is scored as 1 while at least one repetition in a string of 4 entrance is scored 0; Alternate Arm Return, a score index in which while at least one repetition in a string of 3 entrance is scored 1; and Sam Arm Return (SAR), a score index in which two consecutive entries in the same arm is as scored 1.

Radial maze test. The eight-arm radial maze consisted of eight identical arms extending radially from an octagonal platform. It was elevated 80 cm above the floor and surrounded by external cues. A cup containing food was placed at the end of each arm. The protocol was divided into distinct phases: Day 1 - Habituation at the apparatus for 10 min (without
 5 food at the end of the arms). Day 2 - Food deprivation until when the animals had arrived at the 80%–85% of their initial weight; during the experiment the mice had to maintain this weight. Day 3 - Training: put the food in half and at the end of each arm. Release the mouse in the center of the arena, it must eat two of the eight pellets placed at the end of the arms. Day 4-13 (experimental days 1-10 in the Fig 10f) - Test: The pellets are placed only at the
 10 end of the eight arms. The mouse is released in the center of the arena to calculate (i) the time taken to eat the eight pellets and (ii) the percentage of the incorrect choices (the mouse chooses an empty arm) on the total entries. The maze was cleaned with water and 70% ethanol before the next mouse was placed on the apparatus.

Morris water maze test. The mice were inserted in a circular pool with a platform that allows
 15 them to escape the water (max length of each trial 120"). The release site can be in a different quadrant of the pool (see protocol in Fig 9g) with the position of the platform that was the same for the first 3 days and the reverse for the last 2 days of the protocol. The time to complete each trial and the time spent in the platform zone and in the opposite quadrant was quantified.

20 **Neurons from iPSCs**

WT iPSCs were maintained in feeder-free conditions in mTeSR1 (Stem Cell Technologies) supplemented with Pen/Strept and seeded on human embryonic stem cell (HESC)-qualified Matrigel (Corning)-coated six-well plates; cells were fed daily and passaged in cell clumps weekly using Accutase solution (Sigma-Aldrich). At differentiation day -2, 90% confluent
 25 iPSC cultures were infected with the lentiviral vector TetO-Ngn2-T2A-Puro in mTeSR1 medium supplemented with Doxycycline (2 µg/ml, Sigma-Aldrich), overnight. The next day, the medium was replaced with fresh mTeSR1 medium supplemented with antibiotic selection (Puromycin 1 µg/ml, Sigma-Aldrich) and Doxycycline; Doxycycline was maintained for all the experiment. At day 0 medium was replaced with differentiation medium "mTeSR1
 30 + LSBX". Differentiation medium was replaced daily according to the following scheme: Day 0,1: mTeSR1 + LSBX; Day 2,3: mTeSR1 + LSBX + PSD; Day 4,5: 2/3 mTeSR1 + 1/3 N-2 medium + LSX + PSD; Day 6,7: 1/3 mTeSR1 + 2/3 N-2 medium + PSD. At day 8 cells were detached by Accutase solution incubation at 37°C for 20 minutes in order to obtain a single cell suspension. Cells were centrifuged, counted and seeded at a density of 55000 cells/cm²
 35 onto Poly-L-Lysine/Laminin/Fibronectin coated plates or coverslip in neuronal maturation

medium supplemented with ROCK inhibitor Y27632 (10 uM, Selleckchem) for the first 24 hours. Culture medium was replaced the next day to remove ROCK inhibitor, and then half of the medium was replaced with fresh Neuronal maturation medium twice a week.

LSBX: LDN193189 (Stemgent, 250 nm), SB431542 (Sigma-Aldrich, 10 µM) XAV939 (Sigma-Aldrich, 5 µM). PSD: PD0325901 (Sigma-Aldrich, 8 µM), SU5402 (Sigma-Aldrich, 10 µM), DAPT (Sigma-Aldrich, 10 µM). N-2 medium: DMEM/F12 with B-27 supplement (0.5X, ThermoFisher Scientific) and N-2 supplement (0.5X, ThermoFisher Scientific). Neuronal maturation medium: Neurobasal A (ThermoFisher Scientific) supplemented with 1X B-27 supplement, 2 mM Glutamine, 1% Pen/Strept, BDNF (Peprotech, 20 ng/ml), Ascorbic acid (Sigma-Aldrich, 100 nM), Laminin (1 µg/µl), DAPT (10 µM), dbcAMP (Selleckchem, 250 µM).

Primary murine neuronal culture

Primary cultures of mouse embryonic cortical neurons were prepared from E17.5 C57BL/6 Wild-Type mice. Briefly, after dissection, cortices were enzymatically digested with 0.025% trypsin (GIBCO) in Hank's balanced salt solution (HBSS) (Euroclone) for 20 min at 37°C. Successively, HBSS with trypsin was removed and the hippocampi were washed with plating medium (Neurobasal A medium supplemented with 1X B-27 supplement, 3.3 mM glucose, 2 mM glutamine and 1% penicillin/streptomycin) and mechanically dissociated with a P1000-pipette to obtain a homogeneous cell suspension. Cells were then plated on poly-L-lysine (PLL) (0.1 mg/ml) coated glass coverslips.

20

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the disclosed polypeptides, polynucleotides, vectors, cells, compositions, uses and methods of the invention will be apparent to the skilled person without departing from the scope and spirit of the invention. Although the invention has been disclosed in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the disclosed modes for carrying out the invention, which are obvious to the skilled person are intended to be within the scope of the following claims.

25

CLAIMS

1. An epigenetic silencer factor (ESF) comprising a transcription factor DNA-binding domain operably linked to at least one epigenetic effector domain, wherein the transcription factor is an oncogenic transcription factor or a cancer-associated transcription factor.
2. The ESF of claim 1, wherein the transcription factor is selected from the group consisting of SOX2, MYC, MYCN, TEAD1, TEAD2, TEAD3, TEAD4, FOXA1, FOXA2, ELK1, ELK3, ELK4, SRF, FOXM1, FOXC1, FOXC2, TWIST1, SALL4, ELF1, HIF1A, SOX9, SOX12, SOX18, ETS1, PAX3, PAX8, GLI1, GLI2, GLI3, ETV1, ETV2, ETV3, RUNX1, RUNX2, RUNX3, MAFB, TFAP2C and E2F1.
3. The ESF of claim 1 or 2, wherein the transcription factor is Sox2.
4. The ESF of any preceding claim, wherein the at least one epigenetic effector domain is selected from the group consisting of a KRAB domain, a DNMT3A domain, a DNMT3L domain, a ZIM3-KRAB (Z-KRAB) domain, a Chromo Shadow (CS) domain, a YAF2-RYBP (Y-R) domain, an Engrailed Repressor (En-R) domain, a MeCP2 domain, a GLI3RD domain and a MAD1RD domain.
5. The ESF of any preceding claim, wherein the ESF comprises a KRAB domain, a DNMT3A domain and a DNMT3L domain.
6. A polynucleotide comprising a nucleic acid sequence encoding the ESF of any preceding claim.
7. The polynucleotide of claim 6, wherein the polynucleotide further comprises a promoter operably linked to the nucleic acid sequence encoding the ESF, optionally wherein the promoter is a tissue-specific promoter or a constitutive promoter, optionally a cancer cell-specific promoter.
8. A vector comprising the polynucleotide of claim 6 or 7, optionally wherein the vector is a viral vector, optionally wherein the vector is a lentiviral vector or adeno-associated viral (AAV) vector.
9. The ESF, polynucleotide or vector of any preceding claim, wherein the ESF, polynucleotide or vector is comprised in a nanoparticle.
10. A cell comprising the ESF, polynucleotide or vector of any preceding claim.

11. A composition comprising the ESF, polynucleotide, vector or cell of any preceding claim.
12. The ESF, polynucleotide, vector, cell or composition of any preceding claim for use in therapy.
- 5 13. The ESF, polynucleotide, vector, cell or composition of any one of claims 1-11 for use in the treatment of cancer.
14. Use of the ESF, polynucleotide, vector, cell or composition of any one of claims 1-11 for decreasing transcription and/or expression of at least one target gene in a cell.
- 10 15. A method of decreasing transcription and/or expression of at least one target gene in a cell, the method comprising introducing the ESF, polynucleotide, vector or composition of any one of claims 1-9 or 11 into the cell.