SOX11 promotes epithelial/mesenchymal hybrid state and alters tropism of invasive breast cancer cells

Erik Oliemuller[1], Richard Newman[1], Siu Man Tsang[1], Shane Foo[2], Gareth Muirhead[1], Farzana Noor[1], Syed Haider[1], Iskander Aurrekoetxea-Rodríguez[3], Maria dM. Vivanco[3], Beatrice A. Howard[1*]

Affiliations:

[1]The Breast Cancer Now Toby Robins Research Centre, Division of Breast Cancer Research, The Institute of Cancer Research, London, UK

[2]Translational Immunotherapy Team, Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK

[3]CIC bioGUNE, Basque Research and Technology Alliance (BRTA), Derio, Spain

[*]Author for correspondence:
Beatrice A. Howard
e-mail: beatrice.howard@icr.ac.uk
phone: +44-203-437-7359

Keywords:
SOX11, DCIS, breast cancer, brain metastasis, MEX3A

Abstract

SOX11 is an embryonic mammary epithelial marker that is normally silenced prior to birth. High *SOX11* levels in breast tumours are significantly associated with distant metastasis and poor outcome in breast cancer patients. Here, we show that SOX11 confers distinct features to ER-negative DCIS.com breast cancer cells, leading to populations enriched with highly plastic hybrid epithelial/mesenchymal cells, which display invasive features and alterations in metastatic tropism when xenografted into mice. We found that SOX11+DCIS tumour cells metastasize to brain and bone at greater frequency and to lungs at lower frequency compared to cells with lower SOX11 levels. High levels of SOX11 leads to the expression of markers associated with mesenchymal state and embryonic cellular phenotypes. Our results suggest that SOX11 may be a potential biomarker for breast tumours with elevated risk of developing metastases and may require more aggressive therapies.

Introduction

SOX11 is an embryonic mammary factor that is not expressed in normal breast epithelial cells after birth (Wansbury et al., 2011). However, *SOX11* is expressed in many triple negative and HER2+ invasive breast cancers (Wansbury et al., 2011). *SOX11* expression in invasive breast cancer is associated with increased distant metastasis formation (Oliemuller et al., 2017). Inhibition of *SOX11* by siRNA suppressed growth and proliferation of ER- breast cancer cell lines, but had no significant effect on growth and proliferation of ER+ breast cancer cell lines (Shepherd et al., 2016). *SOX11* repression using siRNA reduced both cell migration and invasion in basal-like breast cancer (BLBC) cell lines, supporting a role for SOX11 in promoting breast cancer progression. In addition, *SOX11* inhibition in MDA-MB-468, a BLBC line, resulted in reduced expression of *FOXC1*, *CCNE1*, *KRT14*, *MIA and SFRP1*, all of which are PAM50 genes that are highly expressed in BLBC. This finding suggests SOX11 may modulate key features of basal-like cancer cells, including Keratin 14 expression, a marker expressed by basal cells and some luminal cells within the terminal ductal lobular unit, where many breast cancers arise (Gusterson & Eaves, 2018).

Our previous studies showed that when constitutively expressed by the human breast epithelial cell line MCF10A, SOX11 increased the number of basal/myoepithelial clones formed in standard colony formation assays and led to increased mammosphere formation, suggesting that SOX11 can modulate mammary progenitor features and stem cell activity when expressed in mature postnatal mammary epithelial cells (Oliemuller et al., 2017). Furthermore, we also showed that SOX11 promotes invasive transition of DCIS.com cells that were injected into the mammary duct to mimic formation of DCIS-like lesions within the mammary duct prior to progression to invasive state and tumour formation (Oliemuller et al., 2017). Expression of SOX11 in preinvasive breast lesions and potential sites of microinvasion in samples from DCIS cases supports a role for SOX11 in promoting *in situ* to invasive breast carcinoma transition. These findings led us to hypothesise that reactivation of embryonic mammary developmental programmes mediated by SOX11 in postnatal breast epithelial cells would promote invasive progression

of pre-invasive DCIS breast lesions and acquisition of features associated with poor patient outcome, including formation of distant metastasis.

Therefore, to further explore the role of SOX11 in regulating cancer stem cell states and to determine whether SOX11 promotes metastatic dissemination of invasive breast cancer cells, we have created an inducible model to study the roles of SOX11 in progenitor/stem cell regulation and breast cancer progression *in vitro* and *in vivo*. In the model described here, SOX11 is expressed at higher levels in DCIS.com cells driven from the EF1A promoter after induction with doxycycline (DOX), when compared to another model we have previously used to study DCIS progression, in which low levels of SOX11 expression is constitutively driven by the CMV promoter (Oliemuller et al., 2017). In the inducible model presented here, SOX11 expression confers epithelial/mesenchymal hybrid state, features typical of embryonic mammary cell phenotypes, and influences organ-specific metastatic tropism to breast cancer cells. Finally, we also identify several novel SOX11 targets, many of which are highly correlated with *SOX11* expression in primary breast cancers and breast cancer metastases.

Results

Inducible expression of SOX11 leads to changes in stem cell profiles of DCIS.com cells

To investigate the role of SOX11 in breast cancer progression, we used the pINDUCER21 system to stably transduce DCIS.com cells, an invasive cell line from the MCF10A breast cancer progression series, so that SOX11 was expressed only when induced with Doxycycline (DOX) (referred to as iSOX11 cells) (**Figure 1A-B**). The results show a significantly higher, sustained expression of SOX11 levels compared with the previous constitutive model we have used to study DCIS progression which lost SOX11 expression over time (**Figure 1—figure supplement 1**) (Oliemuller et al., 2017). As expected, SOX11 localised mostly to the nuclei in iSOX11 cells, similar to that observed in SOX11+ DCIS case samples (**Figure 1A-C and Figure 1—figure supplement 1**). SOX11 is also detected in the cytoplasm of iSOX11 cells using Western Blotting (**Figure 1A**), a location that was not observed in the DCIS-SOX11 (data not shown), showing that some differences exist when SOX11 is expressed at different levels in the two models.

To study the role of SOX11 in regulating stem cell state, we used the inducible iSOX11 model and assessed cancer stem cell profiles. Cancer stem cells (CSCs) are subpopulations of cancer cells sharing similar characteristics as normal stem or progenitor cells such as self-renewal ability and multi-lineage differentiation to drive tumour growth and heterogeneity. ALDH1 and CD24 are widely used CSC markers in breast cancer (Liu et al., 2014). A higher proportion of CD24+ and ALDH+ cells are detected iSOX11 cells compared to control iEV cells (DCIS.com cells stably transduced with pINDUCER21 empty vector and induced with DOX) by flow cytometry (**Figure 1D-E and Figure 1—figure supplements 1-2)**. Moreover, CD24 expression levels were significantly increased in a DOX dose-dependent manner in iSOX11 cells (**Figure 1—figure supplement 1**). CD24 is predominantly located in the membrane and CD24 was detected in the cytoplasmic fraction of both iEV and iSOX11 cells, and also detected in the nuclear fraction of iSOX11 cells, which has been previously reported in breast cancer (**Figure 1F**) (Duex et al., 2017).

When cells were grown on collagen, we detected loss of membranous E-Cadherin localisation and acquisition of N-Cadherin expression in iSOX11 cells and formation of bud-like structures (**Figure 1G-H**). Nuclear expression of N-Cadherin has been reported previously, primarily in poorly differentiated breast tumours (Rezaei et al., 2012). We also detected a significant increase in iSOX11 cells expressing Vimentin by western blotting; increased Vimentin was detected in both the cytoplasmic and nuclear fraction. Vimentin was not observed in the nucleus of cells grown in the same conditions when stained by immunofluorescence, suggesting that the nuclear fraction contains proteins attached to the external part of the nuclear envelope since both Lamin B and Vimentin are known to associate with the nuclear envelope (**Figure 1G-H and Figure 1—figure supplement 3**) (Georgatos & Blobel, 1987a, 1987b). We assessed the expression of markers associated with mammary epithelial and mesenchymal states in iEV and iSOX11 as well as non-induced (denoted ni) niEV and niSOX11 DCIS cells. K5, K8, and VIM levels are elevated in iSOX11 cells compared to iEV cells (**Figure 1—figure supplement 3**)**.** These results are consistent with a role for SOX11 in promoting a mesenchymal state and influencing mammary epithelial phenotypes when expressed in epithelial cells.

To further analyse the effects of SOX11 on mammary epithelial phenotypes, we assessed the expression of markers associated with the two major mammary lineages in DCIS cells grown as spheroids that formed from iSOX11 and to iEV, niEV, and niSOX11 control cells. We detected significant increases in expression of both luminal (K8) and basal (K14, SMA) lineage markers in SOX11-expressing spheroids as well as a significant increase in cells co-expressing markers of both lineages, which is suggestive of embryonic mammary phenotypes (**Figure 1I and Figure 1—figure supplements 4-5**).

DCIS cells expressing SOX11 grow more slowly and form more invasive spheroids

Reduced cell growth was detected in iSOX11 cells compared to control iEV cells grown in both 2D or as spheroids (**Figure 2A-B**). Control iEV cells were more clonogenic than iSOX11 cells whether plated in colony-forming assays or when grown from single cells (**Figure 2—figure supplement 1**).

Although iEV cells have greater colony-forming capacity than iSOX11 when large colonies (greater than 50 microns were counted), a greater number of small colonies (less than 50 microns) formed (**Figure 2—figure supplement 1**)**.** We found DCIS-LacZ control cells also exhibited both greater colony and sphere-forming capacity than DCIS-SOX11 cells (**Figure 2—figure supplement 1**). As with iSOX11 cells, a greater number of small colonies form in colony-formation assays with DCIS-SOX11 cells. Spheroids formed from PKH-labelled DCIS-SOX11 cells are smaller and retain PKH dye more than labelled DCIS-LacZ cells (**Figure 2—figure supplement 1**). These results show that SOX11 expression leads to a reduction of larger spheroids formed from DCIS.com cells and increased formation of smaller spheroids that retain label. Multiple attempts to transduce primary human breast epithelial cells to express SOX11 constitutively with a CMV-driven vector were unsuccessful, but are consistent with a possible role for SOX11 in regulating a quiescent state (data not shown).

SOX11 expression is enriched in ER- and HER2+ invasive breast cancers (**Figure 2—figure supplement 2**). To explore whether sustained reactivation of SOX11 promotes tumour progression, we assessed features of iSOX11 spheroids. Spheroids formed from iSOX11 DCIS cells were smaller and formed a higher number of peripheral microspheres than iEV cells (**Figure 2C**). When overlaid with Collagen I, spheroids formed from iSOX11 cells are more invasive compared to iEV, niEV, and niSOX11 control cells (**Figure 2D**). Tumours that formed after mammary fat pad xenografts of iSOX11 cells were smaller than iEV tumours when mice were fed chow with moderate levels (200 or 625 ppm) of DOX (**Figure 2E-F**). Despite their smaller volume, tumours originated from cells with high SOX11 levels displayed greater bioluminescence than control tumours, suggesting iSOX11 tumours contained more viable cells (**Figure 2G**). IHC staining of mammary tumours was performed to detect Cleaved Caspase 3 (CC3). Larger necrotic and CC3+ (apoptotic) areas were observed in the EV tumour tumours and niSOX11 tumours when compared to the iSOX11 tumours, which showed little central necrosis and fewer CC3+ cells (**Figure 2H**). We also observed that extremely small and less luminescent tumours were formed in mice xenografted with iSOX11 cells, compared to control cells that were fed higher

levels (1250 or 2000 ppm) of DOX-chow (**Figure 2—figure supplement 3**). Higher levels of CC3 were detected in these iEV tumours, which displayed central necrosis, whilst almost no CC3 or necrosis was detected in iSOX11 tumours (**Figure 2—figure supplement 3**). Tumours formed from iSOX11 DCIS cells grew out quickly when (1250 or 2000 ppm) DOX-chow was replaced with normal chow, suggesting that high levels of SOX11 could keep tumours in a non-proliferative state and that upon DOX withdrawal, proliferation resumed and the ostensible quiescent state is a reversible condition (**Figure 2—figure supplement 3**).

When injected directly into the mammary duct, iSOX11 DCIS cells formed slightly larger tumours, with similar bioluminescence levels as iEV tumours, which indicates the microenvironment highly influences behaviour of SOX11+ tumour cells (**Figure 2I-K**). A greater proportion of mammary tumours formed from iSOX11 cells expressed moderate to high levels of ALDH1 (6/6) compared to control tumours (1/8). In addition, CD24+ cells were observed at greater frequency in iSOX11 tumours compared to control tumours (**Figure 2—figure supplement 4**). Of note, we observed nuclear CD24 staining in iSOX11 tumour cells, but not in iEV, niEV, or niSOX11 tumours, in line with the observed *in vitro* result (**Figure 2—figure supplement 4**). It has been suggested that cells designated CD24- using flow cytometry maybe expressing CD24 in the nucleus and this promotes aggressive tumour properties, since it has been shown that nuclear CD24 is able to drive tumour growth (Duex et al., 2017).

SOX11 expression promotes expression of developmental pathways frequently activated in cancer

To identify potential SOX11 targets that could confer features of embryonic, epithelial/mesenchymal hybrid state, and metastasis-promoting features to mature breast cells, RNA from iEV, niEV, or niSOX11 control and iSOX11 cells grown in 2D and in 3D (from spheroids at two time-points: 2 days and 5 days after spheroid formation) were sequenced (**Figure 3A, Supplementary File 1 and Figure 3—figure supplement 1**). Uninduced niSOX11 control cells express very low levels of *SOX11,* and suggest the vector is slightly leaky. iSOX11 DCIS cells cultured in 2D showed enrichment of genes

regulating actin filament sequestration, phospholipid catabolism, ERBB signaling, chemotaxis, and epithelial differentiation (**Figure 3B and Supplementary File 2**)**.** iSOX11 DCIS cells grown in 3D showed enrichment of genes regulating ECM disassembly, collagen biosynthesis, glycosaminoglycan metabolism, and platelet degranulation (**Figure 3—figure supplement 1 and Supplementary File 3**). Platelet activation is one of the first steps of tissue repair as part of the wound healing process and Sox4 and Sox11 have recently been shown to reactivate an embryonic epidermal programme during wound repair in mice (Miao et al., 2019). We found substantial overlap of the embryonic wound signature that was shown to be directly regulated by Sox11 and Sox4 (Miao et al., 2019) with iSOX11 spheroids (**Table 1**). In particular, we detected upregulation of embryonic wound signature components with links to actin polymerization and cell adhesion, and one known regulator of embryonic stem cell pluripotency, *RCOR2,* which can function with other transcription factors to induce pluripotent stem cells (**Figure 3—figure supplement 1**) (P. Yang et al., 2011).

RNA sequencing analysis detected *CDH2* (encoding N-Cadherin) expressed at significantly higher levels in iSOX11 cells grown in 2D and as spheroids, whilst other EMT markers, including *CDH1* (encoding E-cadherin) and *VIM* were not significantly changed compared to control cells (**Supplementary File 1**). Other notable downstream targets of SOX11 included *MEX3A*, which encodes an RNA binding protein, that marks slowly proliferating multipotent stem cells in mouse intestine (Barriga et al., 2017) and totipotent cells *in C elegans* (Pereira et al., 2013); *MMP11, ST6GALNAC5,* and *TUBB3,* which are highly expressed in breast cancers that metastasize to brain (Bos et al., 2009; Kim et al., 2015; Lee et al., 2016). We confirmed that a number of putative SOX11 targets of interest (**Figure 3C and Figure 3—figure supplement 1, Supplementary File 2**), including *MEX3A* and *TUBB3* were upregulated in iSOX11 cells when measured by qPCR (**Figure 3D**). We also detected upregulation of MEX3A and TUBB3 protein in iSOX11 cells (**Figure 3E-F**). We found many SOX11+ breast cancer cell lines express high levels of *MEX3A* or *TUBB3* compared to DCIS.com cell line (**Figure 3G**)**.** *MEX3A* and *TUBB3* levels are correlated with *SOX11*

expression in breast cancers in the TCGA dataset (**Figure 3H**) and in breast cancer cell lines in the Broad dataset (Ghandi et al., 2019) (**Figure 3—figure supplement 1**). Notably, with greater increases of *SOX11* levels, a higher percentage of samples with increased *MEX3A* are observed. These findings support a role for SOX11 in mediating developmental signals during breast cancer progression.

DCIS cells expressing SOX11 show alterations in metastatic tropism

To explore whether sustained reactivation of SOX11 promotes tumour progression, we injected luciferase-tagged iEV and iSOX11 cells into the mammary fat pad. Four weeks after orthotopic xenografting with iSOX11 cells, brain micrometastases were detected in two out of six mice by IVIS imaging, whilst none were observed in six mice xenografted with control cells (**Figure 4A**). Liver and lung micrometastases were detected in both cohorts when assessed by IVIS imaging.

After injection into the tail vein, IVIS imaging detected tumour cells in lungs in seven out of the eight mice, bone in three out of eight mice, and in brain in one of eight mice xenografted with iSOX11 cells, whilst no bone or brain metastases were observed in eight mice xenografted with control cells (**Figure 4B**). It was noted that the frequencies of mice with lung metastasis were similar, but a significant reduction of iSOX11 DCIS cells accumulated in the lungs was observed when compared to mice engrafted with iEV cells when quantified by IVIS (**Figure 4C-D**). Macroscopic examination of metastatic lesions confirmed the reduction in tumour burden (data not shown).

SOX11 expression in primary breast cancer is associated with increased metastasis formation at distant sites (**Figure 4E**). *SOX11* is highly expressed in brain metastases and is also detected in bone metastasis from breast cancer patients (**Figure 4—figure supplement 1**)**.** *SOX11* amplification and overexpression has been detected in breast cancer brain metastasis in patients with ER-, ER+, and *BRCA1-/-* tumours in a recent study that comprehensively profiled a small number of cases (**Figure 4—figure supplement 1**) (Saunus et al., 2015). In another dataset in which samples from 21 breast cancer brain metastasis (BCBM) patients were transcriptionally profiled by RNA sequencing, elevated levels of *SOX11* are detected in the

brain metastasis in one third of cases compared to the primary tumour (**Figure 4—figure supplement 2**) (Vareslija et al., 2018).

SOX11+ DCIS.com cells from brain metastasis display a colonisation and growth advantage after intracranial xenografting

iSOX11 cells were isolated from mouse brain after orthotopic injections and expanded (designated iSOX11Br). Induction of SOX11 with DOX led to increased expression of SOX11, MEX3A and TUBB3 compared to untreated cells (**Figure 5A**). Other cells isolated and expanded from bone and lung metastasis after orthotopic injections of iSOX11 did not express SOX11 after induction with DOX.

After xenografting iSOX11Br cells into the tail vein, mice that had been fed normal chow had a greater metastatic burden in the lungs compared to mice fed DOX-containing chow to induce SOX11 expression, similar to results observed using the parental iSOX11 DCIS cell line, (**Figure 5B-D**). After intracranial injections of iSOX11Br cells, higher levels of bioluminescence were detected in the brains of mice fed DOX chow (**Figure 5E**)**.** Induction of SOX11 expression led to larger tumour burden in the brain and reduced survival (**Figure 5F**). These results indicate that iSOX11Br cells have a colonisation and growth advantage in the brain compared to that niSOX11Br cells lacking SOX11 expression**.**

SOX11 regulates proliferative state of ER- breast cancer cells

Next, we examined the effect of reducing SOX11 levels in CAL-148 ER- breast cancer cell line that expresses very high SOX11 levels. Using DOX-inducible lentiviral vectors containing shRNAs to *SOX11* and control non-targeting shRNAs, reductions in *SOX11* levels were obtained, as well as decreases of *TUBB3, MEX3A*, *GPC2*, *MPK4*, *OLFM2, ST6GALNAC5, and NCAD* levels and an increase in *SERPINA3* in CAL-148 cells after *SOX11* knockdown when compared to control (**Figure 6A**). Reduced levels of SOX11, MEX3A and TUBB3 protein were detected after *SOX11* knockdown (**Figure 6B**). Cell viability assays detected greater cell numbers after *SOX11* knockdown in CAL-148 cells, compared to control cells (**Figure 6C**). Colony formation assays detected an increase in clonogenetic potential of CAL-148

cells with reduced *SOX11* levels (**Figure 6D**). These results suggest SOX11 could regulate proliferative state of stem cells in ER- breast cancer cells.

TUBB3, an established SOX11 target, regulates proliferation and invasion of ER- breast cancer cells

As with *SOX11*, both distant metastasis-free survival and overall survival of breast cancer patients are reduced when high levels of *TUBB3,* an established SOX11 target in neural cells*,* are expressed in primary tumours (**Figure 7—figure supplement 1**). We next examined the effects of TUBB3 on triple negative breast cancer (TNBC) growth and invasion. Using siRNA-mediated knockdown, we found that reducing TUBB3 levels in BT-20 cells resulted in reduced growth in both 2D culture and spheroid culture (**Figure 7A-C**). More cells were arrested in G2/M phase of the cell cycle when TUBB3 levels were reduced (**Figure 7D and Figure 7—figure supplement 2**), consistent with a known role of TUBB3 in regulating cell cycle progression of tumour cells. As a result of cell cycle arrest, an increase of multinucleated cells in the >G2/M phase and a higher proportion of dead cells in sub-G1 phase were detected. Spheroid invasion assay s detected less invasion when TUBB3 levels were reduced (**Figure 7E-F**). These results suggest SOX11 can regulate proliferation and invasive growth through TUBB3 in ER- breast cancer cells.

MEX3A, a novel potential SOX11 downstream effector, regulates cell growth and E/M state of ER- breast cancer cells.

Both distant metastasis-free survival and overall survival of breast cancer patients are reduced when high levels of *MEX3A* are expressed in primary tumours, as would be expected for a SOX11 target (**Figure 7—figure supplement 1**). Due to MEX3A's established links with regulation of both EMT and proliferation of intestinal and various types of cancer cells , we knocked down *MEX3A* in ER- BT-20, CAL-148 and HCC1187 breast cancer cells to determine if MEX3A regulates similar processes. In HCC1187 and BT-20 cells, a slight but significant reduction of invasive growth was detected when MEX3A levels were reduced with some, but not all siRNAs (**Figure 8A-B and Figure 8—figure supplement 1**). An increase in cell numbers were

observed after knockdown of *MEX3A* in both CAL-148 and HCC1187 cells grown in 2D or 3D (**Figure 8C-D**). Cell cycle analysis detected a reduction in S phase after *MEX3A* knockdown in both cell lines (**Figure 8E, Figure 8—figure supplement 1**). Several candidate cell cycle regulators from a consensus stem cell quiescence signature (Cheung & Rando, 2013), were downregulated in iSOX11 cell signatures, including *RRM2* and *SURVIVIN* (**Figure 8—figure supplement 1 and Supplementary File 3**). Western blotting found that both RRM2 and SURVIVIN were upregulated when MEX3A levels were reduced in CAL-148 and BT-20 cells (**Figure 8F and Figure 8—figure supplement 1**). High expression levels of *MEX3A* co-occur with increased levels of *E2F3*, *CCNE1* and *CDKN2A,* and decreases in *RB1* levels in the TCGA dataset (**Figure 8—figure supplement 1**); *MEX3A* levels show strong correlation with the levels of *E2F3* and *CCNE1* (**Supplementary File 4**). The addition of EGF to CAL-148 cells growing in either serum-free or low-serum media led to reduction of MEX3A levels (**Figure 8—figure supplement 2**). CAL-148 cells normally form aggregates when grown in 2D culture conditions and after *MEX3A* knockdown, CAL-148 cells displayed a reduced ability to form aggregates and an acquired ability to adhere to plastic (**Figure 8G**). After MEX3A knockdown, CAL-148 cells displayed increased expression of E-CADHERIN and EPCAM (**Figure 8H**)**.** BT-20 and HCC1187 cells lack expression of E-CADHERIN but showed an increase in EPCAM levels after MEX3A knockdown (**Figure 8—figure supplement 2**). Together, these results are consistent with roles for MEX3A in regulation of cell growth and EMT in SOX11+ ER- breast cancer cells.

Discussion

We previously showed that SOX11 promotes invasive transition of DCIS cells using both *in vitro* and *in vivo* models where SOX11 is expressed in DCIS.com cells under the control of a constitutive CMV-driven promoter (Oliemuller et al., 2017). Here, we investigated the role of SOX11 in breast cancer progression after tumour formation using a doxycycline-inducible EF1A promoter to express SOX11 in DCIS.com cells at a level comparable to that observed in clinical DCIS and breast cancer samples. DCIS cells induced to express SOX11 prior to spheroid formation form smaller spheroids, which display more invasion compared to control spheroids, recapitulating the phenotypes we observed with constitutive expression of SOX11 (Oliemuller et al., 2017). In addition, we observed unique phenotypes using these experimental conditions, including substantial morphological changes (cell detachment and formation of multiple satellite spheroids indicative of a highly invasive phenotype without hydrogel), suggesting breast lesions expressing very high levels of SOX11 possess an inherent high potential to form metastatic lesions.

Stem cells and progenitors may serve as a cell of origin for breast cancers. Many SOX (SRY-related HMG-box) transcription factors are expressed in the postnatal breast and some have been shown to control normal and/or cancer stem cells (Domenici et al., 2019; Kogata et al., 2018; Mehta, Khanna, & Gatza, 2019). SOX11 is a unique transcription factor, since it does not appear to be expressed in normal mammary epithelial cells after birth in either mouse or human and therefore is well-poised to reactivate developmental pathways when expressed in breast cancers (Oliemuller et al., 2017; Tsang, Oliemuller, & Howard, 2020; Wansbury et al., 2011; Zvelebil et al., 2013).

CD24 and ALDH1 are widely used cancer stem cell (CSC) markers in breast cancer (Liu et al., 2014). It is now widely accepted that stem cell states fluctuate and are not fixed (Liu et al., 2014) and cells transition between the two states (Liu et al., 2014; O'Brien-Ball & Biddle, 2017). Our results suggest a dose-dependent effect of SOX11 on CD24 levels. It is possible that CMV driven versus EF1A-driven SOX11 has distinct cell context effects. Lower levels of CMV-driven SOX11 expression results in expansion of the ALDH+/CD24- CSC population in DCIS.com cells (Oliemuller et al., 2017).

Meanwhile, EF1A-driven inducible SOX11 expression, presented here, leads to expansion of the CD24+ CSC population and higher levels of SOX11 are associated with an increased size of the ALDH+/CD24+ CSC population. Triple negative breast cancer (TNBC) cells exhibit robust expression of *CD24,* suggesting that the inducible model described here mimics what is observed in human TNBC and may present a potential therapeutic opportunity for some SOX11+ breast cancers (Barkal et al., 2019).

Sox11 is highly expressed in prenatal mammary epithelial cells from the time the mammary organ initially forms (Zvelebil et al., 2013). scRNA-sequencing also detects *Sox11* expression by the majority of Lgr5+ embryonic mammary epithelial cells isolated at E14.5, a stage when embryonic mammary epithelial cells are multipotent and most cells express markers associated with the two major mammary lineages (the basal/myoepithelial and luminal lineages) (Lilja et al., 2018; Wuidart et al., 2018). Prenatal human breast tissues also express markers associated with both basal and luminal mammary lineages (Jolicoeur, Gaboury, & Oligny, 2003). Co-expression of basal and luminal markers are observed in iSOX11 cells and spheroids formed from them. Our findings are consistent with SOX11 marking populations enriched with embryonic mammary cell phenotypes (Bland & Howard, 2018; Wuidart et al., 2018).

Epithelial cells with mesenchymal features have been detected at high frequency during organogenesis in the developing mouse embryo using single cell RNA-seq analysis. *Sox11* positively regulates expression of mesenchymal markers, including *N-cadherin (Cdh2)* and *Fibronectin1 (Fn1)* in several developing organs, including intestine, liver, lung, and skin (Dong et al., 2018; Halbleib & Nelson, 2006). A significant association of *SOX11* and *N-CADHERIN* (*CDH2)* expression is also observed in genomic analyses across six cancer types, suggesting that mesenchymal pathways may be activated by SOX11 in both normal and cancer cells (Vervoort, Lourenco, van Boxtel, & Coffer, 2013). SOX11 has previously been implicated in participating in the regulation of epithelial-mesenchymal transition (EMT) (Venkov et al., 2011). EMT, stemness and plasticity are intertwined (Nieto, Huang, Jackson, & Thiery, 2016; Wahl & Spike, 2017) and SOX11 is well poised to function as a

key regulator of epithelial/mesenchymal cell states during development and in cancers.

The regulation of N-CADHERIN (CDH2) by SOX11 is significant since CDH2 promotes motility in human breast cancer cells regardless of their E-CADHERIN expression (Nieman, Prudoff, Johnson, & Wheelock, 1999). CDH2 is a predictive biomarker for distant metastasis in early-stage breast cancer (Aleskandarany et al., 2015), that is commonly detected in breast cancer cells and provides a mechanism for transendothelial migration (Qi, Chen, Wang, & Siu, 2005).

Micrometastasis to brain is detected in ~30% of mice within 1 month of xenografting DCIS cells induced to express high levels of SOX11. *SOX11* is amplified and overexpressed in ~30% of brain metastases in a recent study of a small cohort of breast cancer patients that were profiled using an integrated genomic and transcriptomic analysis of fresh frozen tumour samples (Saunus, McCart Reed, Lim, & Lakhani, 2017; Saunus et al., 2015) and highly expressed by ~30% of BCBM in another independent study (Vareslija et al., 2018). Our finding that iSOX11 mammary tumours spontaneously metastasize to the brain is clinically relevant.

We have identified a number of genes that are regulated by SOX11 in our mouse models of breast cancer that are also associated with SOX11 expression in both primary breast cancer and breast cancer brain metastasis. A number of these, such as *TUBB3*, encode targetable molecules that are being used in novel pharmacogenetic approaches in combination with other factors (Karki et al., 2013). Others, such as MEX3A, may provide novel biomarkers or therapeutic targets (Bufalieri et al., 2020; Wang et al., 2020; D. Yang, Jiao, Li, & Fang, 2020). MEX3A controls the polarity and stemness and affects the cell cycle of intestinal epithelial cells through the downregulation of the mRNA encoding the CDX2 transcription factor (Pereira et al., 2013) and, in addition, exhibits a transforming activity when overexpressed in gastric epithelial cells (Jiang et al., 2012). The intestine has a reservoir of quiescent stem cells that are resistant to chemotherapy that are marked by Mex3a and are multipotent so have the capacity to produce any kind of intestinal cell and contribute to tumour heterogeneity (Barriga et al., 2017). MEX3A may mark a

rare mammary stem cell in the human breast that could escape traditional chemotherapy treatments, but this remains to be demonstrated.

The analysis of co-occurrence of *SOX11* with *MEX3A* and *TUBB3* in breast cancers profiled in TCGA suggests that *MEX3A* is expressed predominantly in samples with higher levels of *SOX11*. The differences in cell growth observed both *in vitro* and *in vivo* with increasing levels of SOX11 suggest a dose-dependent regulation of SOX11 downstream targets and/or a possible post-translational modification. Also, the discrepancies between the tumour size observed after xenografting iSOX11 cells via the mammary fat pad model and the MIND model, as well as the finding that EGF can decrease MEX3A levels, suggest an influence of microenvironmental factors on SOX11 tumour cell behaviour. This is in agreement with differential expression of *TUBB3* and *MEX3A*, depending if DCIS-SOX11 cells were xenografted in the mammary fat pad or via MIND (Oliemuller 2017). Together, this data suggests it may be necessary to classify SOX11+ tumours, depending on SOX11 expression level, as well as the expression of its effectors in order to stratify breast cancer patients. The subcellular localisation of SOX11 can be informative for cancer classification since high SOX11 mRNA levels and detection of the nuclear protein are reliable markers of mantle cell lymphoma (Mozos et al., 2009). During neurogenesis, SOX11 is detected in both the nucleus and cytoplasm (Balta et al., 2018). Phosphorylation of specific serine residues can modulate SOX11 subcellular localization and prevent its nuclear localization (Balta et al., 2018). We have not determined whether SOX11 localises to the cytoplasm, as well as the nucleus, in breast cancer case samples, and if SOX11 localisation could be useful for classification of breast cancers.

SOX11+ breast cancer cells express markers indicative of phenotypic plasticity and have a high tendency to undergo metastasis (O'Brien-Ball & Biddle, 2017). Together, these data suggest that patients whose DCIS and primary breast cancers express high levels of *SOX11* are among a high-risk metastasis subgroup that should be considered for aggressive therapies in neo-adjuvant settings.

Methods

Cell culture

DCIS.com-Luc cells were generated by transducing cells with lentiviral expression particles for firefly luciferase 2 (LVP325; Amsbio, Abingdon, UK). DCIS.com-Luc-mCherry were created by lentiviral transduction of the mCherry sequence in the DCIS.com-Luc cells. Supplementary material (**Supplementary File 6**) provides details and sources of cell lines and media used. All cell lines were tested and were mycoplasma-free. The DCIS.com have been extensively profiled in (Maguire et al., 2016) and we confirmed PIK3CA mutation status by PCR and Western blotting. Cell lines were authenticated by STR profiling (Eurofins).

Expression vectors

The SOX11 coding sequence (GENEID: 6664) from clone HsCD00295480 19 in the pENTR223.1 plasmid (DNASU) 20 was subcloned into pInducer21 (ORF-EG) plasmid (gift from Stephen Elledge & Thomas Westbrook (Addgene plasmid # 46948; http://n2t.net/addgene: 46948; RRID:Addgene_46948) (Meerbrey et al., 2011). pInducer13-shRNA and pInducer13-NS shRNA were made by subcloning the specific shRNA174 for SOX11 and the NS shRNA from pGIPZ plasmids from Horizon into pInducer13.

RNA isolation

RNA from cells grown in 2D for 48h in presence or absence of 1µM doxycycline (DOX) and from spheroids treated for 2 or 5 days (DOX 2days or DOX 5days) or not treated with DOX (DOX 0Days) (n = 3 for each time point) was isolated with an RNAeasyPlus Micro kit (74034; Qiagen, Manchester, UK) and DNase treatment. RNAClean and concentrator-5 (Zymo Research, Irvine, CA, USA) were used. RNA concentration and purity were determined with a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) and a nanodrop spectrophotometer. RNA integrity number was measured with a bioanalyzer and an Agilent RNA Pico kit (Agilent Technologies, Cheshire, UK).

cDNA synthesis and qPCR

One microgram of each RNA sample was reverse transcribed with QuantiTect Reverse Transcription kit (Qiagen, Manchester, UK) in a final volume of 20µl. cDNA was diluted ten times for subsequent quantitative polymerase chain reaction (qPCR) analysis, as described previously (Oliemuller et al., 2017), with the probes and methods listed in supplementary material (**Supplementary File 7**).

Western blotting

Western blotting was performed as previously described (Zvelebil et al., 2013). Details of the antibodies used are provided in supplementary material (**Supplementary File 8**).

Immunohistochemistry (IHC)

IHC was performed on formalin-fixed paraffin embedded samples. Samples were stained with antibodies as described previously (Oliemuller et al., 2017).

Immunofluorescence

Antibodies and staining protocols are detailed in supplementary material (**Supplementary File 9**). EVOS fluorescence microscope was used for imaging. Confocal images were captured with a Leica Microsystems (Cambridge, UK) TCS- SP2 confocal microscope.

Spheroid formation

Five thousand cells, untreated or treated with 1µM doxycycline (DOX) for 48h, were plated per well in 96-well ultra-low-attachment plates (Corning 7007, Corning, NY, USA) in media containing DOX or not. After 24h, when the spheroids were formed, new media containing 1µM doxycycline was added for 48h to spheres formed in absence of DOX for 2 days (DOX 2days) or spheres that were formed in presence of DOX to a total of 5 days (DOX 5days). Media without doxycycline was added to the control spheroids (No DOX). Images were obtained with a Celigo cytometer (Nexcelom, Manchester, UK).

588

589 **Colony formation assays**

590 DCIS.com cells were plated at 250 per well in six-well (Falcon F3046,
591 Corning, NY, USA) plates. After 7 days, plates were stained with 0.2% crystal
592 violet dissolved in 20% methanol in PBS. Area was measured, and the
593 percentage relative to number of cells plated was calculated.

594 For single cell colony assays and single cell mammosphere assays,
595 cells were FACS sorted and a single living cell plated per well in a 96well or
596 ultra-low attachment plate respectively. After 14 days, plates were analysed in
597 a Celigo cytometer (Nexcelom, Manchester, UK).

598 For mammosphere assays, 5000 DCIS cells/ml were plated in low-
599 attachment six-well plates (Corning 3471) and incubated in medium
600 supplemented with 2% NeuroCult SM1 without vitamin A (StemCell
601 Technologies) and 0.65% methylcellulose (R&D Systems, Abingdon, UK).
602 After 14 days, wells were scanned with a Celigo cytometer. Mammosphere-
603 forming efficiency was calculated by dividing the number of mammospheres
604 by the number of cells plated per well.

605

606 **Cell viability assays**

607 Three thousand cells per well were plated in 96-well plates (655098; Greiner
608 Bio-one, Stonehouse, UK) or in ultra-low attachment plates for 24 h before
609 starting the experiments. CellTiter-Glo (Promega, Southampton, UK) was
610 used according to the manufacturer's protocol. Luminescence was measured
611 with a Victor X5 58 plate reader (Perkin-Elmer, Seer Green, UK). In DCIS.com
612 cells, CellTiter-Glo assays were performed at the time of adding the
613 doxycycline and after 1, 2 and 3 days. When cells were transfected with
614 siRNAs, the transfection was done overnight and next day (day 1), full
615 supplemented media was added. CellTiter-Glo assays were performed on day
616 1 and 5 to allow time for the siRNAs to knock down the genes of interest.
617 CAL-148 cells stably transduced with shRNAs were measured at 0, 3 and 7
618 days, since this cell line has a long doubling time.

619

620 **Invasion assays**

3 days after plating 5000 DCIS.com cells, spheroids were embedded in collagen I (354249; Corning) at 2.2 mg/ml diluted in medium. Complete medium was added on top after 1 h. Images were acquired at this time and after 48 h with a Celigo cytometer. The total area of matrix invaded by cells was calculated with ImageJ after marking of the area manually. For BT-20 and HCC1187 cells, the spheres used in the invasion assays were formed at the same time that they were transfected with the *TUBB3* or *MEX3A* specific siRNAs in Opti-MEM media supplemented with 10% FBS. After 24h, this media was changed for the normal media of this cells. The cells were invaded in collagen 72h after plating.

Tranfections and siRNA

BT-20, HCC1187 and CAL-148 cells were transfected with 50 pmol of each *TUBB3* or *MEX3A* siRNA (siGENOME SMARTpool and four individual siRNAs), control nontargeting siRNA (Thermo Scientific, Waltham, MA, USA) by using Lipofectamine RNAiMAX (Invitrogen, Life Technologies Corporation, Carlsbad, CA, USA) in Opti-MEM media supplemented with 10% FBS (Gibco, Life Technologies Corporation, Carlsbad, CA, USA) media according to the manufacturer's instructions for 16h in a six-well or 96-well plate, depending on the experiment, and then incubated with complete media. For sphere transfections, 5000 cells per well were incubated with 2.5 pmol of each specific siRNA in 150ul of normal media per well in an ultra-low attachment plate for 16h. The next day, media was replaced with fresh media.

Flow cytometry analyses and fluorescence-activated cell sorting (FACS)

Aldehyde dehydrogenase (ALDH) activity was measured with the Aldefluor assay (StemCell Technologies, Cambridge, UK) as described before (Oliemuller et al., 2017). Cells were also co-stained with Aldefluor and anti-CD24–PE–Cy7 (561646) (1:100) and anti-CD44–APC (559942) (1:20) (BD Biosciences, Oxford, UK). A BD FACS LSRII flow cytometer was used and samples were analysed with BD FACS Diva software (BD Biosciences). mCherry+ DCIS cells from the tumours in the tail vein xenograft experiments were sorted by a FACSAriaIII (BD). FACSAriaIII was used for clonogenic and

mammosphere assays that required plating single cells per well. Living and dead cells were distinguished with DAPI 1:5000. Cell cycle analysis was described previously (Zvelebil et al., 2013).

PKH staining

2 x10^{-6} DCIS-LacZ or DCIS-SOX11 cells were resuspended in Diluent C and stained with PKH26 Dye Solution to a final concentration of 5mM following the manufacturer's protocol (MINI26, Sigma). After confirming by flow cytometry that the 100% of the population stained, spheres with 5000 cells were formed in ultra-low attachment 96-well plates.

Animal experiments

All animal work was carried out under UK Home Office project and personal licenses following local ethical approval from The Institute of Cancer Research Ethics Committee and in accordance with local and national guidelines. For xenograft tumour assays, DCIS.com labelled with Luc2-mCherry and stably transduced with empty vector (EV) control or SOX11 (in pInducer21 backbone) were resuspended in PBS for implantation into female NSG-Foxn1^{null} mice. 1.0x10^{6} cells/site in both sides were injected in each mammary fat pad number 4 and 2.0x104 cells/site into the mammary duct via the nipple of mammary gland number 4 as previously described (Oliemuller et al., 2017).

Intracranial injections were performed using a stereotaxic frame by injecting 1x10^{-5} cells (Stoelting, Wood Lane, IL, USA) into the striatum (2-mm right from the midline, 2-mm anterior from bregma, 3-mm deep). Six NSG-Foxn1null mice were used per for each condition.

For tail vein injections, the mice were placed into a hot box set at 38ºC for up to 5 minutes, to dilate the tail veins. The mice were then placed into a restrainer, the lateral tail vein identified, and 2.5x10^{5} cells were slowly infused (over 30 seconds) through the tail vein using a venoflux 25g butterfly needle. Before treatment, mice used were randomized in groups based on their weight.

For the doxycycline induction experiments using iSOX11 or iEV, a week before the injections, the animals injected with transduced cells were

separated into two cohorts and maintained with or without chow containing doxycycline (0.2 g/kg, 0.625, 1.250 or 2.0 g/kg from Envigo) for the duration of the experiment.

For *in vivo* imaging, mice were injected with 200μl of luciferin (XenoLight D-Luciferin Potassium Salt, Perkin Elmer, 30mg/mL). After 5 minutes the mice were anesthetized with isoflurane and the bioluminescence was measured at least at 3 different time points on an IVIS Lumina imaging systems (Perkin-Elmer). For *ex vivo* imaging the organs were resected and imaged for 3 minutes to detect any signal. In the analysis, identical square regions of interest (ROI) were drawn around tumours to measure total and average bioluminescence signal.

Tumour dissociation

iSOX11 or iEV cells from brain, bone and lung metastasis were isolated with a cell dissociation kit following the manufacturer instructions (Miltenyi Biotec, Bergisch Gladbach, Germany) and using the gentleMACS™ Octo Dissociator with Heaters and its 37ºC_h_TDK_3 protocol. To select human cells and discard murine cells, the result of this dissociation was FACS sorted in a FACSAriaIII and mCherry+ cells were selected and grown in normal DCIS media.

RNA sequencing

cDNA library preparation was carried out at Oxford Genomics Centre, The Wellcome Trust Centre for Human Genetics using PolyA+ RNA enrichment method for total RNA from cultured cells. mRNA fraction was selected from the total RNA before conversion to cDNA. Second strand cDNA synthesis incorporated dUTP. The cDNA was end-repaired, A-tailed and adapter-ligated. Prior to amplification, samples underwent uridine digestion. The prepared libraries were size selected, multiplexed and quality checked before paired end sequencing over four lanes of a flow cell.

Sequence files were trimmed by the use of trim_galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default settings. Trimmed data were separately mapped to the GRCh38 and

GRCm38 genome assemblies by the use of hisat2 (v2.0.5) with options --sp 1000,1000 -- omixed--no-discordant, and were filtered to remove non-primary alignments. Species-specific read sets were generated by removing any read that produced a valid alignment in both human and mouse from the results for both species. The remaining data were imported into SeqMonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) with a filter of mapping quality (MAPQ) score ≥ 20. Reads were quantified over the transcript set from Ensembl v78 with annotated mis-spliced, pseudogene and unannotated transcripts removed. Initial quantification was raw read counts from the opposing strand to the transcript, with all exons for each gene being collated into a single measure. This allowed gene-level differential expression to be assessed by the use of DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html), with a cutoff of a false discovery rate of < 0.05. Subsequent visualization was performed by requantifying expression as log2 fragments per million reads of library. RNA sequencing files were submitted to ArrayExpress.

Survival analysis

The prognostic importance of *SOX11* mRNA expression was assessed by the use of survival data using the cBioPortal for Cancer Genomics (http://cbioportal.org) (Cerami et al., 2012; Gao et al., 2013). Data obtained from The Cancer Genome Atlas (https://www.cancer.gov/tcga) was examined by use of the Kaplan–Meier Plotter survival analysis tool (http://kmplot.com) and METABRIC (Curtis et al., 2012), and statistical significance was determined with the Wald test.

Statistical Analysis

The data in the graphs are presented as mean and standard deviation, unless specified otherwise. Experiments were analysed with a two-tailed Student's t-test with a confidence interval of 95% when the number of groups equalled 2, or with a parametric ANOVA and *post hoc* test when the number of groups was >2, unless otherwise specified. *P-value* ≤ 0.0001 is considered as extremely significant (****), $P \leq 0.001$ as highly significant (***), $P \leq 0.01$ as

very significant (**), $P \leq 0.05$ as significant (*), and $P > 0.05$ as not significant (ns), respectively.

Acknowledgements

We thank Breast Cancer Now for funding this work as part of Programme Funding to the Breast Cancer Now Toby Robins Research Centre. This work was supported by a grant from CRUK (CRUK: A21855). This work was supported by grants from the Spanish Ministry of Education and Science (SAF2017-84934-R, to MV) and the Government of the Autonomous Community of the Basque Country, the Department of Industry, Tourism and Trade (Elkartek, IA-R and MV). We acknowledge the role of the Breast Cancer Now Tissue Bank in collecting and making available the normal breast samples used in the generation of this publication. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of the sequencing data.

Conflict of interest

The authors declare no conflict of interest.

Figure 1. Inducible expression of SOX11 leads to changes in cell state profiles of DCIS.com cells.

A. Western blot of SOX11 in cytoplasmic and nuclear fractions of DCIS.com cells containing the pInducer21 empty vector in presence (iEV) or absence (niEV) of 1μM Doxycycline (DOX) or the pInducer21SOX11 with (iSOX11) or without DOX (niSOX11). GAPDH and LAMIN B1 were used as loading control of cytoplasmic and nuclear fractions, respectively. Densitometry results normalised against niSOX11 are shown in brackets.

B. SOX11 expression detected in iSOX11 cells stained by IF after 48 hours of DOX induction. Scale Bar: 200μm.

C. ER- DCIS case sample showing SOX11 staining in DCIS and adjacent normal breast tissue. Scale Bar: 200μm.

D. Results from flow cytometry analysis of Aldefluor assays of niEV and niSOX11 cells (day 0) and iEV and iSOX11 after 2 days treatment with 1μM DOX. Results show the % of ALDH+ cells normalised against niEV. Error bars represent SD. * p = 0.0223. n=5.

E. Results from flow cytometry analysis of CD24 and CD44 of niEV and niSOX11 cells (day 0) and iEV and iSOX11 after treatment with 1μM DOX for 2 days. Results show the average % of cells CD44+/CD24+ in each condition. Error bars represent SD. *** p = 0.0005 (iSOX11 vs niSOX11) and p=0.0009 (iSOX11 vs iEV) n=3.

F. Western blot of CD24 in cytoplasmic and nuclear fractions of niEV, niSOX11, iEV and iSOX11 cells. GAPDH and LAMIN B1 were used as loading control of cytoplasmic and nuclear fractions, respectively. In brackets densitometry results normalised against niSOX11.

G. Confocal images of IF staining of E-CADHERIN, N-CADHERIN, VIMENTIN in niSOX11 and iSOX11 cells. Cells were grown in slides covered with Collagen I. Scale Bar: 200μm.

H. Western blot of N-CADHERIN and VIMENTIN in cytoplasmic and nuclear fractions of niEV, niSOX11, iEV and iSOX11 cells. GAPDH and LAMIN B1 were used as loading control of cytoplasmic and nuclear fractions respectively. Densitometry results normalised against niSOX11 are shown in brackets.

I. Confocal IF images of iEV and iSOX11 spheroids treated with 1μM DOX, stained with luminal marker, K8, and basal markers K14 or SMA, and DAPI. Scale bar: 200μm.

DOX: doxycycline, IF: Immunofluorescence.

Figure 2. DCIS cells expressing SOX11 grow more slowly and form more invasive spheroids.

A. Cell growth assays results for iEV or iSOX11 DCIS cells (induced with 1μM DOX for 72h). Experiments were performed 5 times. Error bars represent SEM. * p=0.0450 and **** p<0.0001.

B. Cell growth assays results for spheroid formed with iEV or iSOX11 DCIS cells induced with 1μM DOX for 3 or 7 days. Experiments were performed 3 times. Error bars represent SEM. p-values (3 days): * p=0.0374 (niEV vs niSOX11), * p=0.0221 (niSOX11 vs iSOX11) *** p=0.0002. p- values (7days): *** p=0.0004 and **** p<0.0001

C. Examples of DCIS iEV and DCIS iSOX11 spheroids grown on low attachment plates. Graph shows the number of microsatellites per sphere. **** p<0.0001.

D. Invasion assay after overlaying niEV, iEV, niSOX11 and iSOX11 DCIS spheroids with Collagen I. Scale bar: 200μm. Graph shows the area invaded in $pixel^2$ normalised against niEV. Both **** p < 0.0001.

E-G. IVIS imaging, tumour volumes and luminescence total flux/volume results after mammary fat pad xenografts. P-value in F: * p=0.0331. p-values in G: * p=0.0252 (niSOX11 vs iSOX11 625ppm), * p=0.0285 (iEV 625ppm vs iSOX11 625ppm), ** p=0.0082.

H. Representative images of IHC staining to detect Cleaved Caspase 3 and H&Es of tumours resected from mice injected with iEV and iSOX11 cells. Scale bar: 500μm.

I-K. IVIS imaging, tumour volumes and luminescence total flux after mammary intraductal injection (MIND) xenografts. * p=0.0286 (U Mann Whitney).

DOX: doxycycline, IHC: Immunohistochemistry

Figure 3. SOX11 expression promotes expression of developmental pathways frequently activated in cancer.

A. Volcano plot representing the RNAs with a log2 fold-change > +/-0.585 in the RNA-sequencing results of iSOX11 cells grown in 2D compared with the controls [(iSOX11-niSOX11)- (iEV-niEV)] to account for effects of DOX treatment on DCIS.com cells.

B. Gene ontology results from A.

C. List of genes overexpressed log2 fold-change >+/- 1.585 times in all 3 RNA-sequencing (cells grown in: 2D, 3D for 2 days, 3D for 5 days) results comparing iSOX11 versus iEV.

D. qRT-PCR results for several potential SOX11 targets in EV and SOX11 cells with and without DOX induction in cells grown in 2D. Experiment was repeated 3 times.

E. Western blot of MEX3A and TUBB3 in cytoplasmic and nuclear fractions of EV or SOX11 cells in presence or absence of 1μM DOX. GAPDH and LAMIN B1 were used as loading control of cytoplasmic and nuclear fractions respectively. In brackets, densitometry results normalised against niEV and niSOX11.

F. IF staining of DCIS iEV and DCIS iSOX11 cells with TUBB3 (green) and MEX3A (red). Scale: 200μm.

G. Western blot of MEX3A and TUBB3 in SOX11+ breast cancer cell lines and SOX11- DCIS.com and MCF10A from the MCF10A mammary cell progression series.

H. Pie charts representing the percentage of breast cancer samples with a log2 fold-change greater than 2 in the levels of *MEX3A* or *TUBB3* RNA when *SOX11* increased between 0.5 and 2-fold, 2 and 4-fold, or greater than 4-fold in the TCGA dataset.

DOX: doxycycline, qRT-PCR: Quantitative real time PCR, IF: Immunofluorescence.

Figure 4. DCIS cells expressing SOX11 show alterations in metastatic tropism.

A. Tabulated results of micrometastasis assessed by *ex vivo* IVIS imaging after orthotopic mammary fat pad xenografting of iEV and iSOX11 cells.
B. Tabulated results of micrometastasis assessed by *ex vivo* IVIS imaging after xenografting iEV and iSOX11 cells via the tail vein.
C. Representative *in vivo* IVIS image 7 days after tail vein injections of iEV and iSOX11 cells.
D. Quantification of *in vivo* lung metastatic burden at day 31. Graph shows the luminescence total flux (p/s) in the lungs 31 days after tail vein injections. ** p=0.0011.
E. Distance metastasis-free survival curve for *SOX11* in breast cancer patients in the Wang cohort (GSE2034).

Figure 5. SOX11+ DCIS cells isolated from brain metastasis display a colonisation and growth advantage after intracranial xenografting.
A. Western blot of SOX11, MEX3A and TUBB3 in total cell lysates of EV and SOX11 cell lines isolated from primary metastasis at indicated sites in presence or absence of DOX.
B. Representative *in vivo* IVIS imaging 7 days after tail vein injections of iSOX11 cells that were isolated from the brain metastasis (SOX11Br) in presence or absence of DOX.
C. Tabulated results of micrometastasis from *in vivo* IVIS imaging 7 days after tail vein injections of SOX11Br cells.
D. Tabulated results of micrometastasis from *ex vivo* IVIS imaging of the tail vein injections of SOX11Br cells.
E. IVIS imaging of mice fed normal chow or DOX-containing chow 10 days after intracranial injections of SOX11Br cells.
F. Survival curve for mice shown in E. * p=0.0195.
DOX: doxycycline

Figure 6. SOX11 regulates growth of ER- breast cancer cells.
A. qRT-PCR results for several potential SOX11 targets in CAL-148 cells transduced with shRNA to *SOX11* or shRNA NS cells with and without DOX induction in cells grown in 2D.

B. Western blot of SOX11, MEX3A and TUBB3 in total cell lysates of CAL-148 cells transduced with shRNA *SOX11* in presence or absence of 1μM DOX after 48h. GAPDH was used as loading control. Densitometry results normalised against no DOX are shown in brackets.

C. Cell growth assay results for CAL-148 shRNA *SOX11* cells induced with 1μM DOX at 3 and 7 days. Experiments were performed 3 times. Error bars represent SEM. * p=0.0106 (day 7).

D. Quantification of clonogenicity in 2D and 3D from single CAL-148 shRNA *SOX11* cells plated in presence or absence of DOX after 21 days. The number in brackets represents the mean in each group of the 3 experimental replicates.

DOX: doxycycline, NS: non-silencing.

Figure 7. TUBB3, an established SOX11 target, regulates cell growth and invasive potential of ER- breast cancer cells.

A. Western blot of TUBB3 in total cell lysates of BT-20 cells transfected with siRNAs specific for *TUBB3*. B-ACTIN was used as loading control. Densitometry results normalised against NS siRNA are shown in brackets.

B. Cell growth assay results 5 days after BT-20 cells were transfected with siRNA specific for *TUBB3* (since siRNAs require 48 hours for efficient knockdown). Results relative to NS siRNA are shown. Experiments performed 3 times. * p=0.0297, *** p=0.0001, **** p<0.0)001.

C. Sphere size measured 3 and 5 days after BT-20 cells were transfected with siRNA specific for *TUBB3*. Results relative to NS siRNA are shown. Experiments performed 3 times. **** p<0.0001.

D. Cell cycle analysis performed by flow cytometry at day 2 after siRNA transfection specific for *TUBB3* in BT-20 cells. Graph shows % of cell in each phase of the cell cycle. * p=0.0396 (Phase G2/M siRNApool vs siRNA NS) and *p=0.0243 (Phase G2/M siRNA5 vs siRNA NS). Experiments were performed 3 times.

E. Invasion assay after overlaying BT-20 spheroids with Collagen I at day 2 after siRNA transfection specific for *TUBB3.* Graph shows the area invaded in $pixel^2$ normalised against NS siRNA. * p=0.0444, ** p=0.0014, **** p<0.0001.

F. Representative images of BT-20 spheroids transfected with indicated siRNAs to *TUBB3* 48h after adding Collagen I and (in small insets) at time 0h (2 days after transfection). Scale bar: 200μm.
NS: non-silencing.

Figure 8. MEX3A, a novel SOX11 downstream effector, regulates cell growth and (E/M) state of ER- breast cancer cells.

A. Invasion assay results after overlaying BT-20 and HCC1187 spheroids with Collagen I, at day 2 after siRNA transfection specific for *MEX3A* or control (NS). Graph shows the area invaded in $pixel^2$ normalised against NS siRNA. * p=0.0181, ** p=0.0014 for BT-20 and * p= 0.0220 for HCC1187 cells.

B. Representative images of BT-20 spheroids 48h after adding Collagen I and at time 0h (shown in small inset, 2 days after transfection) made from cells transfected with either control siRNA or *MEX3A* siRNAs. Scale bar: 200μm.

C. Cell growth assays 5 days after CAL-148 and HCC1187 cells were transfected with siRNA specific for *MEX3A* or NS controls (siRNAs require 48 hours for efficient knockdown; this was taken into account to select day 5 as final point). Relative results to NS siRNA are shown. Experiments performed 3 times. * p=0.0342, ** p= 0.0052 *** p=0.0005 (siRNA2), **** p<0.0003 (siRNA3) for CAL-148 cells and * p=0.0337 (siRNApool) and *p=0.0477 for HCC1187 cells.

D. Sphere size measured 3, and 5 days after HCC1187 cells were transfected with siRNA specific for *MEX3A*. Relative results to NS siRNA are shown. Experiments performed 3 times * p=0.0240.

E. Cell cycle analysis performed by flow cytometry at day 4 after siRNA transfection specific for *MEX3A* in CAL-148 cells. Graph shows % of cell in each phase of the cell cycle. * p=0.0147 (phase S siRNApool vs siRNA NS) and * p=0.0094 (phase S siRNA3 vs siRNA NS) and * p=0.0419 (phase subG1 NS vs siRNA3). Experiments performed 4 times.

F. Western blot of RRM2 and SURVIVIN in total cell lysates of CAL-148 at day 4 after siRNA transfection specific for *MEX3A*. β-ACTIN was used as loading control. Densitometry results normalised against NS siRNA are shown in brackets.

G. Examples of morphological changes observed in CAL-148 cells after 4 days of *MEX3A* knockdown compared to NS control. Scale bar: 200μm.

H. Western Blot of EPCAM and E-CADHERIN in CAL-148 cells transfected with siRNA specific for *MEX3A* or NS control. β-ACTIN was used as loading control. Densitometry results normalised against NS siRNA are in brackets. NS: non-silencing.

Table 1. iSOX11 spheroids express reactivated embryonic wound signature.

Genes upregulated in mouse epidermal cells at E13.5 and at wound edge that are directly regulated by Sox11 and Sox4 in both E16.5 epidermis and keratinocytes in (Miao et al., 2019) are significantly upregulated in iSOX11 cells grown as spheroids.

3D (2 days)	log2 fold-change	p-value
GNG2	2.30823519	4.85E-05
RCOR2	1.57478431	2.90E-05
MARCKSL1	1.06239497	7.15E-08
EVL	0.8482472	9.60E-06
SNN	0.78560135	3.11E-05
FBLIM1	0.78250288	3.12E-05
ETV4	0.77064674	8.27E-05
VCAN	0.67955833	0.00094939
TWIST2	0.64425385	0.01949167
PXDN	0.63841224	1.14E-05
ARHGEF2	0.58601526	2.88E-06
TMSB10	0.56607317	0.00030783
C4orf48	0.54288141	0.03406946

Supplementary Information

Supplementary Files:

Supplementary File 1. RNA-seq results when SOX11 is induced in cells grown in 2D, 3D for 2 days or 3D for 5 days.

Supplementary File 2. Gene ontology analysis of the genes differentially expressed in the 3 RNA-seq datasets when SOX11 is induced in cells grown in 2D, 3D for 2 days or 3D for 5 days.

Supplementary File 3. Genes differentially expressed the 3 RNA-seq datasets when SOX11 is induced in cells grown in 2D, 3D for 2 days or 3D for 5 days.

Supplementary File 4. Expression values of the genes from stem cell quiescence signature in the 3 datasets obtained in DCIS when SOX11 is induced in cells grown in 2D, 3D for 2 days or 3D for 5 days.

Supplementary File 5. Co-occurrence and correlation of *MEX3A* RNA levels with cell cycle related genes in TCGA breast cancer dataset.

Supplementary File 6. Cell lines and culture media.

Supplementary File 7. qPCR probes.

Supplementary File 8. Antibodies used for western blotting.

Supplementary File 9. Antibodies used for IF and IHC.

Figure Supplements:

Figure 1—figure supplement 1.

A. Western Blot to detect SOX11 in nuclear proteins of DCIS.com cells transduced with pLenti6.3 LacZ or SOX11.

B. Western blot to detect SOX11 in total cell lysates to determine the concentration of DOX needed to induce SOX11.

C. Western blot to detect SOX11 with total cell lysates induced with 1μM DOX at different times.

D. Western blot to detect SOX11 with total cell lysates induced with 1μM DOX for 24h and cultured in media without DOX at different times afterwards.

E. Representative flow cytometry analysis of CD44/CD24 cell populations of EV and SOX11 cells before and after 2day treatment with DOX. Experiments were performed 3 times.

F. Fluorescence levels of iEV and iSOX11 after treatment with DOX stained by IF with CD24. The positively stained area for CD24 at different DOX concentrations was normalised by the area stained with DAPI. The results are shown normalised against no DOX.

DOX: Doxycycline, IF: Immunofluorescence.

Figure 1—figure supplement 2.

A. ALDH activity levels in EV and SOX11 cells before and after 2 days treatment with DOX were detected with the Aldefluor assay. Cells were stained and sorted with CD44 and CD24 antibodies, and ALDH activity was measured with the Aldefluor kit. Representative ALDH activities after flow cytometry analysis in CD44/CD24+ EV and SOX11 cell populations before and after 2 days treatment with DOX are shown**.** +DEAB plots display the negative control; cells incubated with diethylaminobenzaldehyde (DEAB), the specific inhibitor of ALDH, were used to establish the baseline fluorescence of these cells are shown in insets. Experiments were performed 5 times.

B. The frequency of CD44+/CD24+ALDH cells are shown in EV and SOX11 cells before and after 2 days treatment with DOX. Error bars represent SD. Experiment performed 5 times. ** p=0.0029.

DOX: Doxycycline.

Figure 1—figure supplement 3.

A. E-CADHERIN, N-CADHERIN, VIMENTIN IF staining of EV or SOX11 cells grown in collagen induced with and without DOX treatment. Scale Bar: 200μm.

B. Fluorescence levels of iEV and iSOX11 stained by IF with SOX11, VIM, K5 and K8. The positively stained area for each protein was normalised by the area stained with DAPI. The results are shown normalised against niEV and niSOX11.

C. VIMENTIN IF and DAPI staining of EV or SOX11 cells induced with DOX and without DOX treatment. Scale Bar: 200μm.DOX: Doxycycline, IF: Immunofluorescence.

Figure 1—figure supplement 4.

Confocal images of spheres formed with niEV, niSOX11, iEV and iSOX11 cells induced for 2 or 5 days with 1μM DOX, and stained by IF with K14, K8 and SOX11. Scale Bar: 200μm.

DOX: Doxycycline, IF: Immunofluorescence.

Figure 1—figure supplement 5.

Confocal images of spheres formed with niEV, niSOX11, iEV and iSOX11 cells induced for 2 or 5 days with 1μM DOX, and stained by IF with K14, SMA and SOX11. Scale Bar: 200μm.

DOX: Doxycycline, IF: Immunofluorescence.

Figure 2—figure supplement 1.

A. Representative colony forming assay results using niEV, iEV, niSOX11 and iSOX11 cells. The area stained with crystal violet was quantified and normalised by the area of niEV or niSOX11 respectively. Representative photo is shown. Experiment was performed 3 times with 3 replicates each.

B. Percentage of iEV and iSOX11 cells plated in two-dimensional (2D) culture that form colonies. Cells were sorted by FACS and single cells were plated per well in a 96-well. Results are shown as the % of clones obtained divided

by the cells seeded in each cell line. Experiment was performed twice with 30 cells for each cell line.

C. Percentage of DCIS-LacZ and DCIS-SOX11 cells plated in two-dimensional (2D) culture that form colonies. Cells were sorted by FACS sorted and single cells were plated per well in a 96-well. Results are shown as the % of clones obtained divided by the cells seeded in each cell line. Experiment was performed 3 times with 30 cells for each cell line. Typical morphologies observed for DCIS-LacZ and DCIS-SOX11 clones are shown.

D. Quantification of sphere-initiating capacity of DCIS-LacZ and DCIS-SOX11 cells plated in three-dimensional (3D) culture that form spheres grown in methylcellulose and ultra-low attachment plates. Results are shown as the % of clones larger than 50um (upper graph) or as a percentage of total clones (down) obtained divided by the cells seeded in each cell line. Experiments were performed twice with six biological replicates each. 10,000 cells were plated per well. Typical morphologies observed for spheroid formed from DCIS-lacZ and DCIS-SOX11 cells grown in methylcellulose and ultra-low attachment plates.

E. Representative images showing PKH26 dye labelling of DCIS-lacZ and DCIS-SOX11 spheres grown in ultra-low attachment 96 well plates after 14 days. Graphs represent the size of the spheres formed (pixel^2) and the percentage of the area that is PKH+ in each sphere. Experiment was replicated twice (n=35).

Figure 2—figure supplement 2.

SOX11 levels in breast cancer METABRIC dataset.

Figure 2—figure supplement 3.

A. Average luminescence results in IVIS images after mammary fat pad xenografts of iEV or iSOX11 in mice that were fed with high doses of DOX.

B. Tumour volumes after mammary fat pad xenografts.

C. Representative images of IHC to detect Cleaved Caspase 3 in tumours resected from mice injected with iEV and iSOX11 tumours. Scale bar: 500μm.

D. Average luminescence results in IVIS images of mice xenografted with iEV and iSOX11 cells at day 6, 14, 21 and 28 after placing half on DOX or off DOX diet after an initial 42 days on DOX.
DOX: Doxycycline.

Figure 2—figure supplement 4.
Panel shows photomicrographs of primary fat pad tumours following immunohistochemical detection of SOX11, ALDH1A1, CD24, and H&E staining. Data shown are from mice xenografted with iEV and iSOX11 cells and fed DOX chow.
DOX: Doxycycline, H&E: Haematoxylin/Eosin.

Figure 3—figure supplement 1.
A. Volcano plots representing the genes with a log2 fold-change > +/-0.585 in the RNA-sequencing results of comparing iSOX11 and iEV cells grown in 3D with DOX for 2 (left) or 5 days (right).
B. Gene ontology results from A.
C. List of genes overexpressed log2 fold-change >+/- 0.585 times in all 3 RNA-sequencing (cells grown in: 2D, 3D for 2 days, 3D for 5 days) results comparing iSOX11 versus iEV.
D. Western blot of RCOR2 in cytoplasmic and nuclear fractions of EV or SOX11 cells in presence or absence of 1μM DOX. GAPDH and LAMIN B1 were used as loading control of cytoplasmic and nuclear fractions, respectively. Densitometry results normalised against niEV and niSOX11 are shown in brackets.
E. Pie charts representing the percentage of samples with a log2 fold-change greater than 2 in the levels of *MEX3A* or *TUBB3* RNA when SOX11 is increased between 0.5 and 2-fold, 2 and 4-fold, or greater than 4-fold in the Broad cell line dataset.
DOX: Doxycycline.

Figure 4—figure supplement 1.

A. *SOX11* expression in breast cancer metastasis (Zhang cohort: GSE140200).

B. *SOX11* is amplified and highly expressed in ~30% breast cancer brain metastasis (BCBM) from ER-, ER+, HER2+ and *BRCA1-/-* breast cancer patients in a study by (Saunus et al., 2015).

Figure 4—figure supplement 2.

A. Graph showing the log2 values of SOX11 in primary tumours and its correlative log2 values in the brain metastasis dataset from (Vareslija et al., 2018).

B. Table resuming the SOX11 log2 fold-change between the values in the brain metastasis and the primary tumour and the ER, PR and HER2 status of each case.

Figure 7—figure supplement 1.

A. Distant metastasis-free survival of breast cancer patients with low or high expression of *CD24*, *TUBB3*, and *MEX3A*.

B. Overall survival of breast cancer patients with low or high expression of *CD24*, *TUBB3*, and *MEX3A*.

Figure 7—figure supplement 2.

Representative flow cytometry histograms of cell cycle analysis of BT-20 cells after transfection with specific *TUBB3* siRNAs.

Figure 8—figure supplement 1.

A. Representative images of HCC1187 spheroids 48h after adding Collagen I and at time 0h (shown in small inset, 2 days after transfection) made from cells transfected with either control siRNA or *MEX3A* siRNAs. Scale bar: 200μm.

B. Percentage of cells in S phase normalised against cells transfected with the NS siRNA in CAL-148 cells at day 4. Experiments performed 3 times.

C. Percentage of cells in S phase normalised against cells transfected with the NS siRNA in HCC1187 cells at day 2. Experiments performed 4 times.

D. Cell cycle analysis performed by flow cytometry at day 2 after siRNA transfection specific for *MEX3A* in HCC1187 cells. Graph shows % of cell in each phase of the cell cycle. Experiments performed 4 times.
E. Representative flow cytometry histograms of cell cycle analysis of CAL-148 and HCC1187 cells after transfection with specific *MEX3A* siRNAs.
F. Heatmap representing the values of a consensus stem cell quiescence signature in the RNA-sequencing results of comparing iSOX11 and iEV cells grown in 2D for 2 days or 3D with DOX for 2 or 5 days.
G. Western blot of RRM2 and SURVIVIN in total cell lysates of BT-20 at day 2 after siRNA transfection specific for *MEX3A*. β-ACTIN was used as loading control. Densitometry results normalised against NS siRNA are shown in brackets.
DOX: Doxycycline, NS: non-silencing.

Figure 8—figure supplement 2.
A. Western Blot to detect MEX3A in CAL-148 cells that were grown in serum free media or media supplemented with 0.2% FBS for 24h in presence or absence of EGF 1ng/ml. GAPDH was used as loading control. Densitometry results normalised against NS siRNA are in brackets.
B. Western Blot to detect MEX3A and EPCAM in BT-20 cells transfected with siRNA specific for *MEX3A* or NS control. β-ACTIN was used as loading control. Densitometry results normalised against NS siRNA are in brackets.
C. Western Blot to detect MEX3A and EPCAM in HCC1187 cells transfected with siRNA specific for *MEX3A* or NS control. LAMIN B1 was used as loading control. Densitometry results normalised against NS siRNA are in brackets.
FBS: Fetal bovine serum, NS: non-silencing.

References

Aleskandarany, M. A., Soria, D., Green, A. R., Nolan, C., Diez-Rodriguez, M., Ellis, I. O., & Rakha, E. A. (2015). Markers of progression in early-stage invasive breast cancer: a predictive immunohistochemical panel algorithm for distant recurrence risk stratification. *Breast Cancer Res Treat, 151*(2), 325-333. doi:10.1007/s10549-015-3406-3

Balta, E. A., Wittmann, M. T., Jung, M., Sock, E., Haeberle, B. M., Heim, B., . . . Lie, D. C. (2018). Phosphorylation Modulates the Subcellular Localization of SOX11. *Front Mol Neurosci, 11*, 211. doi:10.3389/fnmol.2018.00211

Barkal, A. A., Brewer, R. E., Markovic, M., Kowarsky, M., Barkal, S. A., Zaro, B. W., . . . Weissman, I. L. (2019). CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. *Nature, 572*(7769), 392-396. doi:10.1038/s41586-019-1456-0

Barriga, F. M., Montagni, E., Mana, M., Mendez-Lago, M., Hernando-Momblona, X., Sevillano, M., . . . Batlle, E. (2017). Mex3a Marks a Slowly Dividing Subpopulation of Lgr5+ Intestinal Stem Cells. *Cell Stem Cell, 20*(6), 801-816 e807. doi:10.1016/j.stem.2017.02.007

Bland, P., & Howard, B. A. (2018). Mammary lineage restriction in development. *Nat Cell Biol, 20*(6), 637-639. doi:10.1038/s41556-018-0111-6

Bos, P. D., Zhang, X. H., Nadal, C., Shu, W., Gomis, R. R., Nguyen, D. X., . . . Massague, J. (2009). Genes that mediate breast cancer metastasis to the brain. *Nature, 459*(7249), 1005-1009. doi:10.1038/nature08021

Bufalieri, F., Caimano, M., Lospinoso Severini, L., Basili, I., Paglia, F., Sampirisi, L., . . . Di Marcotullio, L. (2020). The RNA-Binding Ubiquitin Ligase MEX3A Affects Glioblastoma Tumorigenesis by Inducing Ubiquitylation and Degradation of RIG-I. *Cancers (Basel), 12*(2). doi:10.3390/cancers12020321

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., . . . Schultz, N. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov, 2*(5), 401-404. doi:10.1158/2159-8290.CD-12-0095

Cheung, T. H., & Rando, T. A. (2013). Molecular regulation of stem cell quiescence. *Nat Rev Mol Cell Biol, 14*(6), 329-340. doi:10.1038/nrm3591

Clayton, H., Titley, I., & Vivanco, M. (2004). Growth and differentiation of progenitor/stem cells derived from the human mammary gland. *Exp Cell Res, 297*(2), 444-460. doi:10.1016/j.yexcr.2004.03.029

Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., . . . Aparicio, S. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature, 486*(7403), 346-352. doi:10.1038/nature10983

Domenici, G., Aurrekoetxea-Rodriguez, I., Simoes, B. M., Rabano, M., Lee, S. Y., Millan, J. S., . . . Vivanco, M. D. (2019). A Sox2-Sox9 signalling axis maintains human breast luminal progenitor and breast cancer stem cells. *Oncogene, 38*(17), 3151-3169. doi:10.1038/s41388-018-0656-7

Dong, J., Hu, Y., Fan, X., Wu, X., Mao, Y., Hu, B., . . . Tang, F. (2018). Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome Biol, 19*(1), 31. doi:10.1186/s13059-018-1416-2

1250 Duex, J. E., Owens, C., Chauca-Diaz, A., Dancik, G. M., Vanderlinden, L. A., Ghosh, D.,
1251 . . . Theodorescu, D. (2017). Nuclear CD24 Drives Tumor Growth and Is
1252 Predictive of Poor Patient Prognosis. *Cancer Res, 77*(18), 4858-4867.
1253 doi:10.1158/0008-5472.CAN-17-0367
1254 Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., . . . Schultz,
1255 N. (2013). Integrative analysis of complex cancer genomics and clinical
1256 profiles using the cBioPortal. *Sci Signal, 6*(269), pl1.
1257 doi:10.1126/scisignal.2004088
1258 Georgatos, S. D., & Blobel, G. (1987a). Lamin B constitutes an intermediate
1259 filament attachment site at the nuclear envelope. *J Cell Biol, 105*(1), 117-
1260 125. doi:10.1083/jcb.105.1.117
1261 Georgatos, S. D., & Blobel, G. (1987b). Two distinct attachment sites for vimentin
1262 along the plasma membrane and the nuclear envelope in avian
1263 erythrocytes: a basis for a vectorial assembly of intermediate filaments. *J*
1264 *Cell Biol, 105*(1), 105-115. doi:10.1083/jcb.105.1.105
1265 Ghandi, M., Huang, F. W., Jane-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E.
1266 R., 3rd, . . . Sellers, W. R. (2019). Next-generation characterization of the
1267 Cancer Cell Line Encyclopedia. *Nature, 569*(7757), 503-508.
1268 doi:10.1038/s41586-019-1186-3
1269 Ginestier, C., Hur, M. H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., . . .
1270 Dontu, G. (2007). ALDH1 is a marker of normal and malignant human
1271 mammary stem cells and a predictor of poor clinical outcome. *Cell Stem*
1272 *Cell, 1*(5), 555-567. doi:10.1016/j.stem.2007.08.014
1273 Gusterson, B., & Eaves, C. J. (2018). Basal-like Breast Cancers: From Pathology to
1274 Biology and Back Again. *Stem Cell Reports, 10*(6), 1676-1686.
1275 doi:10.1016/j.stemcr.2018.04.023
1276 Halbleib, J. M., & Nelson, W. J. (2006). Cadherins in development: cell adhesion,
1277 sorting, and tissue morphogenesis. *Genes Dev, 20*(23), 3199-3214.
1278 doi:10.1101/gad.1486806
1279 Jiang, H., Zhang, X., Luo, J., Dong, C., Xue, J., Wei, W., . . . Yang, C. (2012).
1280 Knockdown of hMex-3A by small RNA interference suppresses cell
1281 proliferation and migration in human gastric cancer cells. *Mol Med Rep,*
1282 *6*(3), 575-580. doi:10.3892/mmr.2012.943
1283 Jolicoeur, F., Gaboury, L. A., & Oligny, L. L. (2003). Basal cells of second trimester
1284 fetal breasts: immunohistochemical study of myoepithelial precursors.
1285 *Pediatr Dev Pathol, 6*(5), 398-413. doi:10.1007/s10024-003-1125-y
1286 Karki, R., Mariani, M., Andreoli, M., He, S., Scambia, G., Shahabi, S., & Ferlini, C.
1287 (2013). betaIII-Tubulin: biomarker of taxane resistance or drug target?
1288 *Expert Opin Ther Targets, 17*(4), 461-472.
1289 doi:10.1517/14728222.2013.766170
1290 Kim, J. W., Young, J. S., Solomaha, E., Kanojia, D., Lesniak, M. S., & Balyasnikova, I.
1291 V. (2015). A novel single-chain antibody redirects adenovirus to
1292 IL13Ralpha2-expressing brain tumors. *Sci Rep, 5*, 18133.
1293 doi:10.1038/srep18133
1294 Kogata, N., Bland, P., Tsang, M., Oliemuller, E., Lowe, A., & Howard, B. A. (2018).
1295 Sox9 regulates cell state and activity of embryonic mouse mammary
1296 progenitor cells. *Commun Biol, 1*, 228. doi:10.1038/s42003-018-0215-3

Lee, J. Y., Park, K., Lee, E., Ahn, T., Jung, H. H., Lim, S. H., . . . Park, Y. H. (2016). Gene Expression Profiling of Breast Cancer Brain Metastasis. *Sci Rep, 6*, 28623. doi:10.1038/srep28623

Lilja, A. M., Rodilla, V., Huyghe, M., Hannezo, E., Landragin, C., Renaud, O., . . . Fre, S. (2018). Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland. *Nat Cell Biol, 20*(6), 677-687. doi:10.1038/s41556-018-0108-1

Lim, E., Vaillant, F., Wu, D., Forrest, N. C., Pal, B., Hart, A. H., . . . Lindeman, G. J. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med, 15*(8), 907-913. doi:10.1038/nm.2000

Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., . . . Wicha, M. S. (2014). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports, 2*(1), 78-91. doi:10.1016/j.stemcr.2013.11.009

Maguire, S. L., Peck, B., Wai, P. T., Campbell, J., Barker, H., Gulati, A., . . . Natrajan, R. (2016). 3D modelling identifies novel genetic dependencies associated with breast cancer progression in the isogenic MCF10 model. *J Pathol*. doi:10.1002/path.4778

Meerbrey, K. L., Hu, G., Kessler, J. D., Roarty, K., Li, M. Z., Fang, J. E., . . . Elledge, S. J. (2011). The pINDUCER lentiviral toolkit for inducible RNA interference in vitro and in vivo. *Proc Natl Acad Sci U S A, 108*(9), 3665-3670. doi:10.1073/pnas.1019736108

Mehta, G. A., Khanna, P., & Gatza, M. L. (2019). Emerging Role of SOX Proteins in Breast Cancer Development and Maintenance. *J Mammary Gland Biol Neoplasia, 24*(3), 213-230. doi:10.1007/s10911-019-09430-6

Miao, Q., Hill, M. C., Chen, F., Mo, Q., Ku, A. T., Ramos, C., . . . Nguyen, H. (2019). SOX11 and SOX4 drive the reactivation of an embryonic gene program during murine wound repair. *Nat Commun, 10*(1), 4042. doi:10.1038/s41467-019-11880-9

Mozos, A., Royo, C., Hartmann, E., De Jong, D., Baro, C., Valera, A., . . . Campo, E. (2009). SOX11 expression is highly specific for mantle cell lymphoma and identifies the cyclin D1-negative subtype. *Haematologica, 94*(11), 1555-1562. doi:10.3324/haematol.2009.010264

Nieman, M. T., Prudoff, R. S., Johnson, K. R., & Wheelock, M. J. (1999). N-cadherin promotes motility in human breast cancer cells regardless of their E-cadherin expression. *J Cell Biol, 147*(3), 631-644. doi:10.1083/jcb.147.3.631

Nieto, M. A., Huang, R. Y., Jackson, R. A., & Thiery, J. P. (2016). EMT: 2016. *Cell, 166*(1), 21-45. doi:10.1016/j.cell.2016.06.028

O'Brien-Ball, C., & Biddle, A. (2017). Reprogramming to developmental plasticity in cancer stem cells. *Dev Biol, 430*(2), 266-274. doi:10.1016/j.ydbio.2017.07.025

Oliemuller, E., Kogata, N., Bland, P., Kriplani, D., Daley, F., Haider, S., . . . Howard, B. A. (2017). SOX11 promotes invasive growth and ductal carcinoma in situ progression. *J Pathol, 243*(2), 193-207. doi:10.1002/path.4939

Pereira, B., Sousa, S., Barros, R., Carreto, L., Oliveira, P., Oliveira, C., . . . Almeida, R. (2013). CDX2 regulation by the RNA-binding protein MEX3A: impact on

intestinal differentiation and stemness. *Nucleic Acids Res, 41*(7), 3986-3999. doi:10.1093/nar/gkt087

Qi, J., Chen, N., Wang, J., & Siu, C. H. (2005). Transendothelial migration of melanoma cells involves N-cadherin-mediated adhesion and activation of the beta-catenin signaling pathway. *Mol Biol Cell, 16*(9), 4386-4397. doi:10.1091/mbc.e05-03-0186

Raouf, A., Zhao, Y., To, K., Stingl, J., Delaney, A., Barbara, M., . . . Eaves, C. (2008). Transcriptome analysis of the normal human mammary cell commitment and differentiation process. *Cell Stem Cell, 3*(1), 109-118. doi:10.1016/j.stem.2008.05.018

Rezaei, M., Friedrich, K., Wielockx, B., Kuzmanov, A., Kettelhake, A., Labelle, M., . . . Breier, G. (2012). Interplay between neural-cadherin and vascular endothelial-cadherin in breast cancer progression. *Breast Cancer Res, 14*(6), R154. doi:10.1186/bcr3367

Saunus, J. M., McCart Reed, A. E., Lim, Z. L., & Lakhani, S. R. (2017). Breast Cancer Brain Metastases: Clonal Evolution in Clinical Context. *Int J Mol Sci, 18*(1). doi:10.3390/ijms18010152

Saunus, J. M., Quinn, M. C., Patch, A. M., Pearson, J. V., Bailey, P. J., Nones, K., . . . Lakhani, S. R. (2015). Integrated genomic and transcriptomic analysis of human brain metastases identifies alterations of potential clinical significance. *J Pathol, 237*(3), 363-378. doi:10.1002/path.4583

Shepherd, J. H., Uray, I. P., Mazumdar, A., Tsimelzon, A., Savage, M., Hilsenbeck, S. G., & Brown, P. H. (2016). The SOX11 transcription factor is a critical regulator of basal-like breast cancer growth, invasion, and basal-like gene expression. *Oncotarget, 7*(11), 13106-13121. doi:10.18632/oncotarget.7437

Tanei, T., Morimoto, K., Shimazu, K., Kim, S. J., Tanji, Y., Taguchi, T., . . . Noguchi, S. (2009). Association of breast cancer stem cells identified by aldehyde dehydrogenase 1 expression with resistance to sequential Paclitaxel and epirubicin-based chemotherapy for breast cancers. *Clin Cancer Res, 15*(12), 4234-4241. doi:10.1158/1078-0432.CCR-08-1479

Tsang, S. M., Oliemuller, E., & Howard, B. A. (2020). Regulatory roles for SOX11 in development, stem cells and cancer. *Semin Cancer Biol*. doi:10.1016/j.semcancer.2020.06.015

Vareslija, D., Priedigkeit, N., Fagan, A., Purcell, S., Cosgrove, N., O'Halloran, P. J., . . . Young, L. S. (2018). Transcriptome Characterization of Matched Primary Breast and Brain Metastatic Tumors to Detect Novel Actionable Targets. *J Natl Cancer Inst*. doi:10.1093/jnci/djy110

Venkov, C., Plieth, D., Ni, T., Karmaker, A., Bian, A., George, A. L., Jr., & Neilson, E. G. (2011). Transcriptional networks in epithelial-mesenchymal transition. *PLoS One, 6*(9), e25354. doi:10.1371/journal.pone.0025354

Vervoort, S. J., Lourenco, A. R., van Boxtel, R., & Coffer, P. J. (2013). SOX4 mediates TGF-beta-induced expression of mesenchymal markers during mammary cell epithelial to mesenchymal transition. *PLoS One, 8*(1), e53238. doi:10.1371/journal.pone.0053238

Wahl, G. M., & Spike, B. T. (2017). Cell state plasticity, stem cells, EMT, and the generation of intra-tumoral heterogeneity. *NPJ Breast Cancer, 3*, 14. doi:10.1038/s41523-017-0012-z

Wang, X., Shan, Y. Q., Tan, Q. Q., Tan, C. L., Zhang, H., Liu, J. H., . . . Liu, X. B. (2020). MEX3A knockdown inhibits the development of pancreatic ductal adenocarcinoma. *Cancer Cell Int, 20*, 63. doi:10.1186/s12935-020-1146-x

Wansbury, O., Mackay, A., Kogata, N., Mitsopoulos, C., Kendrick, H., Davidson, K., . . . Howard, B. A. (2011). Transcriptome analysis of embryonic mammary cells reveals insights into mammary lineage establishment. *Breast Cancer Res, 13*(4), R79. doi:10.1186/bcr2928

Wuidart, A., Sifrim, A., Fioramonti, M., Matsumura, S., Brisebarre, A., Brown, D., . . . Blanpain, C. (2018). Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat Cell Biol, 20*(6), 666-676. doi:10.1038/s41556-018-0095-2

Yang, D., Jiao, Y., Li, Y., & Fang, X. (2020). Clinical characteristics and prognostic value of MEX3A mRNA in liver cancer. *PeerJ, 8*, e8252. doi:10.7717/peerj.8252

Yang, P., Wang, Y., Chen, J., Li, H., Kang, L., Zhang, Y., . . . Gao, S. (2011). RCOR2 is a subunit of the LSD1 complex that regulates ESC property and substitutes for SOX2 in reprogramming somatic cells to pluripotency. *Stem Cells, 29*(5), 791-801. doi:10.1002/stem.634

Zvelebil, M., Oliemuller, E., Gao, Q., Wansbury, O., Mackay, A., Kendrick, H., . . . Howard, B. A. (2013). Embryonic mammary signature subsets are activated in Brca1-/- and basal-like breast cancers. *Breast Cancer Res, 15*(2), R25. doi:10.1186/bcr3403

1

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
gene (*Homo sapiens*)	*SOX11*	DNASU	Gene ID: 6664 Clone HsCD00295480 19	
strain, strain background (*Mus Musculus*, Female)	NSG-Foxn1^{null}	in house		from breeding colony at ICR Biological Services Unit
cell line (*Homo sapiens*)	MCF10DCIS.com	Dr. Gillian Farnie		
cell line (*Homo sapiens*)	CAL-148	DMSZ	ACC 460	
cell line (*Homo sapiens*)	BT-20	ATCC	HTB-19	
cell line (*Homo sapiens*)	HCC1187	ATCC	CRL-2322	
cell line (*Homo sapiens*)	MCF10A	ATCC	CRL-10317	
cell line (*Homo sapiens*)	BT474	ATCC	HTB-20	
cell line (*Homo sapiens*)	BT549	ATCC	HTB-122	
cell line (*Homo sapiens*)	HCC202	Dr. Paul Huang, ICR, London	CRL-2316	
cell line (*Homo*	MX-1	DMSZ	CVCL_4774	

sapiens)				
cell line (*Homo sapiens*)	UACC893	ATCC	CRL-1902	
transfected construct (*Homo sapiens*)	pInducer21-SOX11	this paper	SOX11 coding sequence (GENEID: 6664) was subcloned into pInducer21 (ORF-EG) plasmid Cat# 46948 (See Expression vectors in Material and methods)	Lentiviral construct to transfect and express SOX11 sequence
transfected construct (*Homo sapiens*)	pInducer13-SOX11 shRNA 174	this paper	specific shRNA174 for SOX11 from pGIPZ plasmids from Horizon was subcloned into pInducer13 (miR-LUP) plasmid Cat# 46936 (See Expression vectors in Material and methods)	Lentiviral construct to transfect and express the shRNA
transfected construct (*Homo sapiens*)	pInducer13-shRNA NS	this paper	NS: non-silencing shRNA from pGIPZ plasmid from Horizon was subcloned into pInducer13 (miR-LUP) plasmid Cat# 46936 (See Expression vectors in Material and methods)	Lentiviral construct to transfect and express the shRNA
transfected construct (*Homo sapiens*)	siRNA: siGENOME Non-Targeting siRNA #1	Horizon Discovery	D-001210-01-20	
transfected construct (*Homo sapiens*)	siRNA: siGENOME SMARTpool TUBB3 siRNA	Horizon Discovery	MQ-020099-03-0020	
transfected construct (*Homo sapiens*)	siRNA: siGENOME TUBB3 siRNA #3	Horizon Discovery	MQ-020099-03-0020	

transfected construct (*Homo sapiens*)	siRNA: siGENOME TUBB3 siRNA #4	Horizon Discovery	MQ-020099-03-0020	
transfected construct (*Homo sapiens*)	siRNA: siGENOME TUBB3 siRNA #5	Horizon Discovery	MQ-020099-03-0020	
transfected construct (*Homo sapiens*)	siRNA: siGENOME TUBB3 siRNA #19	Horizon Discovery	MQ-020099-03-0020	
transfected construct (*Homo sapiens*)	siRNA: siGENOME SMARTpool MEX3A siRNA	Horizon Discovery	MQ-022355-01-0020	
transfected construct (*Homo sapiens*)	siRNA: siGENOME MEX3A siRNA #1	Horizon Discovery	MQ-022355-01-0020	
transfected construct (*Homo sapiens*)	siRNA: siGENOME MEX3A siRNA #2	Horizon Discovery	MQ-022355-01-0020	
transfected construct (*Homo sapiens*)	siRNA: siGENOME MEX3A siRNA #3	Horizon Discovery	MQ-022355-01-0020	
transfected construct (*Homo sapiens*)	siRNA: siGENOME MEX3A siRNA #4	Horizon Discovery	MQ-022355-01-0020	
Antibody	Anti-SOX11 (Rabbit monoclonal)	Abcam	Clone EPR8191(2) Cat# ab78078	WB: (1:1000)
Antibody	Anti-SOX11 (Mouse monoclonal)	EBioscience	SOX11-C1 Cat# 50-9773-82	IF: (1:200)
Antibody	Anti-TUBB3 (Mouse monoclonal)	Abcam	Clone 2G10 Cat# ab78078	WB: (1:1000) IF: (1:100)

Antibody	Anti-MEX3A (Rabbit polyclonal)	Abcam	Cat# ab79046	WB: (1:1000) IF: (1:100)
Antibody	Anti-CD24 (Mouse monoclonal)	Creative Biolabs	Cat# SWA11	WB: (1:1000)
Antibody	Anti-VIMENTIN (Rabbit monoclonal)	Abcam	Cat# Ab92547 (EPR3776)	WB: (1:1000) IF: (1:400)
Antibody	Anti-RCOR2 (CoREST2) (Rabbit polyclonal)	Abcam	Cat# ab37113	WB: (1:1000)
Antibody	Anti-N-CADHERIN (Rabbit monoclonal)	Cell Signaling	Cat# 13116	WB: (1:1000) IF: (1:50)
Antibody	Anti-E-CADHERIN (Mouse monoclonal)	BD Bioscience	Clone 36 Cat# 610182	WB: (1:1000) IF: (1:200)
Antibody	Anti-EPCAM (Rabbit monoclonal)	Cell Signaling	Clone D1B3 Cat# 2626	WB: (1:1000)
Antibody	Anti-RRM2 (R2) (Mouse monoclonal)	Santa Cruz	Clone A-5 Cat# sc-398294	WB: (1:1000)
Antibody	Anti-SURVIVIN (Rabbit monoclonal)	Cell Signaling	Clone 71G4B7 Cat# 2808	WB: (1:1000)
Antibody	Anti-GAPDH (Rabbit monoclonal)	Cell Signaling	Clone D16H11 Cat# 5174	WB: (1:5000)
Antibody	Anti-LAMINB1 (Rabbit polyclonal)	Abcam	Cat# ab16048	WB: (1:1000)
Antibody	Anti-β-TUBULIN (Mouse monoclonal)	Sigma	Cat# T4026	WB: (1:1000)

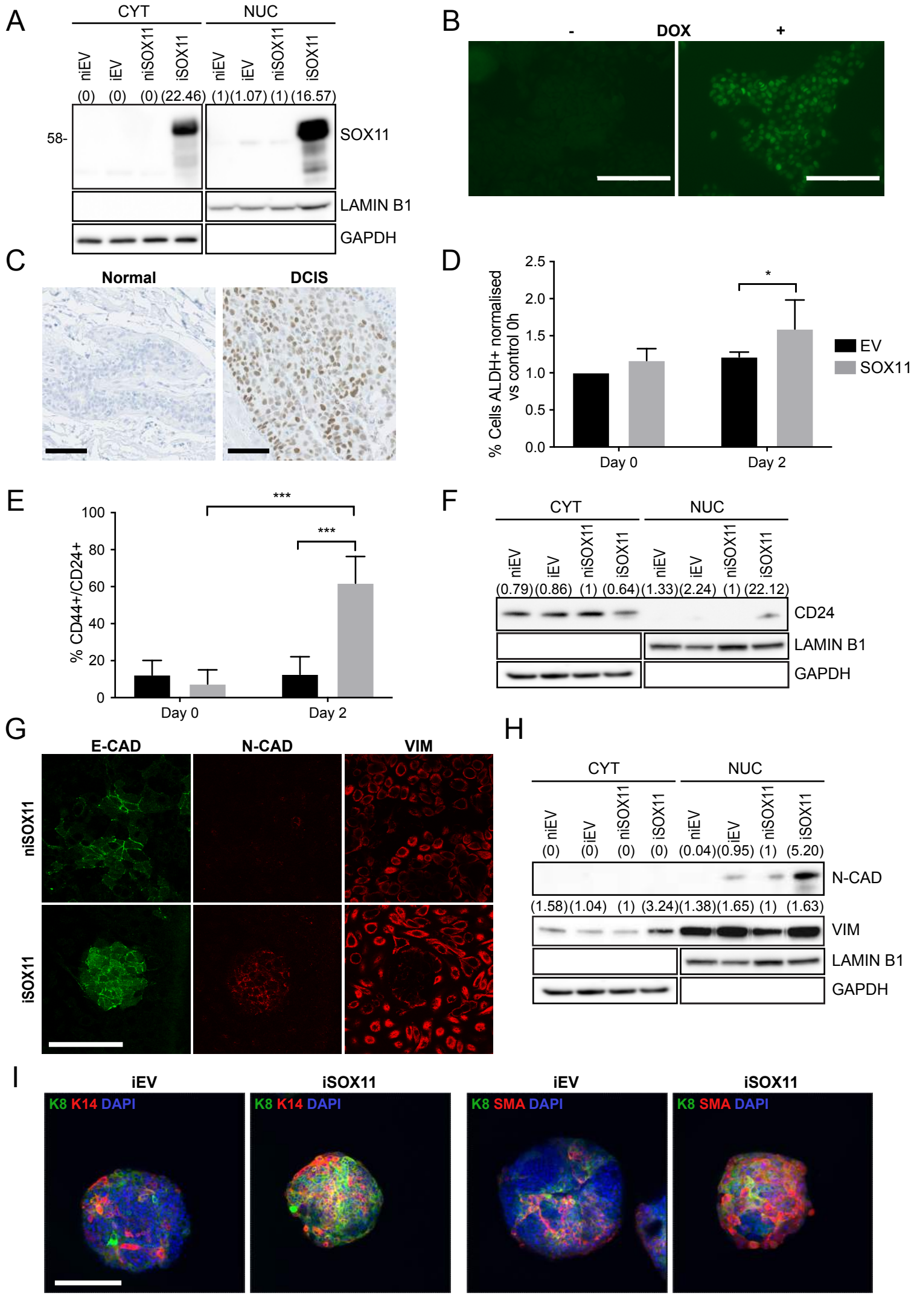
Antibody	Anti-β-ACTIN (Mouse monoclonal)	Cell Signaling	Clone 8H10D10 Cat# 3700	WB: (1:1000)
Antibody	Anti-β-K5 (Rabbit polyclonal)	Biolegend	Cat# PRB-160P	IF: (1:200)
Antibody	Anti-β-K14 (Rabbit polyclonal)	Biolegend	Cat# PRB-155P	IF: (1:200)
Antibody	Anti-β-CD24 (Mouse monoclonal)	Invitrogen/ Thermofisher	Cat# SN3	IF: (1:50)
Antibody	Anti-β-SMA (Rabbit monoclonal)	Invitrogen/ Thermofisher	EPR5368 Cat# Ab202509	IF: (1:50)
Antibody	Anti-CD24–PE–Cy7 (Mouse monoclonal)	BD Bioscience	Cat# 561646	Flow cytometry: (1:50)
Antibody	Anti-CD44–APC (Mouse monoclonal)	BD Bioscience	Cat# 559942	Flow cytometry: (1:50)
recombinant DNA reagent	pInducer21 (Plasmid)	Addgene	Cat# 46948	
recombinant DNA reagent	pInducer13 (Plasmid)	Addgene	Cat# 46936	
recombinant DNA reagent	Firefly Luciferase 2 lentiviral particles	Amsbio	Cat# LVP325	
recombinant DNA reagent	pLV-mCherry (Plasmid)	Addgene	Cat# 36084	
sequence-based reagent	TaqMan probe SOX11	Thermofisher Scientific	Hs00846583_s1	
sequence-based reagent	TaqMan probe TUBB3	Thermofisher Scientific	Hs00801390_s1	

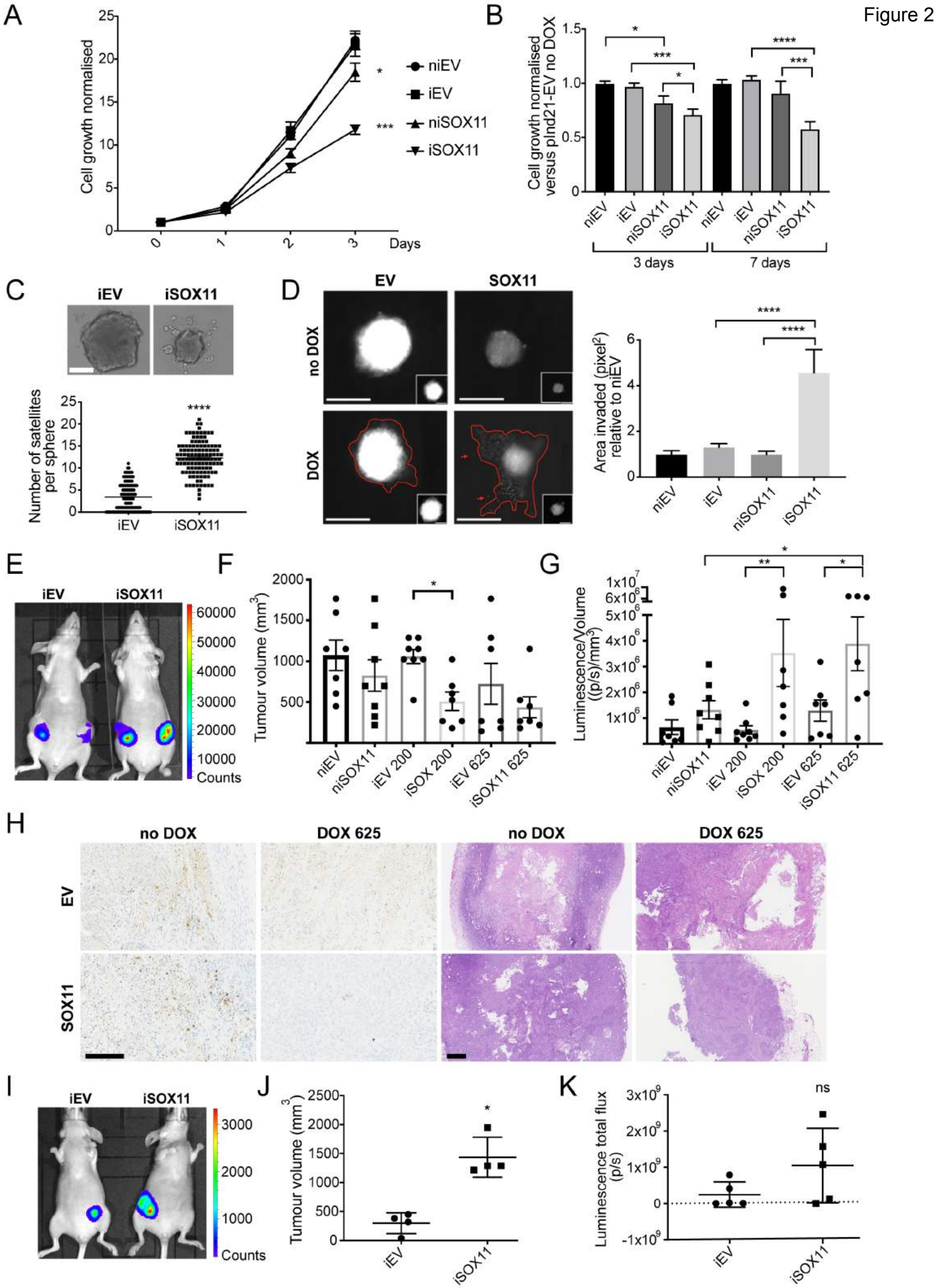
sequence-based reagent	TaqMan probe MEX3A	Thermofisher Scientific	Hs00863536_m1	
sequence-based reagent	TaqMan probe GPC2	Thermofisher Scientific	Hs00415099_m1	
sequence-based reagent	TaqMan probe MAPK4	Thermofisher Scientific	Hs00969401_m1	
sequence-based reagent	TaqMan probe LBH	Thermofisher Scientific	Hs00368853_m1	
sequence-based reagent	TaqMan probe SERPINA3	Thermofisher Scientific	Hs00153674_m1	
sequence-based reagent	TaqMan probe OLFM2	Thermofisher Scientific	Hs01017934_m1	
sequence-based reagent	TaqMan probe N-CADHERIN	Thermofisher Scientific	Hs00983056_m1	
sequence-based reagent	TaqMan probe ST6GALNAC5	Thermofisher Scientific	Hs05018504_s1	
sequence-based reagent	TaqMan probe GAPDH	Thermofisher Scientific	Hs02786624_g1	
peptide, recombinant protein	Animal-Free Recombinant Human EGF	Peprotech	Cat# AF-100-15	
commercial assay or kit	RNAeasyPlus Micro kit	Qiagen	Cat# 74034	

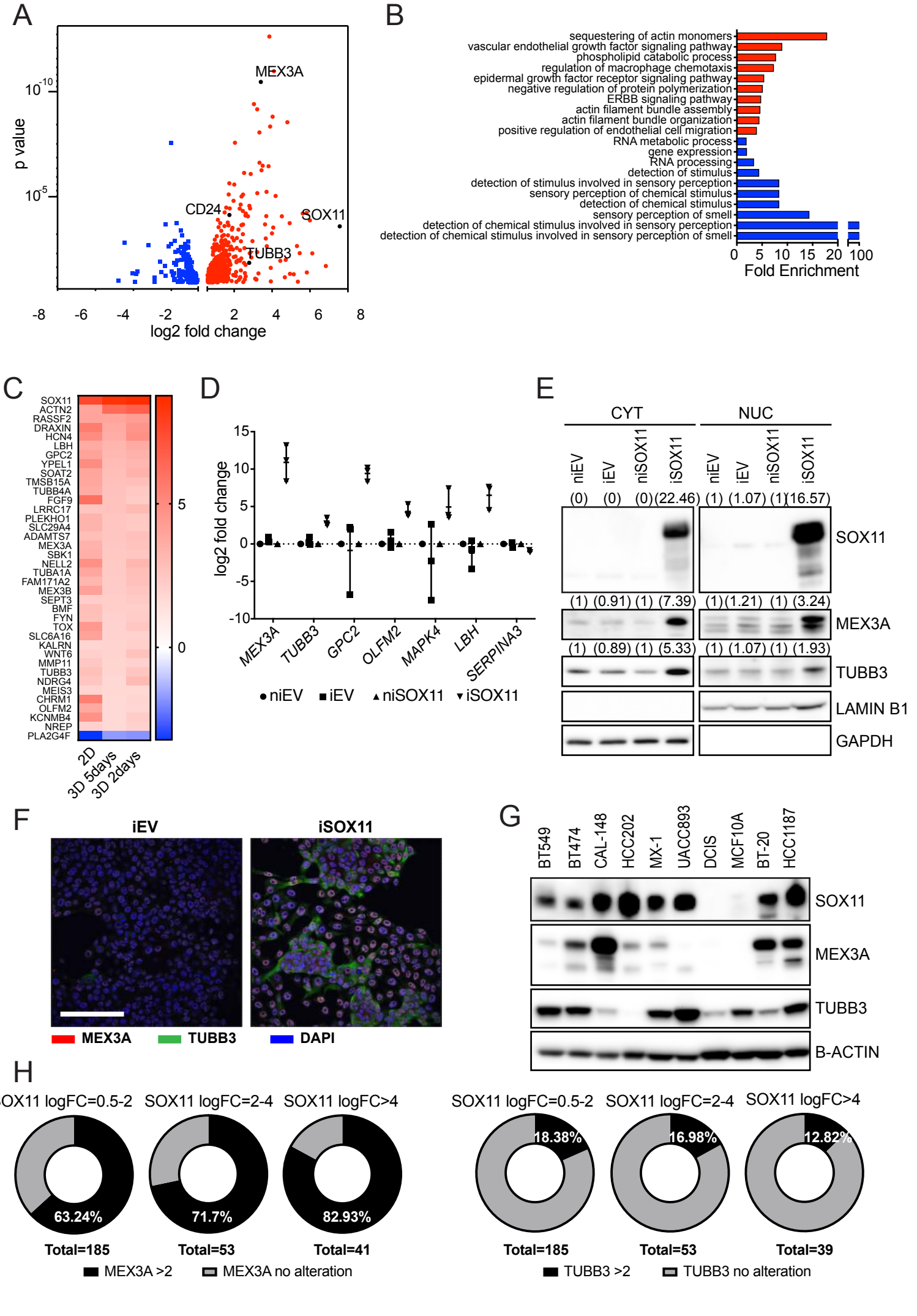
commercial assay or kit	RNAClean and concentrator-5	Zymo Research	Cat# R1013	
commercial assay or kit	Agilent RNA Pico kit	Agilent Technologies	Cat# 5067-1513	
commercial assay or kit	QuantiTect Reverse Transcription kit	Qiagen	Cat# 205311	
commercial assay or kit	TaqMan™ Gene Expression Master Mix	Thermofisher Scientific	Cat# 4369016	
commercial assay or kit	Aldefluor assay	StemCell Technologies	Cat# 01700	
commercial assay or kit	PKH26 Dye Solution	SIGMA	Cat# MINI26	
commercial assay or kit	Tumour dissociation kit, human	Miltenyi	Cat# 130-095-929	
chemical compound, drug	Doxycycline hyclate	Sigma	Cat# D9891	
chemical compound, drug	NeuroCult SM1 without vitamin A	StemCell Technologies	Cat # 05731	
chemical compound, drug	Methylcellulose	R&D Systems	Cat # HSC002	
chemical compound, drug	Cell titer-Glo	Promega	Cat # G7572	
chemical compound, drug	Collagen I, High Concentration, Rat Tail	Corning	Cat # 354249	

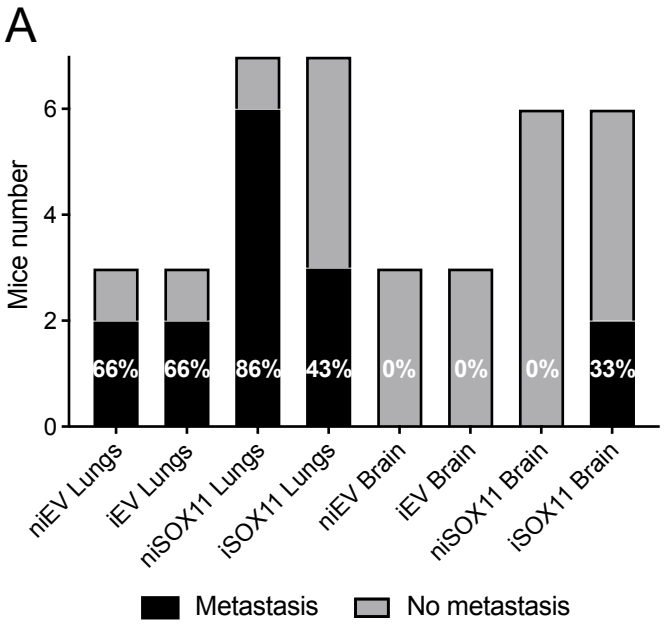
chemical compound, drug	XenoLight D-Luciferin Potassium Salt	Perkin Elmer	Cat # 122799	
chemical compound, drug	Lipofectamine 2000	Invitrogen	11668019	
chemical compound, drug	Lipofectamine RNAiMAX	Invitrogen	13778075	
software, algorithm	PRISM	Graphpad		
software, algorithm	BD FACS Diva software	BD Bioscience		
software, algorithm	Image J	National Institutes of Health (NIH)		
other	EVOS FL microscope	Thermofisher Scientific		
other	Confocal microscope	Leica	Model TCS-SP2	
other	Celigo cytometer	Nexcelom		
other	96-well ultra-low-attachment plates	Corning	Cat # 7007	
other	Luminescence plate reader	Perkin Elmer	Victor X5 58	
other	FACS	BD Bioscience	FACSAriaIII	
other	Flow cytometer	BD Bioscience	BD FACS LSRII	
other	Stereotaxic frame	Stoelting		
other	IVIS Lumina imaging systems	Perkin Elmer		

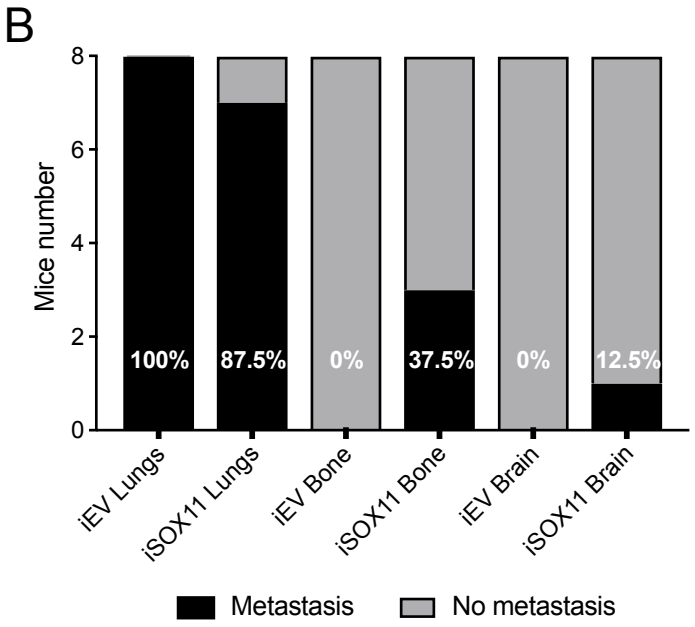
other	gentleMACS™ Octo Dissociator with Heaters	Perkin Elmer		
other	DAPI	Sigma		IF: 1:5000 FC/FACS: 1:5000

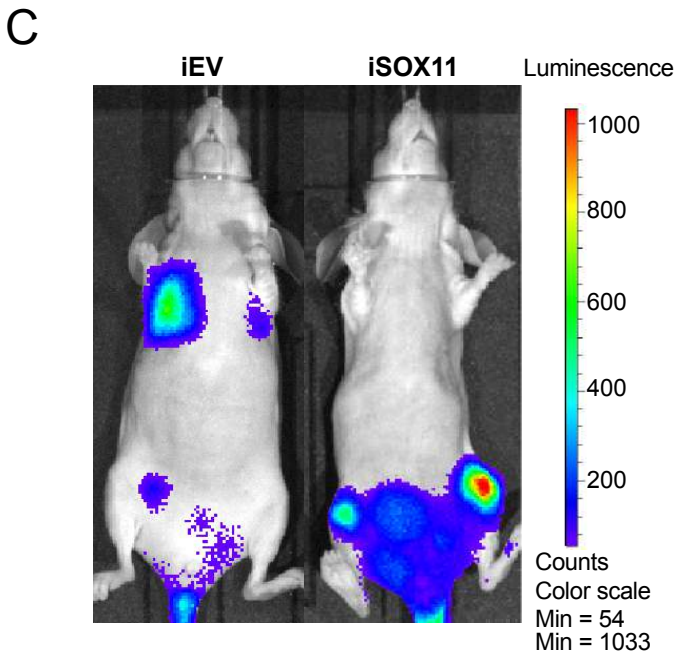

A
CYT
NUC
niEV
iEV
niSOX11
iSOX11
(0) (0) (0) (22.46) (1) (1.07) (1) (16.57)
58-
SOX11
LAMIN B1
GAPDH
B
-
DOX
+
C
Normal
DCIS
D
% Cells ALDH+ normalised vs control 0h
2.5
2.0
1.5
1.0
0.5
0.0
*
Day 0
Day 2
EV
SOX11
E
% CD44+/CD24+
100
80
60
40
20
0
***
***
Day 0
Day 2
F
CYT
NUC
niEV
iEV
niSOX11
iSOX11
(0.79) (0.86) (1) (0.64) (1.33) (2.24) (1) (22.12)
CD24
LAMIN B1
GAPDH
G
E-CAD
N-CAD
VIM
niSOX11
iSOX11
H
CYT
NUC
niEV
iEV
niSOX11
iSOX11
(0) (0) (0) (0) (0.04) (0.95) (1) (5.20)
N-CAD
(1.58) (1.04) (1) (3.24) (1.38) (1.65) (1) (1.63)
VIM
LAMIN B1
GAPDH
I
iEV
iSOX11
iEV
iSOX11
K8 K14 DAPI
K8 K14 DAPI
K8 SMA DAPI
K8 SMA DAPI

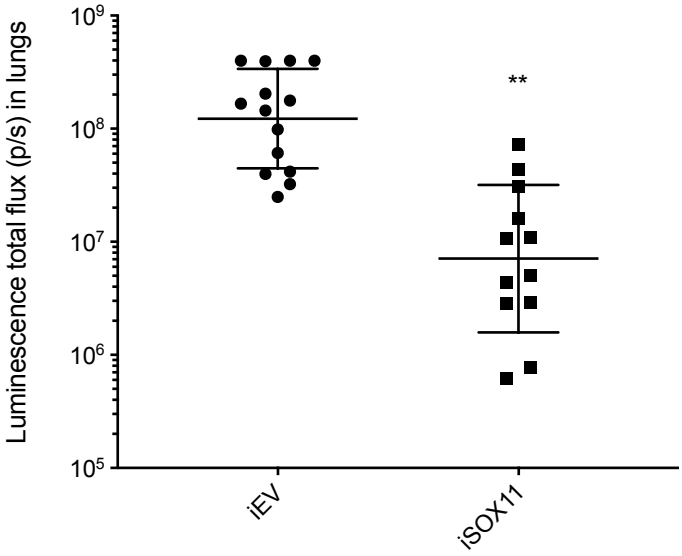

A
Cell growth normalised
Days
niEV
iEV
niSOX11
iSOX11
B
Cell growth normalised versus pInd21-EV no DOX
niEV
iEV
niSOX11
iSOX11
3 days
7 days
C
iEV
iSOX11
Number of satellites per sphere
D
EV
SOX11
no DOX
DOX
Area invaded (pixel²) relative to niEV
E
iEV
iSOX11
Counts
F
Tumour volume (mm³)
niEV
niSOX11
iEV 200
iSOX 200
iEV 625
iSOX11 625
G
Luminescence/Volume ((p/s)/mm³)
niEV
niSOX11
iEV 200
iSOX 200
iEV 625
iSOX11 625
H
no DOX
DOX 625
no DOX
DOX 625
EV
SOX11
I
iEV
iSOX11
Counts
J
Tumour volume (mm³)
iEV
iSOX11
K
Luminescence total flux (p/s)
ns
iEV
iSOX11

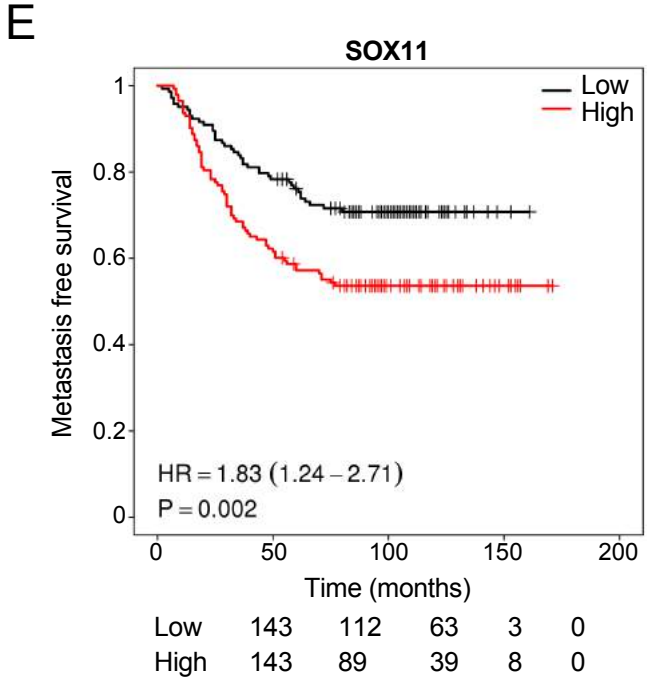

A
p value
log2 fold change
MEX3A
CD24
SOX11
TUBB3
B
sequestering of actin monomers
vascular endothelial growth factor signaling pathway
phospholipid catabolic process
regulation of macrophage chemotaxis
epidermal growth factor receptor signaling pathway
negative regulation of protein polymerization
ERBB signaling pathway
actin filament bundle assembly
actin filament bundle organization
positive regulation of endothelial cell migration
RNA metabolic process
gene expression
RNA processing
detection of stimulus
detection of stimulus involved in sensory perception
sensory perception of chemical stimulus
detection of chemical stimulus
sensory perception of smell
detection of chemical stimulus involved in sensory perception
detection of chemical stimulus involved in sensory perception of smell
Fold Enrichment
C
SOX11
ACTN2
RASSF2
DRAXIN
HCN4
LBH
GPC2
YPEL1
SOAT2
TMSB15A
TUBB4A
FGF9
LRRC17
PLEKHO1
SLC29A4
ADAMTS7
MEX3A
SBK1
NELL2
TUBA1A
FAM171A2
MEX3B
SEPT3
BMF
FYN
TOX
SLC6A16
KALRN
WNT6
MMP11
TUBB3
NDRG4
MEIS3
CHRM1
OLFM2
KCNMB4
NREP
PLA2G4F
2D
3D 5days
3D 2days
D
log2 fold change
MEX3A
TUBB3
GPC2
OLFM2
MAPK4
LBH
SERPINA3
niEV
iEV
niSOX11
iSOX11
E
CYT
NUC
niEV
iEV
niSOX11
iSOX11
(0) (0) (0) (22.46) (1) (1.07) (1) (16.57)
SOX11
(1) (0.91) (1) (7.39) (1) (1.21) (1) (3.24)
MEX3A
(1) (0.89) (1) (5.33) (1) (1.07) (1) (1.93)
TUBB3
LAMIN B1
GAPDH
F
iEV
iSOX11
MEX3A
TUBB3
DAPI
G
BT549
BT474
CAL-148
HCC202
MX-1
UACC893
DCIS
MCF10A
BT-20
HCC1187
SOX11
MEX3A
TUBB3
B-ACTIN
H
SOX11 logFC=0.5-2
SOX11 logFC=2-4
SOX11 logFC>4
63.24%
71.7%
82.93%
Total=185
Total=53
Total=41
MEX3A >2
MEX3A no alteration
SOX11 logFC=0.5-2
SOX11 logFC=2-4
SOX11 logFC>4
18.38%
16.98%
12.82%
Total=185
Total=53
Total=39
TUBB3 >2
TUBB3 no alteration

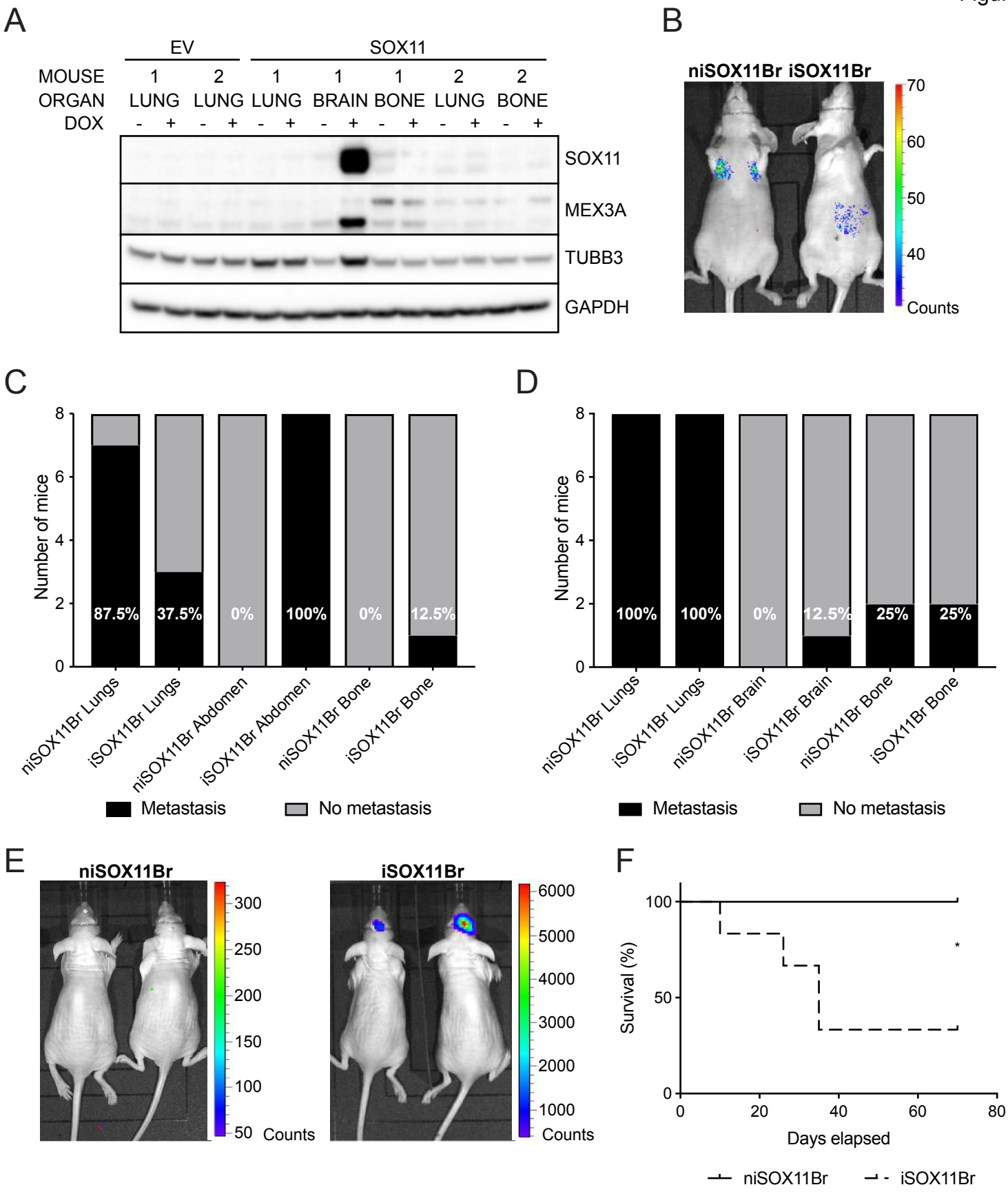

A
Mice number
0
2
4
6
66%
66%
86%
43%
0%
0%
0%
33%
niEV Lungs
iEV Lungs
niSOX11 Lungs
iSOX11 Lungs
niEV Brain
iEV Brain
niSOX11 Brain
iSOX11 Brain
Metastasis
No metastasis

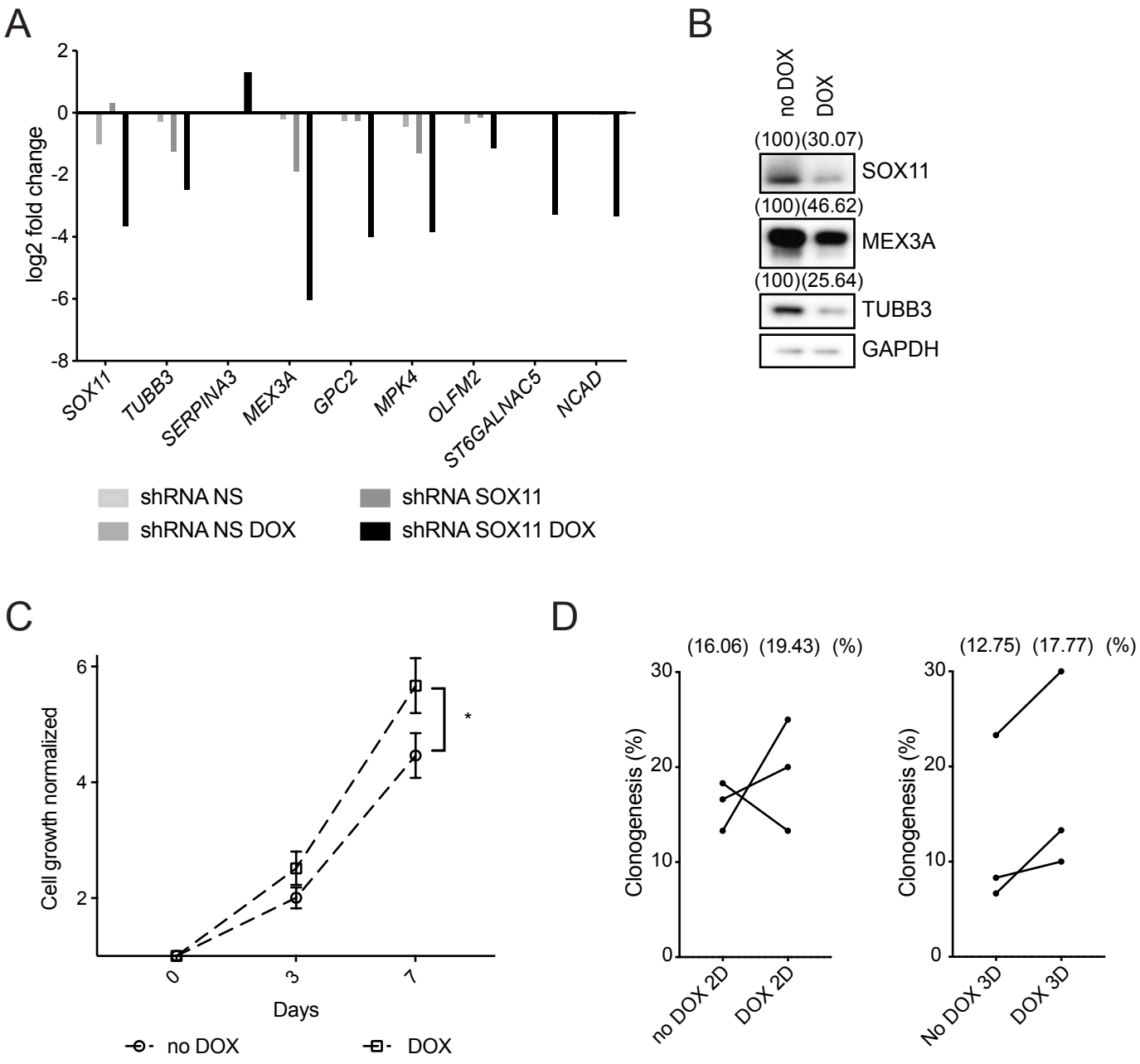

B
Mice number
0
2
4
6
8
100%
87.5%
0%
37.5%
0%
12.5%
iEV Lungs
iSOX11 Lungs
iEV Bone
iSOX11 Bone
iEV Brain
iSOX11 Brain
Metastasis
No metastasis

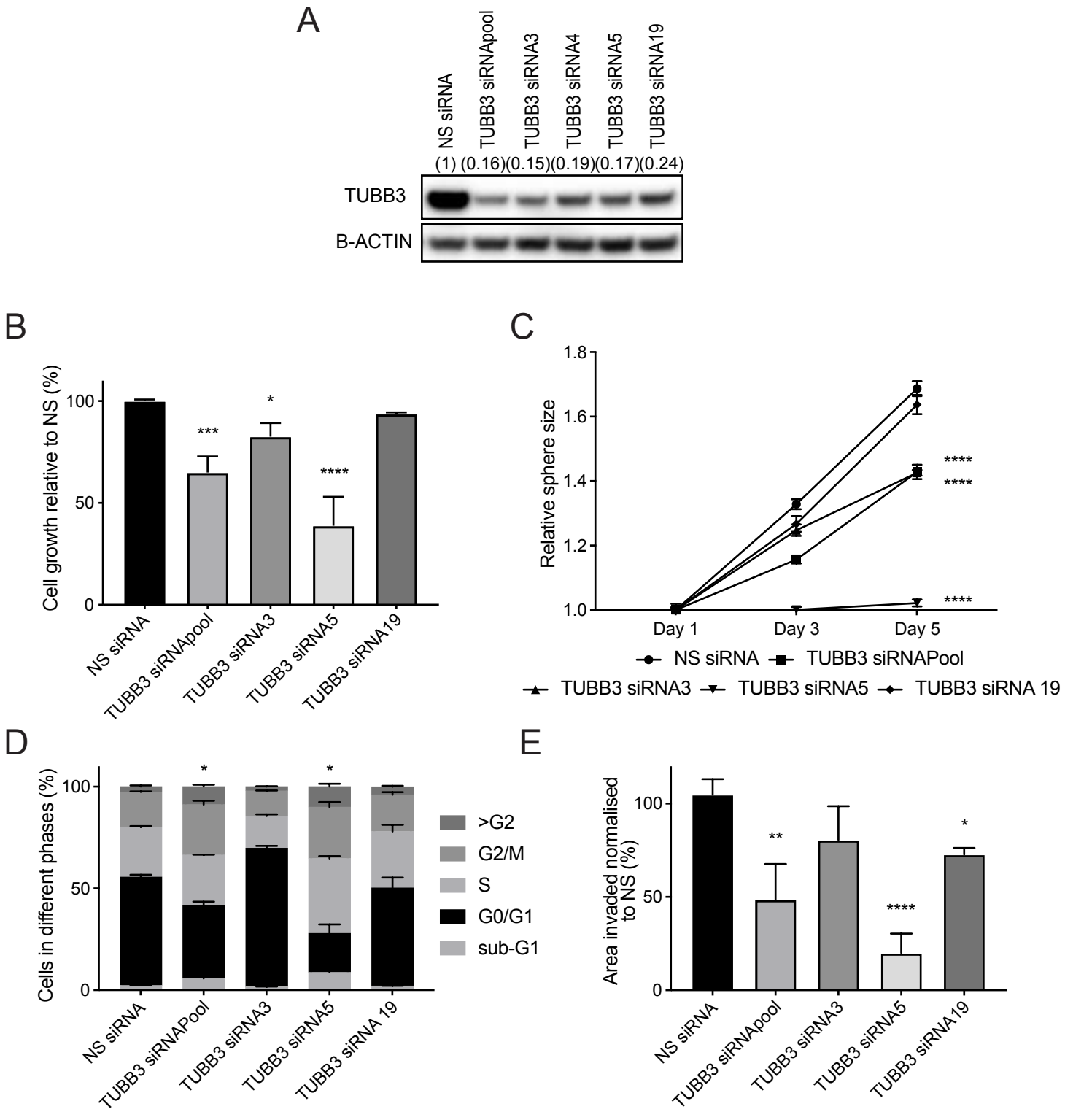

C
iEV
iSOX11
Luminescence
1000
800
600
400
200
Counts
Color scale
Min = 54
Min = 1033

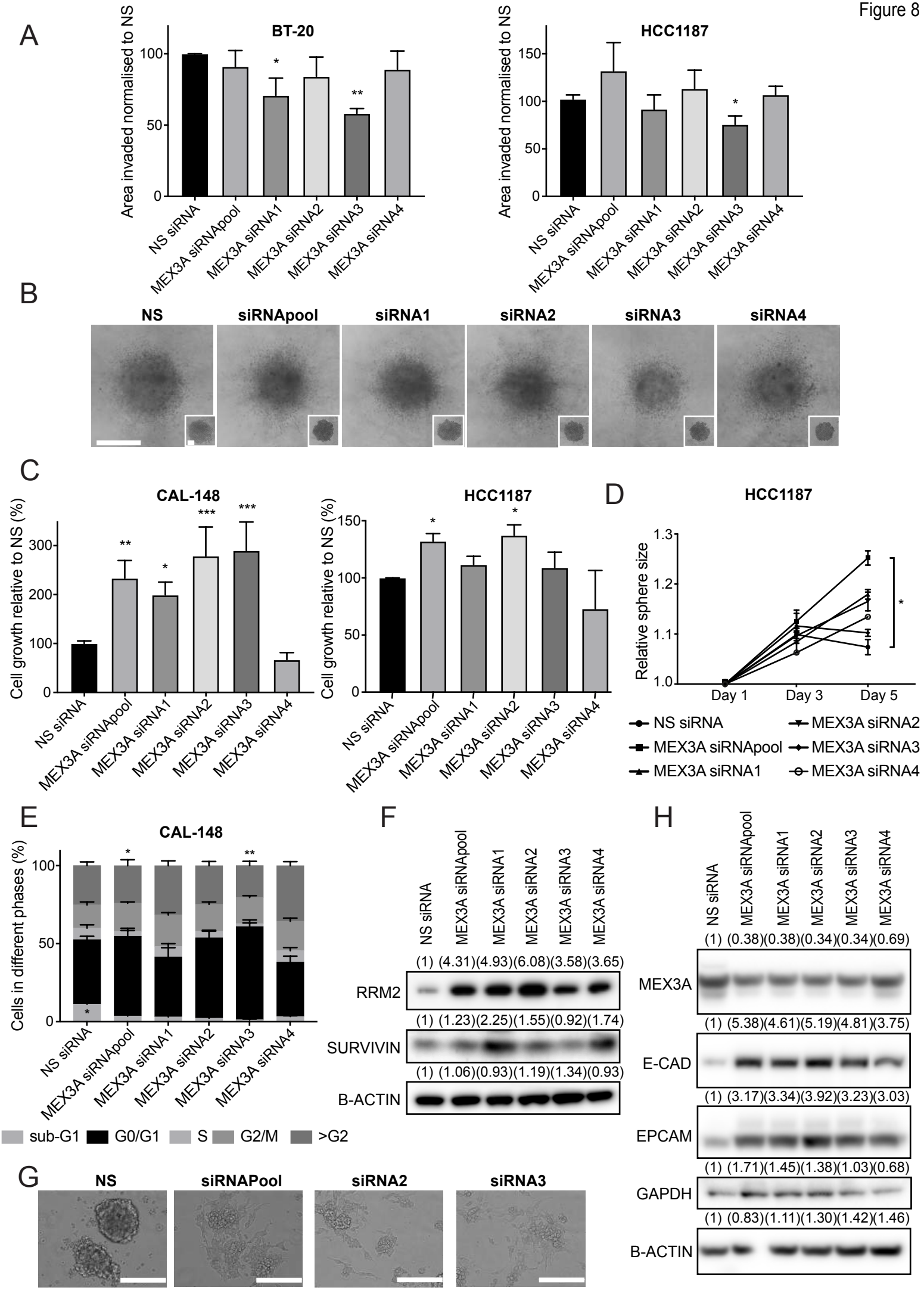

D
Luminescence total flux (p/s) in lungs
10^9
10^8
10^7
10^6
10^5
**
iEV
iSOX11

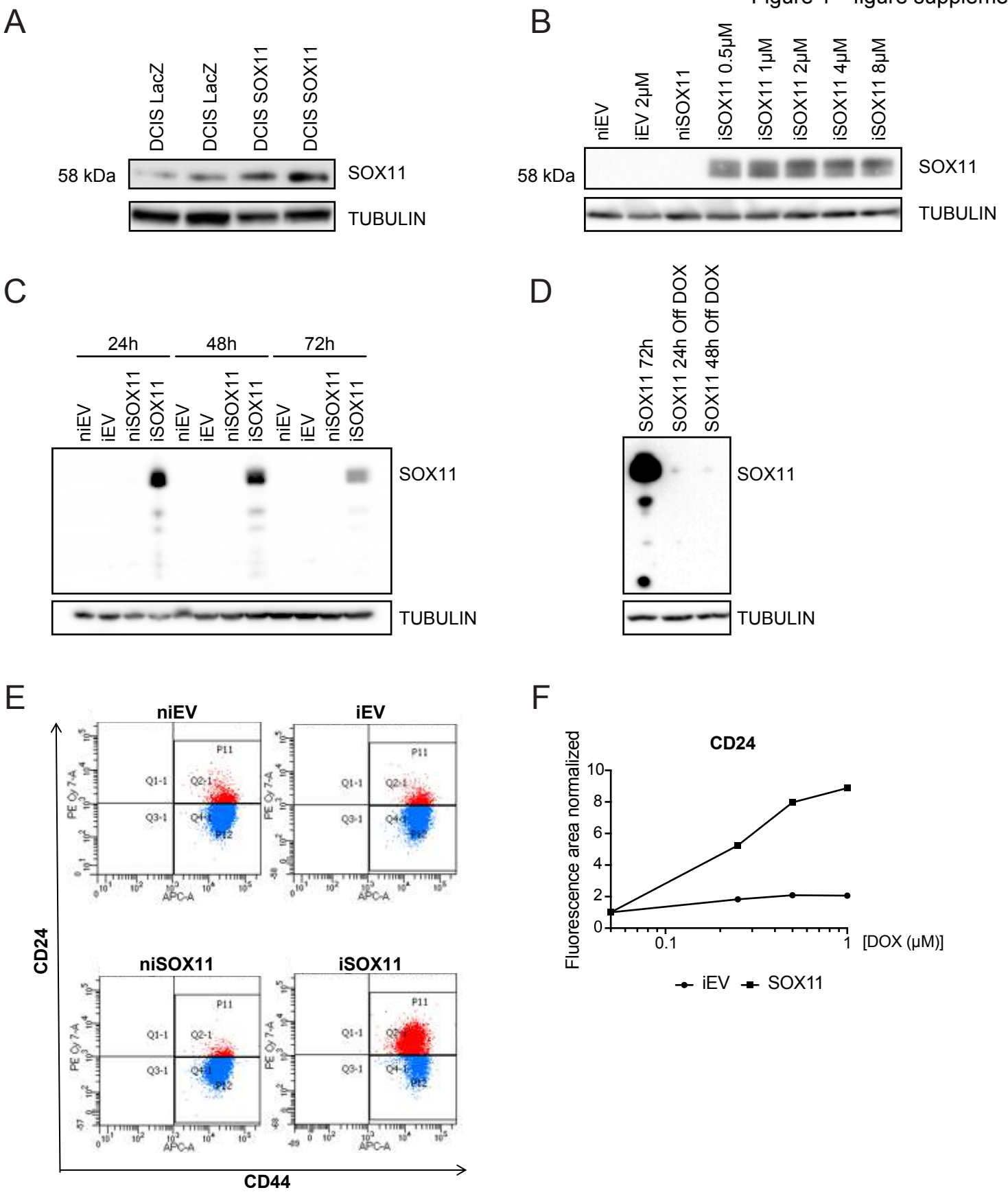

E
SOX11
Low
High
Metastasis free survival
1
0.8
0.6
0.4
0.2
0
HR = 1.83 (1.24 – 2.71)
P = 0.002
0
50
100
150
200
Time (months)
Low 143 112 63 3 0
High 143 89 39 8 0


A
EV
SOX11
MOUSE 1 2 1 1 1 2 2
ORGAN LUNG LUNG LUNG BRAIN BONE LUNG BONE
DOX - + - + - + - + - + - + - +
SOX11
MEX3A
TUBB3
GAPDH
B
niSOX11Br iSOX11Br
70
60
50
40
Counts
C
Number of mice
87.5%
37.5%
0%
100%
0%
12.5%
niSOX11Br Lungs
iSOX11Br Lungs
niSOX11Br Abdomen
iSOX11Br Abdomen
niSOX11Br Bone
iSOX11Br Bone
Metastasis
No metastasis
D
Number of mice
100%
100%
0%
12.5%
25%
25%
niSOX11Br Lungs
iSOX11Br Lungs
niSOX11Br Brain
iSOX11Br Brain
niSOX11Br Bone
iSOX11Br Bone
Metastasis
No metastasis
E
niSOX11Br
iSOX11Br
300
250
200
150
100
50
Counts
6000
5000
4000
3000
2000
1000
Counts
F
Survival (%)
Days elapsed
niSOX11Br
iSOX11Br

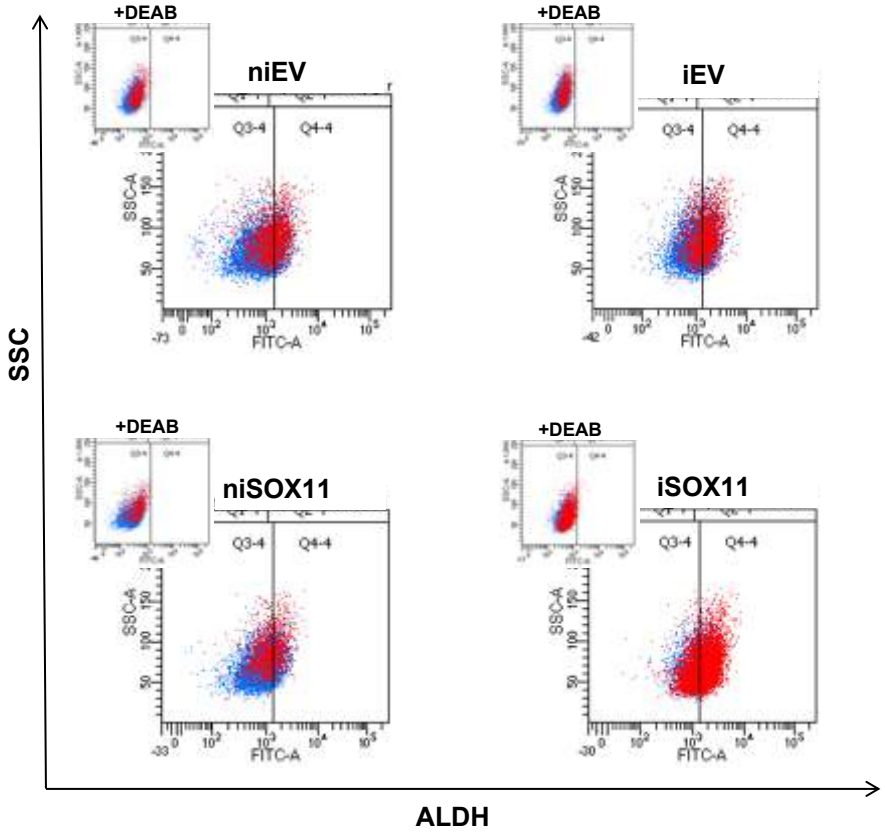

A
log2 fold change
2
0
-2
-4
-6
-8
SOX11
TUBB3
SERPINA3
MEX3A
GPC2
MPK4
OLFM2
ST6GALNAC5
NCAD
shRNA NS
shRNA NS DOX
shRNA SOX11
shRNA SOX11 DOX
B
no DOX
DOX
(100)(30.07)
SOX11
(100)(46.62)
MEX3A
(100)(25.64)
TUBB3
GAPDH
C
Cell growth normalized
6
4
2
0
3
7
Days
*
no DOX
DOX
D
(16.06) (19.43) (%)
Clonogenesis (%)
30
20
10
0
no DOX 2D
DOX 2D
(12.75) (17.77) (%)
Clonogenesis (%)
30
20
10
0
No DOX 3D
DOX 3D


A
NS siRNA
TUBB3 siRNApool
TUBB3 siRNA3
TUBB3 siRNA4
TUBB3 siRNA5
TUBB3 siRNA19
(1) (0.16)(0.15)(0.19)(0.17)(0.24)
TUBB3
B-ACTIN
B
Cell growth relative to NS (%)
100
50
0
***
*
****
NS siRNA
TUBB3 siRNApool
TUBB3 siRNA3
TUBB3 siRNA5
TUBB3 siRNA19
C
Relative sphere size
1.8
1.6
1.4
1.2
1.0
Day 1
Day 3
Day 5
****
****
****
NS siRNA
TUBB3 siRNAPool
TUBB3 siRNA3
TUBB3 siRNA5
TUBB3 siRNA 19
D
Cells in different phases (%)
100
50
0
*
*
>G2
G2/M
S
G0/G1
sub-G1
NS siRNA
TUBB3 siRNAPool
TUBB3 siRNA3
TUBB3 siRNA5
TUBB3 siRNA 19
E
Area invaded normalised to NS (%)
100
50
0
**
****
*
NS siRNA
TUBB3 siRNApool
TUBB3 siRNA3
TUBB3 siRNA5
TUBB3 siRNA19
F
NS
siRNApool
siRNA3
siRNA5
siRNA19

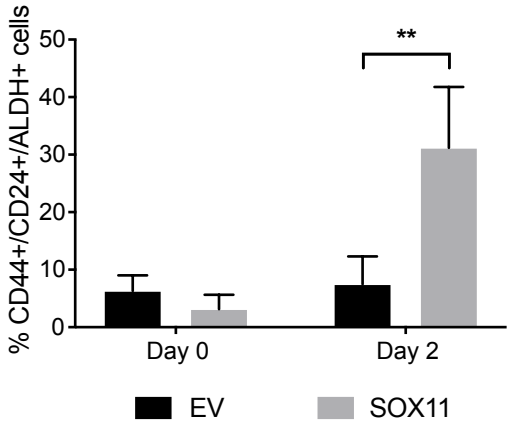

A
BT-20
Area invaded normalised to NS
NS siRNA
MEX3A siRNApool
MEX3A siRNA1
MEX3A siRNA2
MEX3A siRNA3
MEX3A siRNA4
HCC1187
B
NS
siRNApool
siRNA1
siRNA2
siRNA3
siRNA4
C
CAL-148
Cell growth relative to NS (%)
HCC1187
D
HCC1187
Relative sphere size
Day 1
Day 3
Day 5
E
CAL-148
Cells in different phases (%)
sub-G1
G0/G1
S
G2/M
>G2
F
RRM2
SURVIVIN
B-ACTIN
H
MEX3A
E-CAD
EPCAM
GAPDH
B-ACTIN
G
NS
siRNAPool
siRNA2
siRNA3

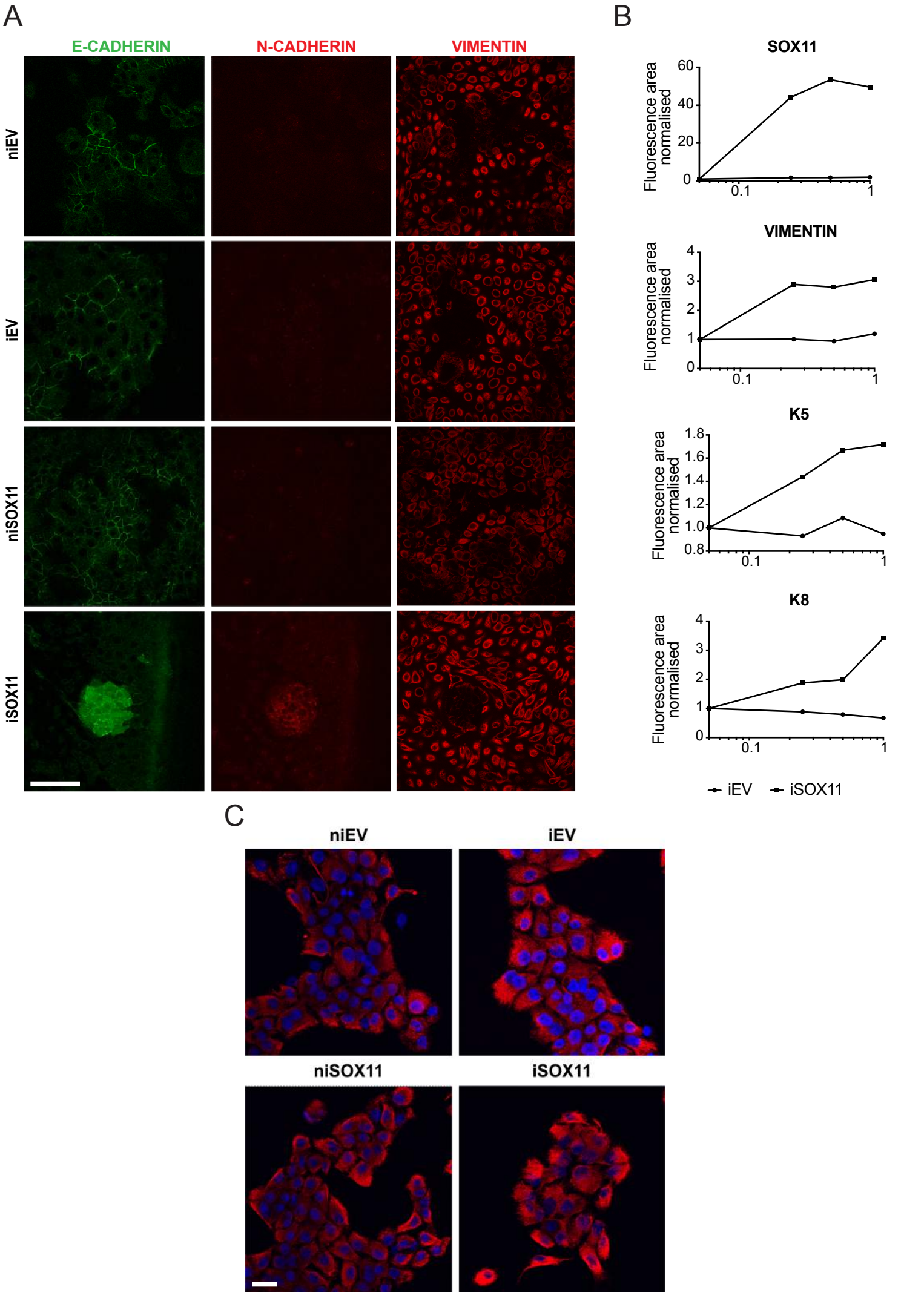

A
DCIS LacZ
DCIS LacZ
DCIS SOX11
DCIS SOX11
58 kDa
SOX11
TUBULIN
B
niEV
iEV 2μM
niSOX11
iSOX11 0.5μM
iSOX11 1μM
iSOX11 2μM
iSOX11 4μM
iSOX11 8μM
58 kDa
SOX11
TUBULIN
C
24h
48h
72h
niEV
iEV
niSOX11
iSOX11
SOX11
TUBULIN
D
SOX11 72h
SOX11 24h Off DOX
SOX11 48h Off DOX
SOX11
TUBULIN
E
niEV
iEV
niSOX11
iSOX11
CD24
CD44
P11
Q1-1
Q2-1
Q3-1
Q4-1
P12
PE Cy 7-A
APC-A
F
CD24
Fluorescence area normalized
10
8
6
4
2
0
0.1
1
[DOX (μM)]
iEV
SOX11

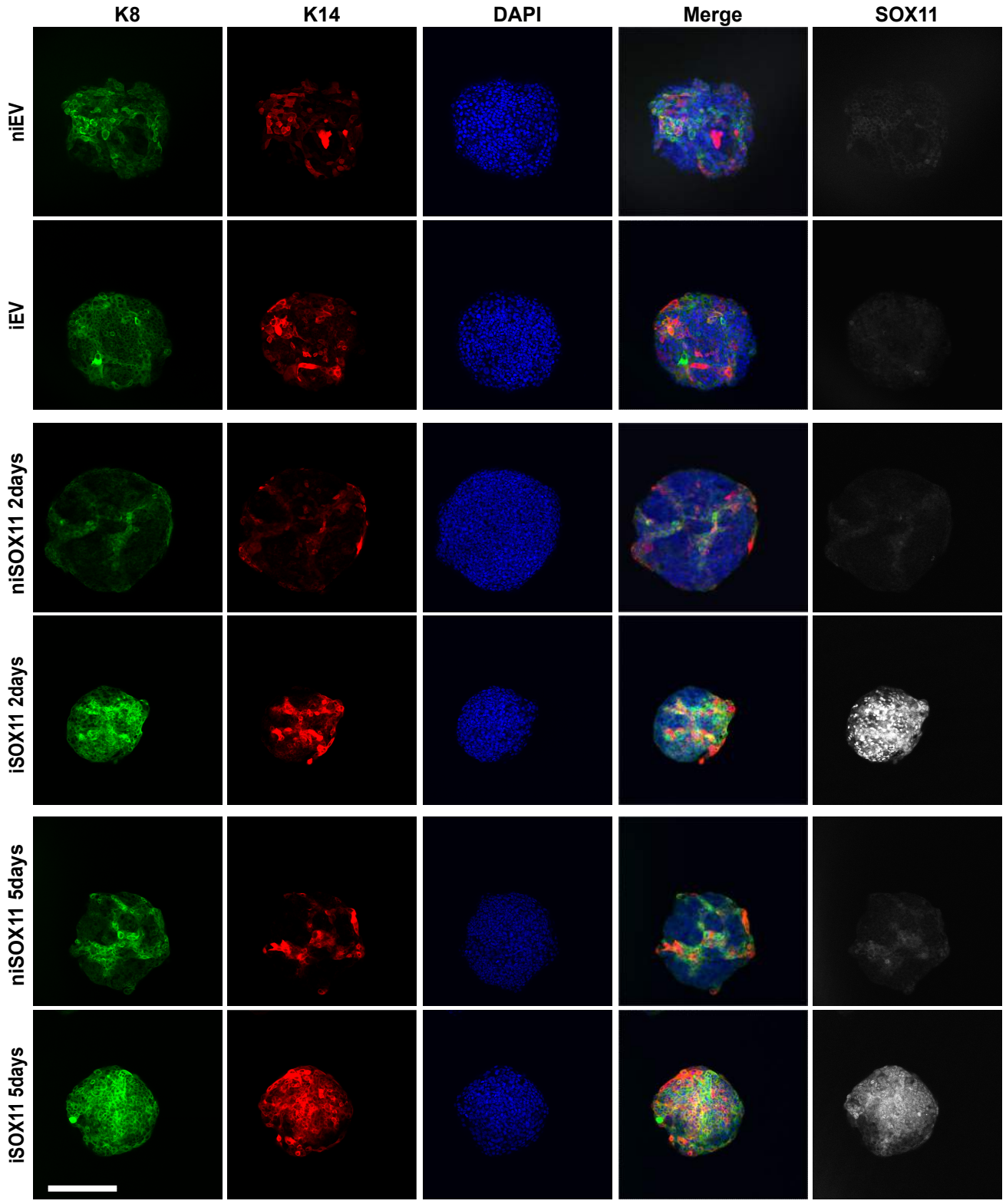
A


+DEAB
niEV
Q3-4
Q4-4
SSC-A
FITC-A
+DEAB
iEV
Q3-4
Q4-4
SSC-A
FITC-A
+DEAB
niSOX11
Q3-4
Q4-4
SSC-A
FITC-A
+DEAB
iSOX11
Q3-4
Q4-4
SSC-A
FITC-A
SSC
ALDH

B


% CD44+/CD24+/ALDH+ cells
50
40
30
20
10
0
**
Day 0
Day 2
EV
SOX11


A
E-CADHERIN
N-CADHERIN
VIMENTIN
niEV
iEV
niSOX11
iSOX11
B
SOX11
VIMENTIN
K5
K8
Fluorescence area normalised
iEV
iSOX11
C
niEV
iEV
niSOX11
iSOX11


K8
K14
DAPI
Merge
SOX11
niEV
iEV
niSOX11 2days
iSOX11 2days
niSOX11 5days
iSOX11 5days

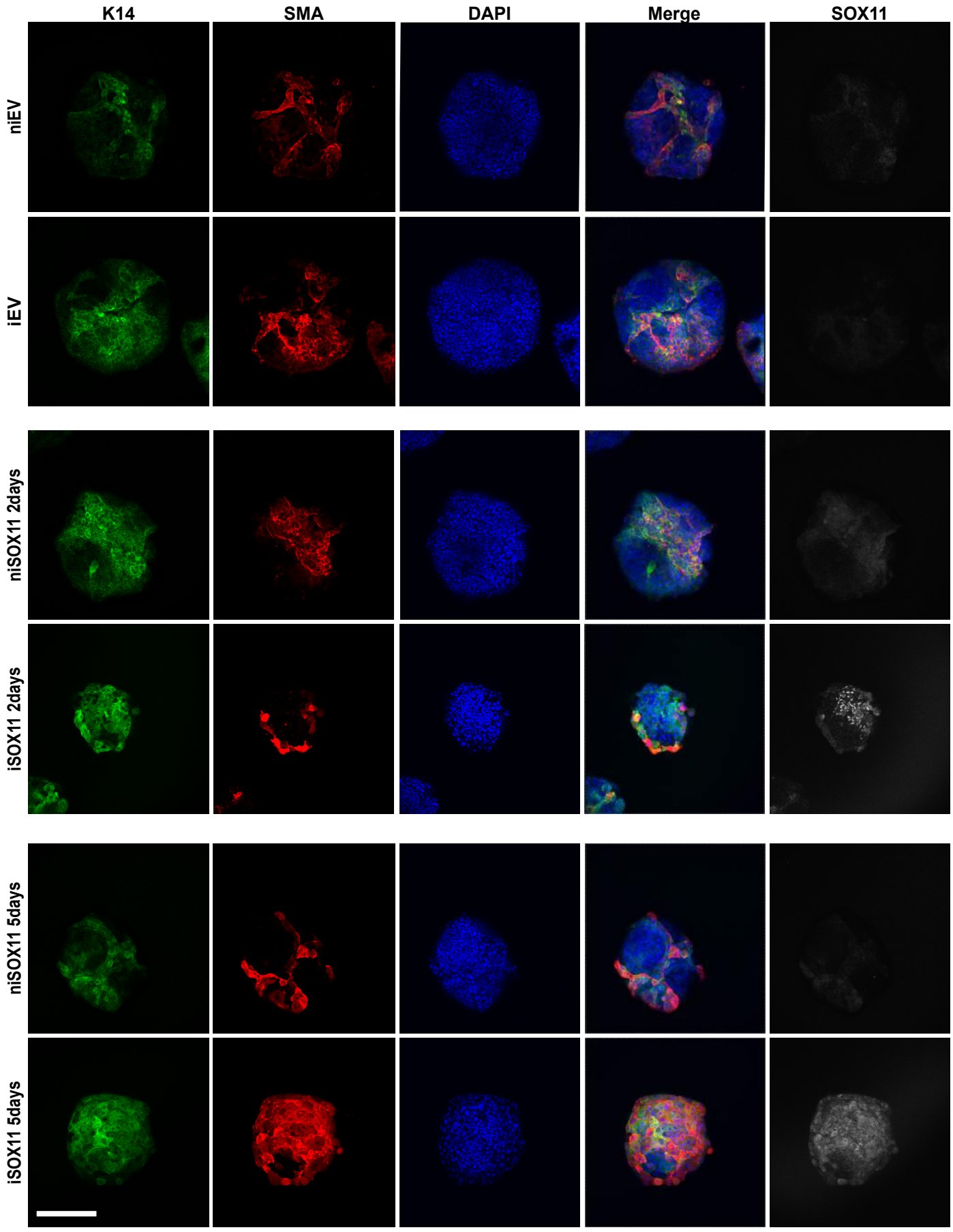
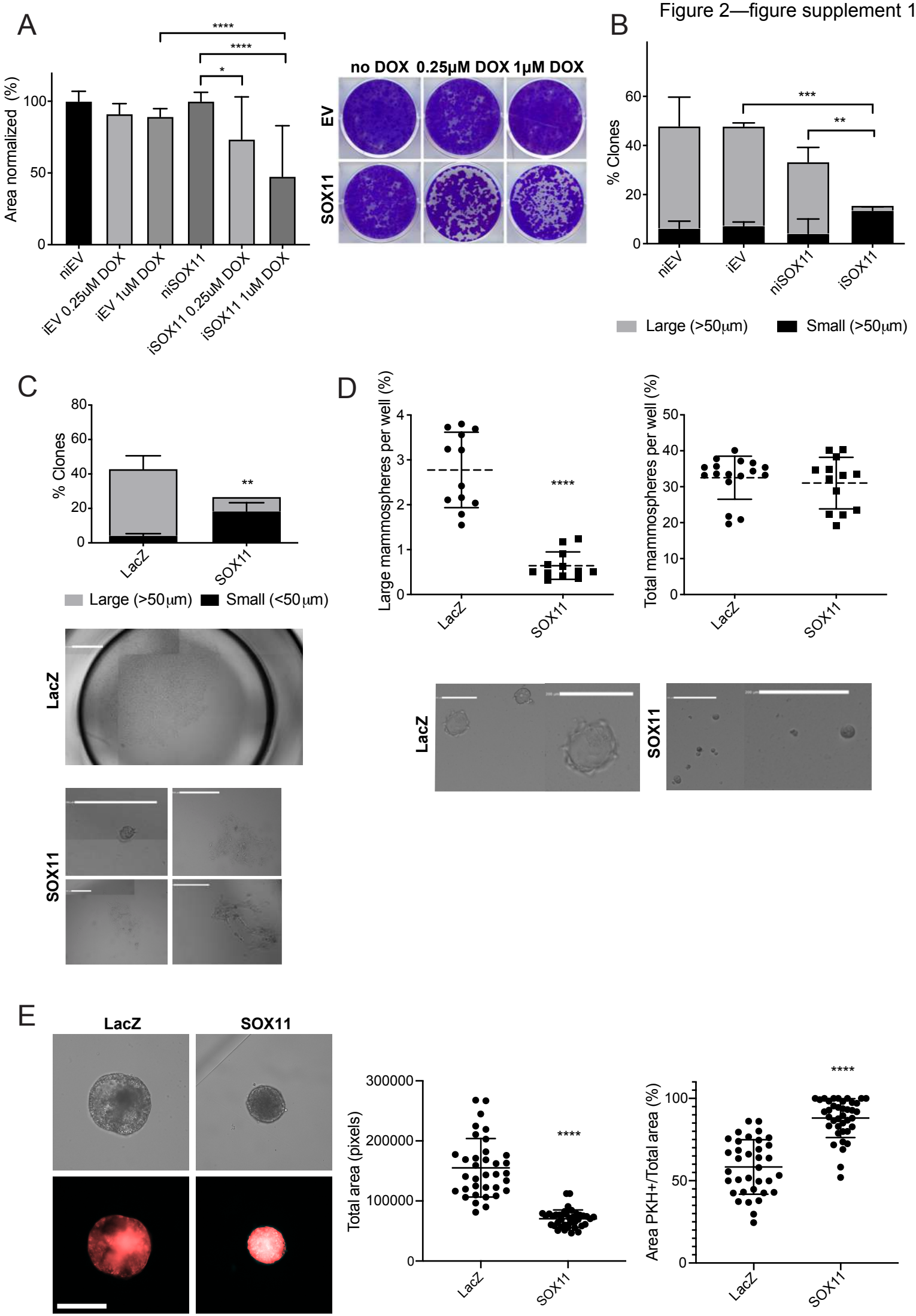
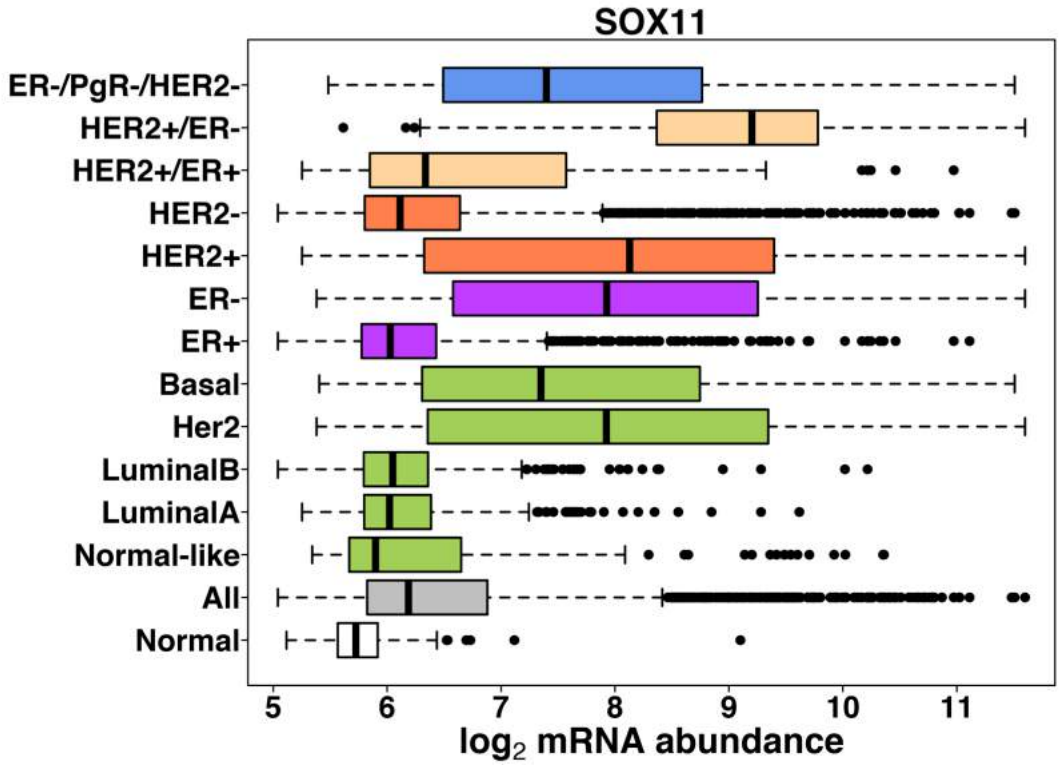

K14
SMA
DAPI
Merge
SOX11
niEV
iEV
niSOX11 2days
iSOX11 2days
niSOX11 5days
iSOX11 5days

Figure 2—figure supplement 1




SOX11
ER-/PgR-/HER2-
HER2+/ER-
HER2+/ER+
HER2-
HER2+
ER-
ER+
Basal
Her2
LuminalB
LuminalA
Normal-like
All
Normal
5
6
7
8
9
10
11
log2 mRNA abundance

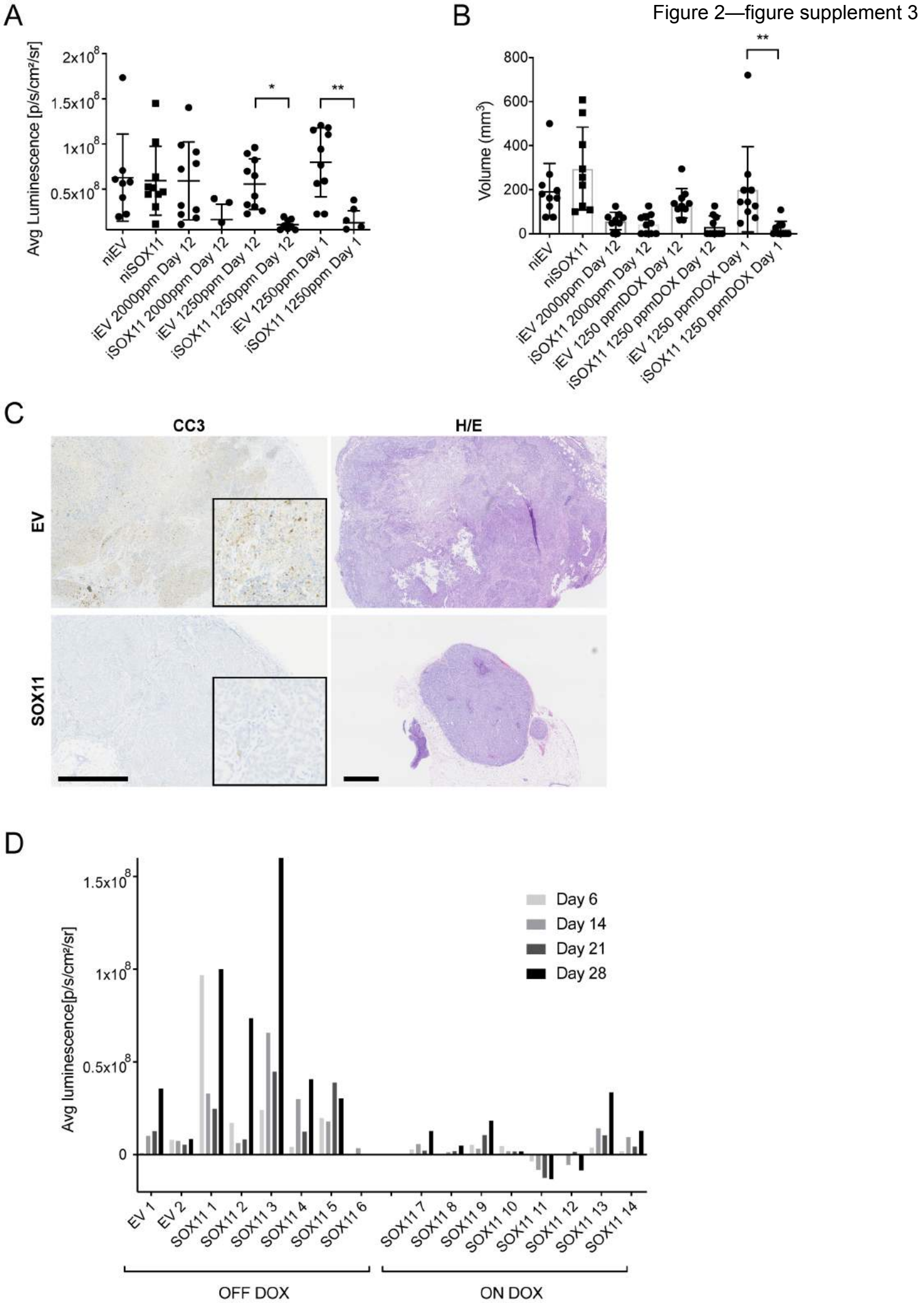
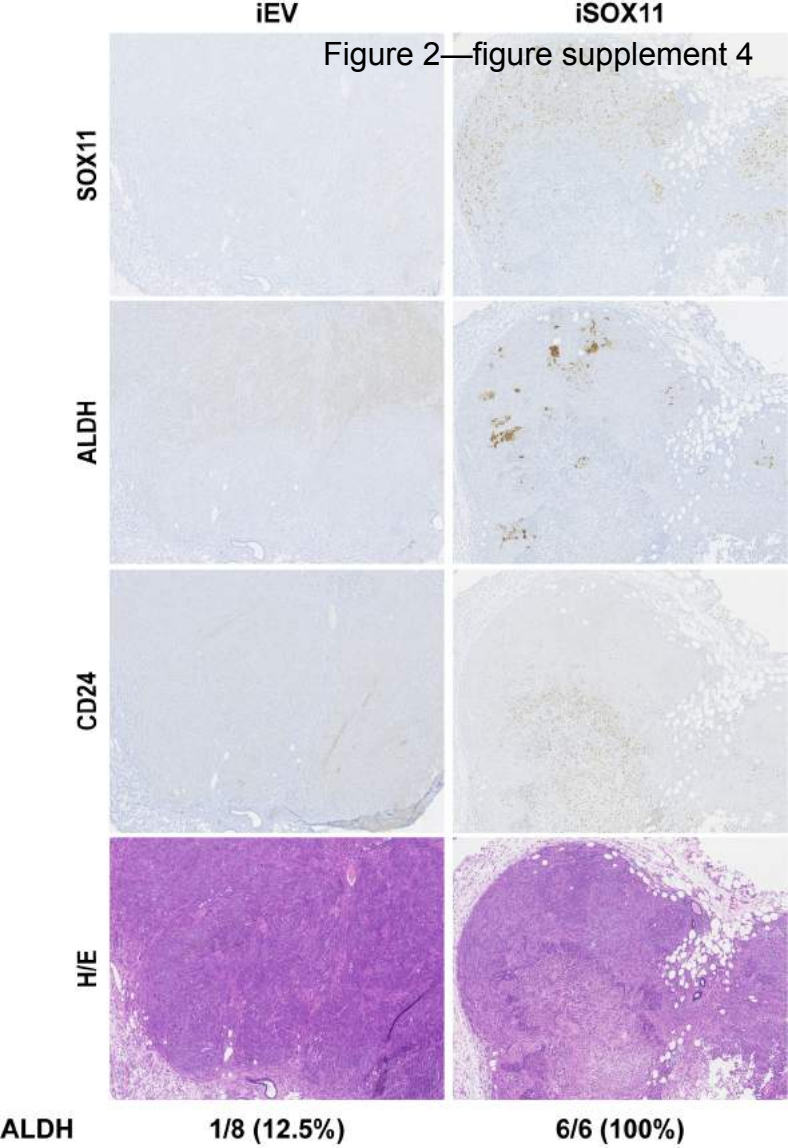
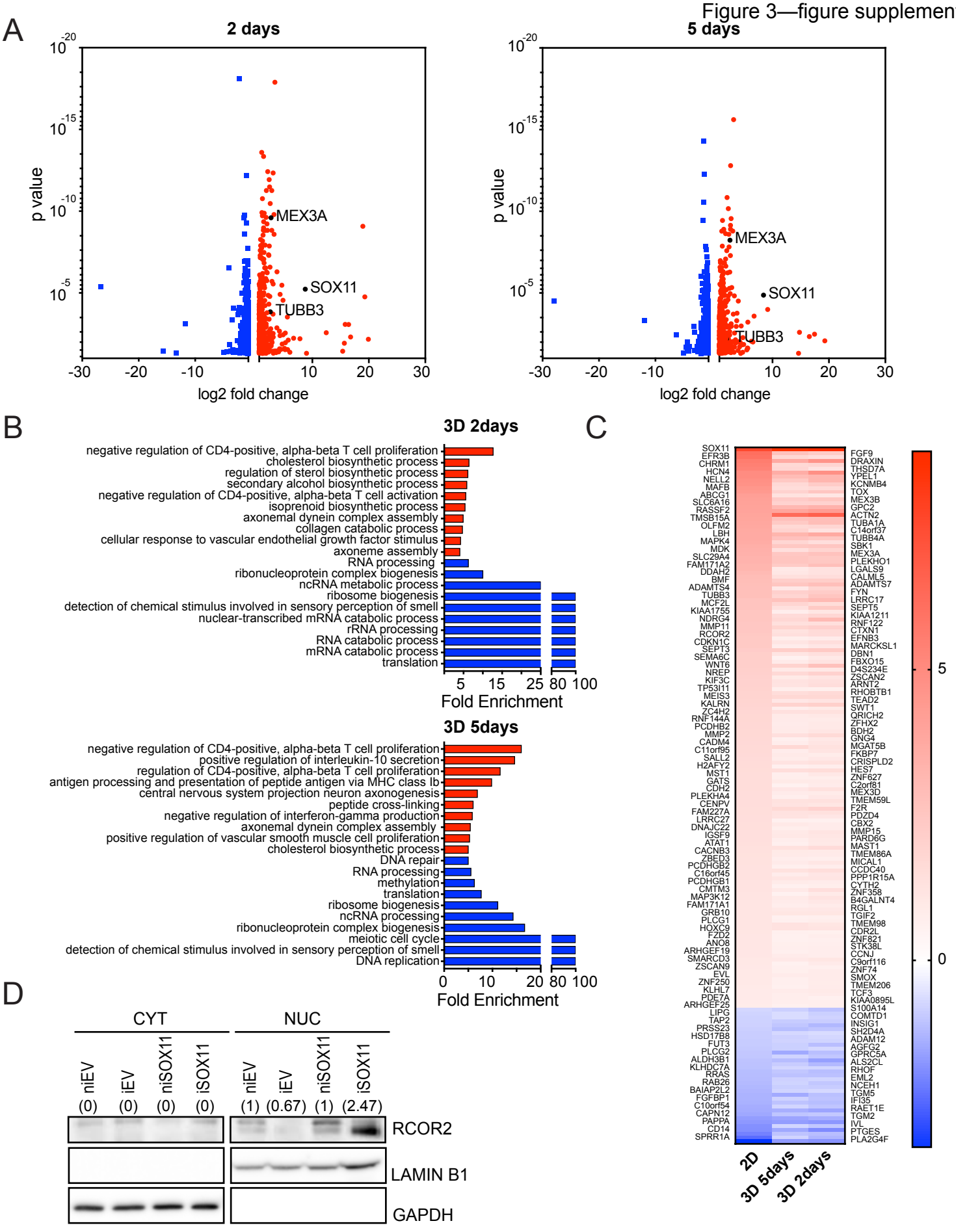
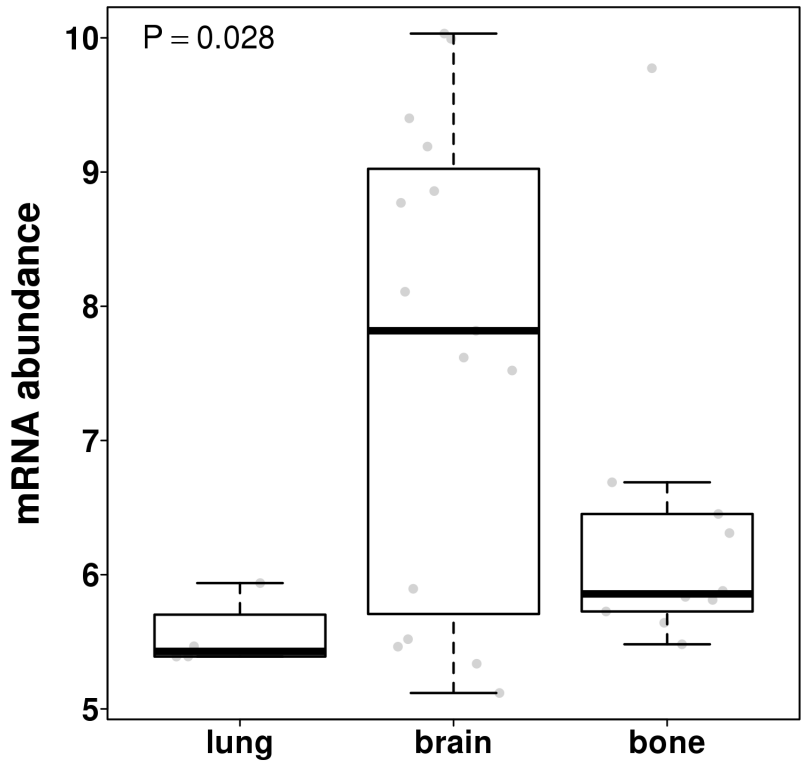

A
Avg Luminescence [p/s/cm²/sr]
2x10^8
1.5x10^8
1x10^8
0.5x10^8
*
**
niEV
niSOX11
iEV 2000ppm Day 12
iSOX11 2000ppm Day 12
iEV 1250ppm Day 12
iSOX11 1250ppm Day 12
iEV 1250ppm Day 1
iSOX11 1250ppm Day 1
B
Volume (mm³)
800
600
400
200
0
**
niEV
niSOX11
iEV 2000ppm Day 12
iSOX11 2000ppm Day 12
iEV 1250 ppmDOX Day 12
iSOX11 1250 ppmDOX Day 12
iEV 1250 ppmDOX Day 1
iSOX11 1250 ppmDOX Day 1
C
CC3
H/E
EV
SOX11
D
Avg luminescence[p/s/cm²/sr]
1.5x10^8
1x10^8
0.5x10^8
0
Day 6
Day 14
Day 21
Day 28
EV 1
EV 2
SOX11 1
SOX11 2
SOX11 3
SOX11 4
SOX11 5
SOX11 6
SOX11 7
SOX11 8
SOX11 9
SOX11 10
SOX11 11
SOX11 12
SOX11 13
SOX11 14
OFF DOX
ON DOX

Figure 2—figure supplement 4




A
2 days
5 days
p value
log2 fold change
MEX3A
SOX11
TUBB3
B
3D 2days
negative regulation of CD4-positive, alpha-beta T cell proliferation
cholesterol biosynthetic process
regulation of sterol biosynthetic process
secondary alcohol biosynthetic process
negative regulation of CD4-positive, alpha-beta T cell activation
isoprenoid biosynthetic process
axonemal dynein complex assembly
collagen catabolic process
cellular response to vascular endothelial growth factor stimulus
axoneme assembly
RNA processing
ribonucleoprotein complex biogenesis
ncRNA metabolic process
ribosome biogenesis
detection of chemical stimulus involved in sensory perception of smell
nuclear-transcribed mRNA catabolic process
rRNA processing
RNA catabolic process
mRNA catabolic process
translation
Fold Enrichment
3D 5days
positive regulation of interleukin-10 secretion
antigen processing and presentation of peptide antigen via MHC class Ib
central nervous system projection neuron axonogenesis
peptide cross-linking
negative regulation of interferon-gamma production
positive regulation of vascular smooth muscle cell proliferation
DNA repair
methylation
meiotic cell cycle
DNA replication
C
2D
3D 5days
3D 2days
D
CYT
NUC
niEV
iEV
niSOX11
iSOX11
(0)
(0.67)
(1)
(2.47)
RCOR2
LAMIN B1
GAPDH
E
SOX11 logFC=0.5-2
SOX11 logFC=2-4
SOX11 logFC>4
60%
94.12%
49.1%
30%
23.53%
Total=110
Total=30
Total=17
MEX3A >2
MEX3A no alteration
TUBB3 >2
TUBB3 no alteration

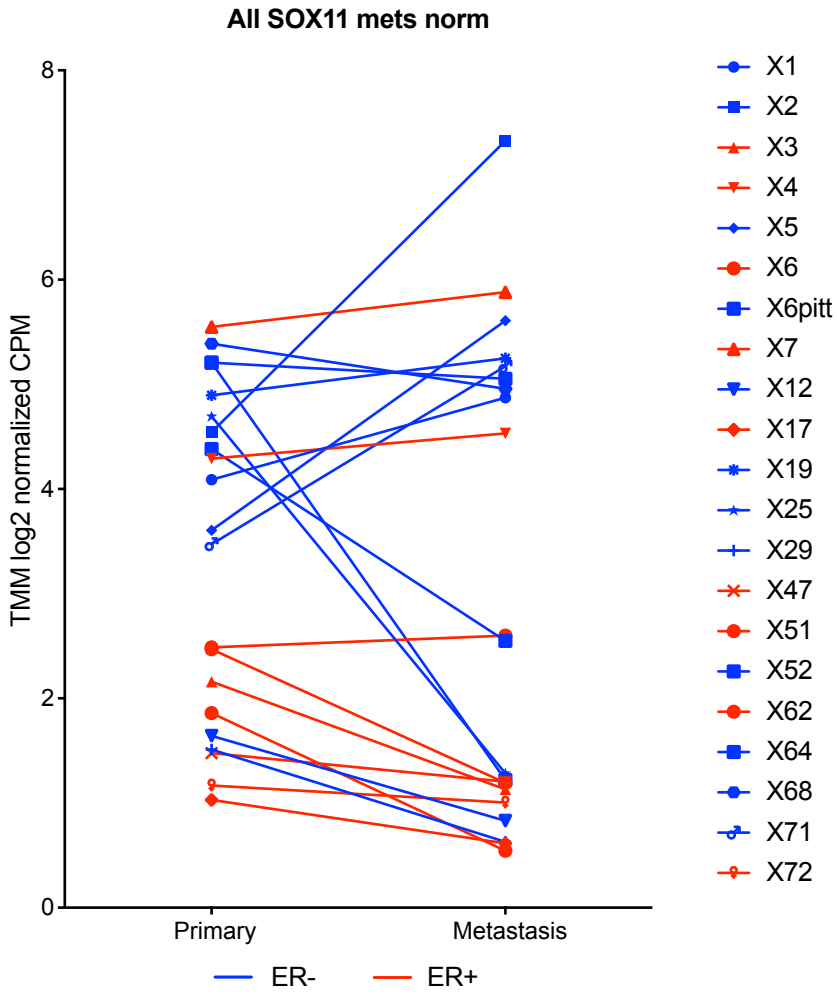
A



B

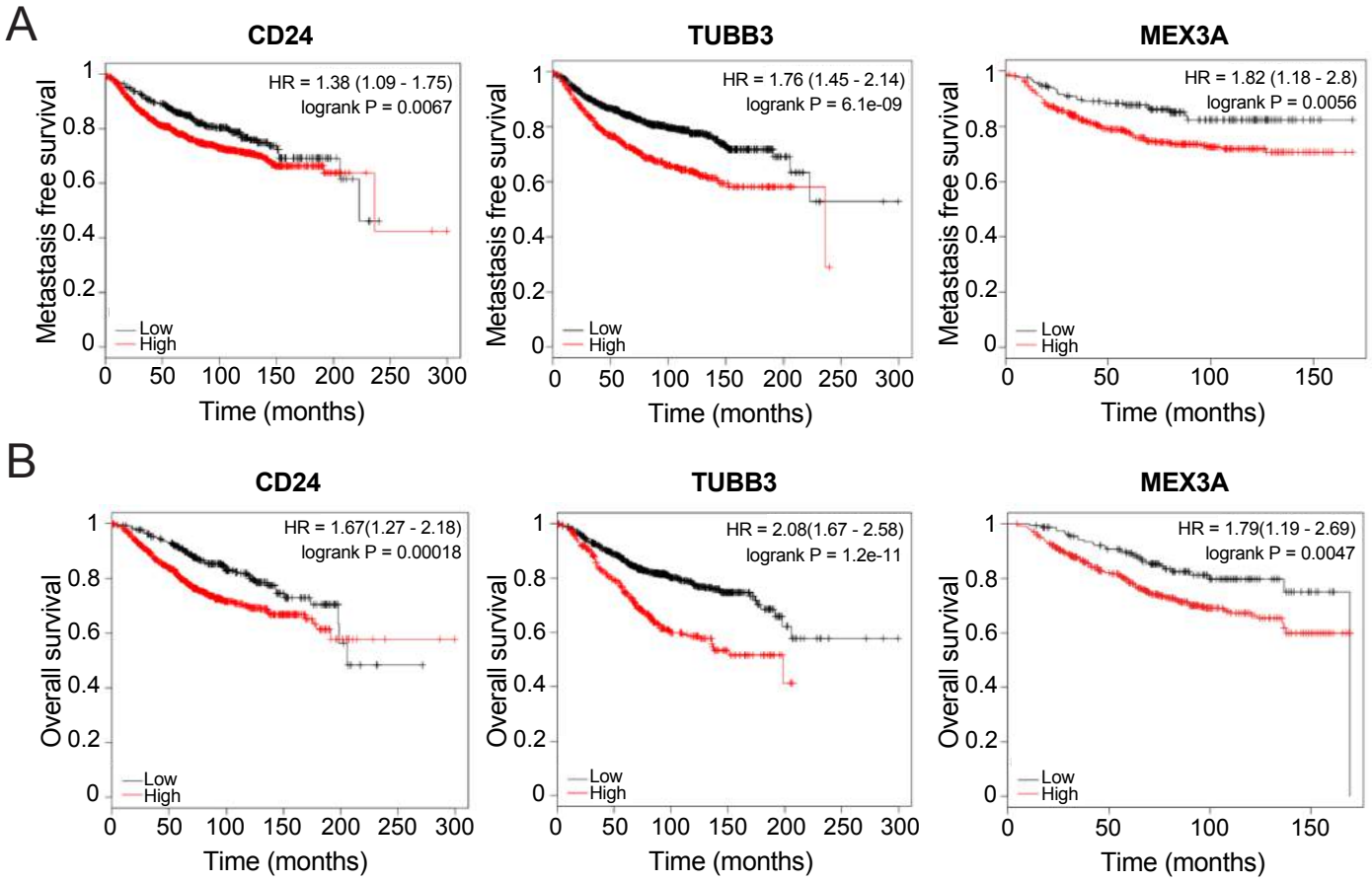
SOX11												
q values	**Residual q values after removing segments shared with higher peaks**	**Amplitude Threshold**	**Q678_B**	**Q751_B**	**Q349_B**	**Q635_B**	**Q851_B**	**Q452_B**	**Q030_B**	**Q896_B**	**Q639_B**	**Q772_B**	**Q898_B**
1.34E-21	2.68E-21	Actual Copy Change Given	0.5601	0.4731	-0.0071	0.40344	0.40894	0.0289	0.0944	0.3489	0.8643	0.0649	0.2906
1.34E-21	2.68E-21	0: t<0.1; 1: 0.1<t< 0.9; 2: t>0.9	1	1	0	1	1	0	0	1	1	0	1
	primary breast cancer		ER+	ER+	ER+	ER-	ER+	ER-	ER-	ER-	ER?	ER-	ER-
			HER2+	HER2?	HER2-	HER2+	HER2-	HER2+	HER2+	HER2+	HER2-	HER2-	HER2-
							BRCA1-/-				*BRCA1-/-*		
	brain metastasis		ER+	ER+	ER+	**ER+**	ER+	ER-	ER-	ER-	ER?	ER-	ER-
			HER2-	HER2-	HER2-	HER2+	**HER2+**	HER2+	HER2+	HER2+	HER2-	HER2-	HER2-
Amplitude levels are derived from GISTIC (copy-number analysis algorithm) and indicate the copy-number level per gene:													
0 is diploid													
1 or Gain indicates a low-level gain (a few additional copies copies, often broad)													

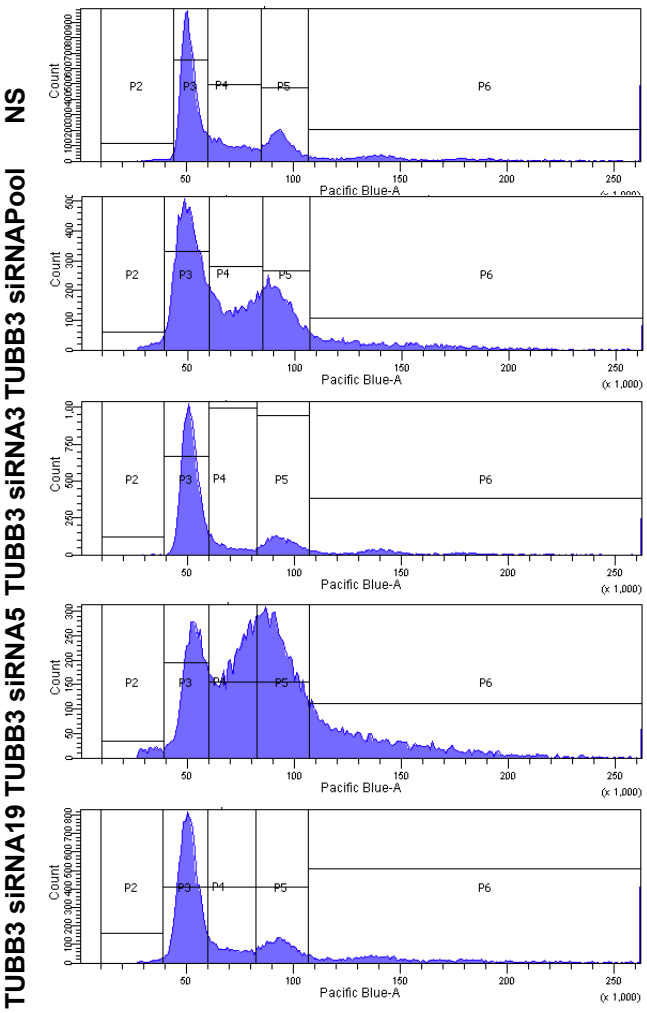
A

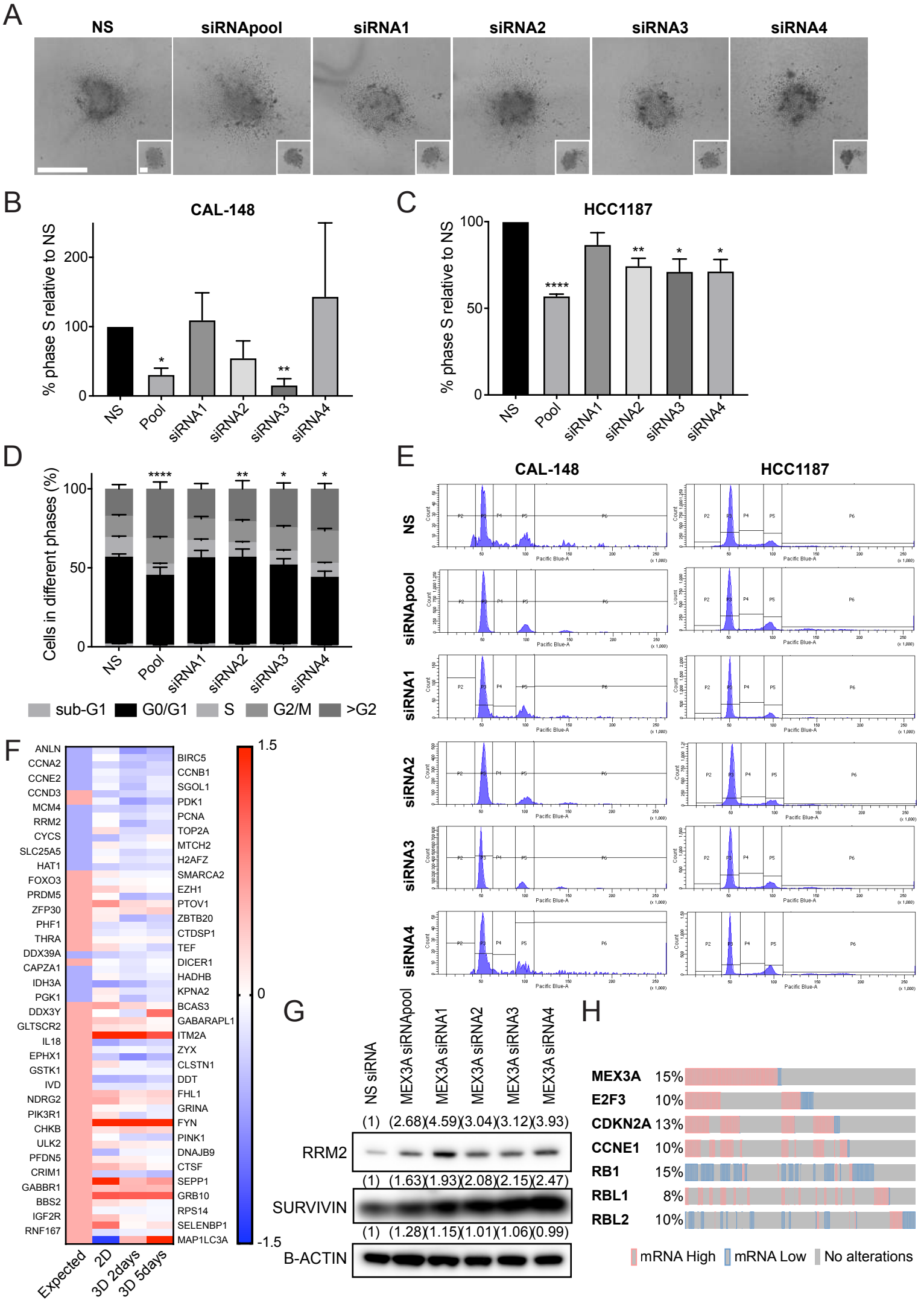


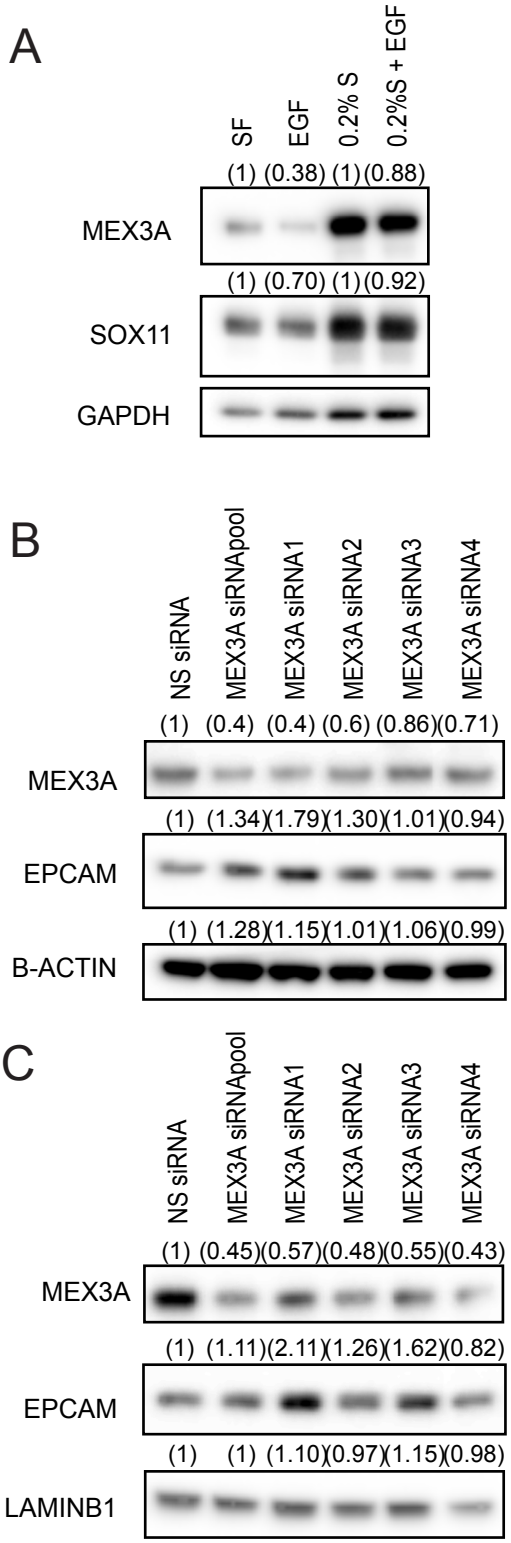
B

Case	SOX11 log2FC	ER	PR	HER2
X2	2.78232292	-	-	+
X5	2.004237269	-	-	-
X71	1.70187714	-	-	-
X1	0.782848917	-	-	+
X19	0.352422256	-	-	+
X7	0.33262191	+	-	+
X4	0.241737514	+	-	-
X51	0.113636244	+	-	-
X52	-0.151978814	-	+	+
X72	-0.161114792	+	+	-
X47	-0.271066812	+	+	+
X17	-0.41514682	+	-	+
X68	-0.430151202	-	-	-
X12	-0.810650677	-	-	-
X29	-0.881355706	-	-	-
X3	-1.028735668	+	-	+
X6	-1.277244249	+	-	-
X62	-1.316106927	+	+	-
X6pitt	-1.833267182	-	-	-
X25	-3.409142201	-	-	-
X64	-3.986661122	-	-	-


A
CD24
HR = 1.38 (1.09 - 1.75)
logrank P = 0.0067
Metastasis free survival
Time (months)
Low
High
TUBB3
HR = 1.76 (1.45 - 2.14)
logrank P = 6.1e-09
Metastasis free survival
Time (months)
Low
High
MEX3A
HR = 1.82 (1.18 - 2.8)
logrank P = 0.0056
Metastasis free survival
Time (months)
Low
High
B
CD24
HR = 1.67(1.27 - 2.18)
logrank P = 0.00018
Overall survival
Time (months)
Low
High
TUBB3
HR = 2.08(1.67 - 2.58)
logrank P = 1.2e-11
Overall survival
Time (months)
Low
High
MEX3A
HR = 1.79(1.19 - 2.69)
logrank P = 0.0047
Overall survival
Time (months)
Low
High


NS
TUBB3 siRNAPool
TUBB3 siRNA3
TUBB3 siRNA5
TUBB3 siRNA19
Count
Pacific Blue-A
(x 1,000)
P2
P3
P4
P5
P6


A
NS
siRNApool
siRNA1
siRNA2
siRNA3
siRNA4
B
CAL-148
% phase S relative to NS
200
100
0
*
**
NS
Pool
siRNA1
siRNA2
siRNA3
siRNA4
C
HCC1187
% phase S relative to NS
100
50
0
****
**
*
*
NS
Pool
siRNA1
siRNA2
siRNA3
siRNA4
D
Cells in different phases (%)
100
50
0
****
**
*
*
NS
Pool
siRNA1
siRNA2
siRNA3
siRNA4
sub-G1
G0/G1
S
G2/M
>G2
E
CAL-148
HCC1187
NS
siRNApool
siRNA1
siRNA2
siRNA3
siRNA4
Count
Pacific Blue-A
P2
P3
P4
P5
P6
F
ANLN
CCNA2
CCNE2
CCND3
MCM4
RRM2
CYCS
SLC25A5
HAT1
FOXO3
PRDM5
ZFP30
PHF1
THRA
DDX39A
CAPZA1
IDH3A
PGK1
DDX3Y
GLTSCR2
IL18
EPHX1
GSTK1
IVD
NDRG2
PIK3R1
CHKB
ULK2
PFDN5
CRIM1
GABBR1
BBS2
IGF2R
RNF167
BIRC5
CCNB1
SGOL1
PDK1
PCNA
TOP2A
MTCH2
H2AFZ
SMARCA2
EZH1
PTOV1
ZBTB20
CTDSP1
TEF
DICER1
HADHB
KPNA2
BCAS3
GABARAPL1
ITM2A
ZYX
CLSTN1
DDT
FHL1
GRINA
FYN
PINK1
DNAJB9
CTSF
SEPP1
GRB10
RPS14
SELENBP1
MAP1LC3A
1.5
0
-1.5
Expected
2D
3D 2days
3D 5days
G
NS siRNA
MEX3A siRNApool
MEX3A siRNA1
MEX3A siRNA2
MEX3A siRNA3
MEX3A siRNA4
(1) (2.68)(4.59)(3.04)(3.12)(3.93)
RRM2
(1) (1.63)(1.93)(2.08)(2.15)(2.47)
SURVIVIN
(1) (1.28)(1.15)(1.01)(1.06)(0.99)
B-ACTIN
H
MEX3A 15%
E2F3 10%
CDKN2A 13%
CCNE1 10%
RB1 15%
RBL1 8%
RBL2 10%
mRNA High
mRNA Low
No alterations


A
SF
EGF
0.2% S
0.2%S + EGF
(1) (0.38) (1) (0.88)
MEX3A
(1) (0.70) (1) (0.92)
SOX11
GAPDH
B
NS siRNA
MEX3A siRNApool
MEX3A siRNA1
MEX3A siRNA2
MEX3A siRNA3
MEX3A siRNA4
(1) (0.4) (0.4) (0.6) (0.86) (0.71)
MEX3A
(1) (1.34) (1.79) (1.30) (1.01) (0.94)
EPCAM
(1) (1.28) (1.15) (1.01) (1.06) (0.99)
B-ACTIN
C
NS siRNA
MEX3A siRNApool
MEX3A siRNA1
MEX3A siRNA2
MEX3A siRNA3
MEX3A siRNA4
(1) (0.45) (0.57) (0.48) (0.55) (0.43)
MEX3A
(1) (1.11) (2.11) (1.26) (1.62) (0.82)
EPCAM
(1) (1) (1.10) (0.97) (1.15) (0.98)
LAMINB1