Global transcriptional analysis identifies a novel role for SOX4 in tumor-induced angiogenesis

Stephin J Vervoort†, Olivier G de Jong†, M Guy Roukens†, Cynthia L Frederiks†, Jeroen F Vermeulen†, Ana Rita Lourenço†, Laura Bella†, Ana Tufegdzic Vidakovici, José L Sandoval†, Cathy Moelans†, Miranda van Amersfoort†, Margaret J Dallman†, Alejandra Bruna†, Carlos Caldas†, Edward Nieuwenhuis†, Elsken van der Wall†, Patrick Derksen†, Paul van Diest†, Marianne C Verhaar†, Eric W-F Lam†, Michal Mokry†, Paul J Coffer†,‡

†Department of Cell Biology, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands; ‡Department of Nephrology and Hypertension, University Medical Center Utrecht, Utrecht, The Netherlands; †Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands; †Department of Surgery and Cancer, Imperial Centre for Translational and Experimental Medicine, Imperial College London, Hammersmith Hospital Campus, London, United Kingdom; †Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Cambridge, United Kingdom; †Department of Life Sciences, Division of Cell and Molecular Biology, Imperial College London, London, United Kingdom; †Division of Pediatrics, Wilhelmina Children’s Hospital, University Medical Center Utrecht, Utrecht, The Netherlands; †Cancer Center, University Medical Center Utrecht, Utrecht, The Netherlands

Abstract The expression of the transcription factor SOX4 is increased in many human cancers, however, the pro-oncogenic capacity of SOX4 can vary greatly depending on the type of tumor. Both the contextual nature and the mechanisms underlying the pro-oncogenic SOX4 response remain unexplored. Here, we demonstrate that in mammmary tumorigenesis, the SOX4 transcriptional network is dictated by the epigenome and is enriched for pro-angiogenic processes. We show that SOX4 directly regulates endothelin-1 (ET-1) expression and can thereby promote tumor-induced angiogenesis both in vitro and in vivo. Furthermore, in breast tumors, SOX4 expression correlates with blood vessel density and size, and predicts poor-prognosis in patients with breast cancer. Our data provide novel mechanistic insights into context-dependent SOX4 target gene selection, and uncover a novel pro-oncogenic role for this transcription factor in promoting tumor-induced angiogenesis. These findings establish a key role for SOX4 in promoting metastasis through exploiting diverse pro-tumorigenic pathways.

DOI: https://doi.org/10.7554/eLife.27706.001

Introduction Transcription factor-mediated control of gene expression networks is crucial during embryonic development and thereafter for tissue homeostasis. Consequently, dysregulation of transcription factor function has been observed to result in a wide-variety of developmental defects and is often responsible for the pathogenesis of disease. A restricted subset of transcription factors is often
expressed in human cancers, aberrantly re-activating developmental pathways which then contribute to tumor-progression and metastasis (Darnell, 2002).

SOX4 is a prominent cancer-associated transcription factor and its mRNA expression is elevated in a large number of human cancers, being part of a general human cancer-associated gene expression signature (Rhodes et al., 2004). It is a member of the SRY-related HMG-box (SOX) family of transcription factors that have important roles in embryonic development and tissue homeostasis, for example by controlling differentiation and maintaining adult tissue stem cell populations (Lourenço and Coffer, 2017; Vervoort et al., 2013a; Sarkar and Hochedlinger, 2013). Sox4-deficient mice die during embryogenesis due to defective cardio-vasculature development (Schilham et al., 1996). In addition, SOX4 controls the development of specific tissues, including lymphoid, pancreatic, brain and bone (Vervoort et al., 2013a). In contrast to its wide-spread expression during embryogenesis, the expression of SOX4 in adult tissues is mostly restricted to adult stem- and progenitor cell populations including intestinal and hematopoietic stem cells (Sarkar and Hochedlinger, 2013; Vervoort et al., 2013a).

SOX4 expression in human cancers has been positively correlated with tumor-progression in a wide-variety of solid and hematopoietic tumors (Lourenço and Coffer, 2017; Vervoort et al., 2013a). Accordingly, SOX4 hypomorphic mice have decreased cancer-incidence and a resistance to carcinogen-induced skin cancer (Foronda et al., 2014).

The pro-oncogenic function of SOX4 has been attributed to a number of key cell-intrinsic processes including cell proliferation, cell-cycle regulation and tumor stemness (Vervoort et al., 2013a). A recurring theme is that SOX4 endows tumor cells with a more migratory and invasive phenotype. This has been shown using in vitro assays employing a large variety of different tumor types, such as breast cancer (Tavazoie et al., 2008; Zhang et al., 2012), hepatocellular carcinoma (Liao et al., 2008), ovarian cancer (Yeh et al., 2013), prostate cancer (Wang et al., 2013) and lung cancer (Zhou et al., 2015). Moreover, SOX4 expression correlates with increased depth of invasion in clinical specimens (Fang et al., 2012; Lin et al., 2013). For a limited number of tumor types, downstream targets of SOX4 have been identified that were important for invasion such as NRP1 and SEMA3C (hepatocellular carcinoma; Liao et al., 2008), TEAD2 and RBP1 (lung cancer; Castillo et al., 2012) and EGFR, Tenascin C (prostate cancer; Scharer et al., 2009). However, despite the similarity in phenotype that SOX4 confers in the various cell types, the overlap of transcriptional targets in the different studies has proven to be very limited (Vervoort et al., 2013a) suggesting that SOX4 has context-dependent effects on tumor development.

A number of studies have indicated a role for SOX4 in mammary tumor progression. In breast cancer, SOX4 is directly controlled by miRNA-335, the loss of which is associated with disease progression and poor metastasis-free survival (Tavazoie et al., 2008). SOX4 has also been demonstrated to be a part of gene signatures associated with metastasis of breast tumors to the brain and lungs (Minn et al., 2005; Bos et al., 2009). Moreover, SOX4 has been shown to control the TGF-β-induced epithelial-to-mesenchymal transition (EMT), a process associated with increases in tumor-initiating cells, in invasive and migratory capacity, in metastasis and in drug resistance (Zhang et al., 2012; Tiwari et al., 2013; Kalluri, 2009; Vervoort et al., 2013b; Lourenço and Coffer, 2017). In mice, SOX4-mediated induction of EMT has been proposed to be mediated by direct regulation of the histone-lysine N-methyltransferase EZH2, resulting in alterations in local histone three lysine (K) 27 trimethyl (H3K27me3) occupancy followed by changes in expression of EMT-associated genes (Tiwari et al., 2013). In line with its increased metastasis-associated expression and role in EMT, depletion of SOX4 has also been shown to reduce metastasis formation in mouse models of breast cancer (Tiwari et al., 2013; Tavazoie et al., 2008). The context-dependent nature of the SOX4 transcriptional response can in part be explained by cell-type-specific expression of cooperating transcription factors (Wilson and Koopman, 2002). Recently, it has been demonstrated in murine pancreatic ductal adenocarcinoma (PDA) that, the cell-type-specific expression of KLF5 can redirect the SOX4-mediated TGF-β transcriptional network from a pro-apoptotic to a pro-tumorigenic response, thus preventing lethal EMT (David et al., 2016). However, the wider relevance of this process in tumor progression remains to be elucidated.

Despite these observations, the transcriptional network and molecular mechanisms through which SOX4 contributes to the pathogenesis of breast cancer remain largely unknown. Moreover, SOX4 protein expression in both primary and metastatic tumor-tissue of patients with breast cancer has not been systematically studied. Here, by using an integrative genome-wide analysis approach, we
provide novel insight into the mechanisms determining SOX4 target gene selection. We define a ‘core’ SOX4-transcriptional network in mammary epithelial cells that reveals an unexpected role of SOX4 in tumor-induced angiogenesis through direct regulation of Endothelin-1 (ET-1). Our in vitro and in vivo experiments demonstrate a novel pro-oncogenic function of SOX4 in mammary tumor development as a promoter of tumor-induced angiogenesis, and furthermore we show that SOX4-high breast tumors are highly vascularized, aggressive and therapy-resistant. These findings provide both a resource for understanding SOX4 function as well as defining a novel role for this transcription factor in regulating tumor-induced angiogenesis.

Results

SOX4 is an epigenome-guided transcriptional activator

To better understand the functional role of SOX4 in tumorigenesis, it is imperative to understand how target gene specificity and activation are achieved. We aimed to characterize SOX4 chromatin-binding on a genome-wide level by using chromatin-immunoprecipitation and sequencing (ChIP-seq). To identify an antibody suitable for ChIP-seq, we first analyzed the ability of a number of SOX4 antibodies to immunoprecipitate (IP) SOX4 in combination with Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (RIME) (Figure 1—figure supplement 1A; Figure 1—source data 1) (Mohammed et al., 2013). This analysis yielded a single suitable antibody (Diagenode, CS-129–100), which was subsequently used for ChIP-seq analyses.

To globally evaluate SOX4 transcriptional function, the well-characterized human mammary epithelial cell line (HMLE) was utilized as a model system (Taube et al., 2010). We have previously demonstrated that HMLE cells express low but detectable levels of SOX4 and rapidly gain metastatic traits associated with EMT upon conditional activation of SOX4 (Vervoort et al., 2013b). In order to investigate the transcriptional output of SOX4, a doxycycline (DOX)-inducible SOX4 HMLE cell line (HMEL-S4) was generated, allowing rapid and conditional expression of SOX4 (Figure 1—figure supplement 1B). SOX4 ChIP-seq was subsequently performed in wild-type HMLE cells and in doxycycline-treated or -untreated HMEL-S4 cells. We simultaneously generated genome-wide binding profiles for RNA-polymerase II (POL2), the transcription start site (TSS)-associated histone mark histone H3 lysine (K) four trimethylation (H3K4me3), the repressive mark H3K27me3 and the active regulatory mark H3K27 acetyl (H3K27ac), in the latter (Figure 1—source data 2).

SOX4-bound loci were successfully identified in both HMLE wild-type (WT), untreated and DOX-treated HMEL-S4 cells and, as expected, the largest number of peaks was uniquely bound in DOX-treated HMEL-S4 cells (Figure 1A), as reflected by an increased SOX4 occupancy on the HDAC2 promoter (Figure 1—figure supplement 1C). To analyze the genomic distribution of SOX4 binding, we determined the distance of SOX4 peaks to the closest TSS. In HMLE WT cells, untreated and DOX-treated HMEL-S4 cells, SOX4 binding was enriched at TSS-proximal sites, compared to random genomic regions (Figure 1B and Figure 1—figure supplement 1D–E), indicating that SOX4 binding is enriched at promoter regions.

To both validate and gain further insight into the sequence requirements for SOX4 binding, de novo motif analysis was performed. As expected, the SOX4-consensus DNA-binding motif was the most highly enriched motif in DOX-treated HMEL-S4 cells (Figure 1C) and was present in over 67% of all peaks. The SOX4 consensus motif was also centrally enriched in the identified peaks, further confirming sequence-dependent binding (Figure 1D). In addition to the SOX4 motif, a number of co-occurring motifs were identified corresponding to AP-1, CTCF, ETS1 and SMAD3 (Figure 1C).

To further define the features that underlie binding of SOX4 to chromatin, the epigenetic profile of SOX4-bound sites in HMEL-S4 cells was analyzed. SOX4-binding sites were centrally enriched in histone marks associated with active and open chromatin (H3K4me3 and H3K27ac), as well as in high levels of POL2, whereas the repressive mark H3K27me3 was absent from SOX4-bound sites (Figure 1E). To quantify these effects, the discriminating features between SOX4-bound and -unbound sites within regions of open chromatin, as defined by H3K27ac signal, were assessed. Quantitative analysis of SOX4-bound versus -unbound H3K27ac sites revealed that SOX4-bound sites have significantly higher occupancy for H3K27ac, H3K4me3 and POL2 compared to unbound sites, and are devoid of H3K27me3 occupancy (Figure 1F). These findings indicate that SOX4 preferentially binds to active/open chromatin.
Figure 1. SOX4 is a transcriptional activator, preferentially binding to open/active chromatin. (A) Venn-diagram showing the overlap of SOX4 binding sites identified by ChIP-seq in WT-HMLE and DOX-induced HMLE-S4 cells. (B) Genomic distribution of SOX4 binding sites in WT-HMLE cells with respect to annotated genes, represented as distance from the nearest TSS. Random genomic regions were used as background. (C) De novo motif analysis of SOX4-bound sites in DOX-treated HMLE-S4 cells. Selected significantly enriched motifs are represented. SOX4 was identified as the most enriched motif.
To assess whether SOX4-induction in HMLE-S4 cells can itself result in alterations in active histone marks and POL2, all H3K27ac loci were ranked according to their relative change in SOX4-signal, or H3K27ac-signal in DOX-treated versus -untreated HMLE-S4 cells. This revealed that induction of SOX4 results in a concomitant increase in POL2 and H3K27ac occupancy (Figure 1G), which is also apparent at the genomic locus surrounding the validated SOX4 target gene TEAD2 (Figure 1H) (Bhattaram et al., 2010). No change was observed for the subset of sites where SOX4 binding remains unaltered (Figure 1—figure supplement 1F). In agreement, the most highly induced H3K27ac sites showed increased SOX4 and POL2 occupancy (Figure 1—figure supplement 1G).

Taken together, these findings show that increased binding of SOX4 to regions of open chromatin promotes an active chromatin-state.

The reliance on a pre-existing chromatin-state for SOX4 binding implies that SOX4 transcriptional effects are likely to be highly context dependent. Indeed, analysis of overlap between SOX4-bound sites in HMLE cells and sites identified in two distinct breast cancer cell lines, the HER2-amplified HCC1954 cells and highly metastatic triple-negative MDA-MB-231 cells, revealed a minor degree of overlap in SOX4-bound loci, although the overlap on the associated genes was higher (Figure 1—figure supplement 1H–I). SOX4 thus utilizes a pre-existing chromatin-landscape by binding to open/active chromatin, and binding to these sites subsequently increases the occupancy of active histone-marks and POL2.

Characterization of the SOX4 transcriptional network

To characterize the SOX4 transcriptional network, RNA-sequencing (RNA-seq) was subsequently performed. To this end, HMLE cells stably expressing SOX4 fused to the estrogen-receptor hormone-binding domain (ERSOX4 HMLE) were utilized, allowing for extremely rapid activation of SOX4 by 4-hydroxytamoxifen (4-OHT) (Vervoort et al., 2013b). ER-control HMLE and ERSOX4 HMLE cells were stimulated with 4-OHT (8 hr) after which total RNA was isolated and subjected to RNA-seq to identify alterations in gene expression. Data analysis revealed that 1378 genes were significantly regulated (q-value < 0.05) and differentially expressed by at least two-fold upon conditional SOX4 activation (Figure 2A), whereas no significantly altered genes were identified in ER HMLE control cells.

To identify SOX4 direct transcriptional targets, the RNA-seq dataset was overlapped with genes containing a SOX4 -binding site within 20 kb of their TSS (Figure 2B). This revealed a core set of 650 genes, which were shown to be bound and transcriptionally regulated by SOX4 (Figure 2B–C). Notably, the largest fraction of genes showed increased expression upon SOX4 activation (63%), while a considerably smaller percentage were found to be repressed targets (37%).

In order to independently validate the targets identified in our RNA-seq dataset, quantitative real-time PCR (qRT-PCR) was performed on ER-control and ERSOX4 HMLE cells stimulated with 4-OHT (8 hr and 24 hr). For all selected target genes, SOX4 conditional activation resulted in a time-dependent increase in expression, with no significant change in the expression in ER-control HMLE cells.
Figure 2. Identification of a novel core-SOX4 target gene signature with prognostic value and pro-angiogenic networks. (A) Gene expression analysis of SOX4 induced changes in HMLE ERSOX4 cells conditionally activated by 4-OHT for 8 hr and analyzed by RNA-seq. No significantly changed genes were observed in ER-control HMLE cells (data not shown). Significantly changed genes (q-value < 0.05) are indicated in red. (B) Venn diagram showing the overlap of genes regulated by SOX4 on the RNA-level (FC > 2, q-value < 0.05) with genes possessing a SOX4 binding site within 20 kb from an TSS.

Vervoort et al. eLife 2018;7:e27706. DOI: https://doi.org/10.7554/eLife.27706

Vervoort et al. eLife 2018;7:e27706. DOI: https://doi.org/10.7554/eLife.27706

6 of 30 Cancer Biology | Cell Biology
annotated TSS in DOX-treated HMLE-S4 cells (650 genes). (C) Heatmap visualizing gene expression changes for core-SOX4 target genes. (D) Quantitative real-time PCR analysis of gene expression changes in 4-OHT treated ER-control and ERSOX4 HMLE cells for selected SOX4 target genes identified by RNA-seq. Data represented as mean ± SD, normalized for β2M of three independent biological replicates. (E) Gene-set enrichment analysis (GSEA) representing the enrichment of SOX4 bound sites (2 kb from TSS) in the RNA-seq expression dataset ranked by log2 fold change upon conditional activation of SOX4. (F) Quantitative analysis of changes in SOX4 occupancy in peaks associated with most highly induced or repressed genes derived from the RNA-seq dataset generated in ERSOX4 HMLE cells (****p<0.0001, Wilcoxon-signed-rank test). (G) Analysis of the prognostic value of the core-SOX4 gene expression signature (650 genes) in claudinlow breast cancer patients (METABRIC).

Figure supplement 1. Overlap of ChiP-seq and RNA-seq datasets.
DOI: https://doi.org/10.7554/eLife.27706.007

SOX4 positively regulates EDN1 expression and promotes ET-1 secretion

To identify enriched cellular processes in the core-SOX4 gene set (650 genes), gene-ontology (GO) analysis was performed. GO-term analysis of SOX4 core genes showed a significant association with blood vessel morphogenesis, in addition to cell cycle- and cell migration-associated processes (Figure 3A). SOX4 target genes were also enriched for genes encoding secreted proteins that are involved in cell migration, extra-cellular composition, angiogenesis and inflammation (Figure 3—figure supplement 1A–B). The association with vascular development and cell migration was restricted to the positively regulated gene-set, whereas terms related to cell cycle and apoptosis were enriched in repressed genes (Figure 3—figure supplement 1C–D).

A role for SOX4 in tumor-induced angiogenesis remains unexplored and could be an essential component of its pro-tumorigenic effects. To this end, we examined the genes associated with angiogenesis more closely, and 13 genes were found to be induced upon conditional SOX4 activation (Figure 3B). We chose to further evaluate the role of Endothelin-1 (EDN1) since its expression has been demonstrated to positively correlate with breast cancer tumor-grade and distant metastasis, and has been suggested to promote tumor-induced angiogenesis, migration and invasion (Salani et al., 2000; Wülfing et al., 2004). EDN1 encodes preproendothelin, which through intracellular cleavage by convertases and Endothelin Converting Enzyme-1 (ECE-1) is converted into a 21-amino acid mature form (Rosanò et al., 2013). Endothelin-1 (ET-1) directly induces angiogenesis by binding to its cognate receptors (ETAR and ETBR) expressed on endothelial cells (Rosanò et al., 2013). Interestingly, both EDN1 and its converting enzyme ECE-1 were observed to be transcriptionally induced, indicating that SOX4 targets dual aspects of this pro-angiogenic pathway.
Figure 3. SOX4 directly regulates EDN1 mRNA expression and promotes production of the mature ET-1 peptide. (A) Gene Ontology (GO) analysis on core-SOX4 target genes. Top terms are shown. (B) Heatmap of gene expression changes upon conditional SOX4 activation in HMLE cells for genes associated with the GO-term Blood vessel morphogenesis (HMLE regulated and bound). Genes with a SOX4-binding site within 20 kb of a TSS in MDA-MB-231 cells are annotated in red. (C) Normalized ChiP-seq tracks for SOX4, H3K27me3, H3K4me3, POL2, and H3K27ac surrounding the EDN1 locus in Figure 3 continued on next page.
In line with our previous observations, analysis of ChIP-seq and RNA-seq profiles of EDN1 locus indicated a strong increase in POL2 and H3K27ac upon DOX-mediated activation of SOX4 in HMLE-S4 cells (Figure 3C) and an increase in mRNA expression of EDN1 upon conditional activation of SOX4 in ERSOX4 HMLE cells (Figure 3D). This induction of EDN1 expression was also observed by qRT-PCR in ERSOX4 HMLE cells, and was absent in control ER HMLE cells (Figure 3E). EDN1 was also identified as a direct SOX4 target by ChiP in both MDA-MB-231 and HCC1954 cell lines (Figure 1-figure supplement 1H and data not shown). Indeed upon shRNA-mediated knockdown of SOX4 in these two cell lines we also observed that EDN1 expression was decreased (Figure 6, Figure 6—figure supplement 1A–B for MDA-MB-231 and Figure 3—figure supplement 2C for HCC1954).

Closer examination of the EDN1 promoter overlapping SOX4 and H3K27ac profiles identified two conserved SOX4 binding sites at 70 bp and 350 bp from the TSS (Figure 3—figure supplement 2A). To assess whether the SOX4 binding region (−961 bp - + 3 bp) upstream of the TSS is sufficient for transcriptional activation of EDN1 by this factor, luciferase assays were performed using a reporter construct containing the EDN1 promoter region. Ectopic expression of SOX4 strongly induced luciferase expression as compared to the empty vector (EV) control, which could be inhibited by the co-transfection of a dominant negative truncated SOX4 lacking the transactivation domain (Figure 3—figure supplement 2B). Activation of the luciferase reporter was significantly inhibited when the SOX4 binding site at −350 bp was mutated, whereas no significant reduction was observed upon mutation of the −70 bp site, indicating that transcriptional activation of EDN1 by SOX4 under these conditions is mediated by the site at −350 bp from the TSS (Figure 3F). To determine whether SOX4 can indeed directly bind these sites, pull-down assays were performed with biotinylated DNA-probes containing the identified EDN1 promoter binding sites or a consensus SOX4 binding site (positive control) and mutated versions thereof. SOX4 was found to specifically bind to the conserved motifs at −350 bp from the EDN1 TSS, which could be inhibited by mutation of this site (Figure 3G).

To assess whether transcriptional induction of EDN1 by SOX4 also results in increased production of the mature ET-1 peptide, enzyme-linked immunosorbent assays (ELISA) were performed. Conditional activation of SOX4 in HMLE resulted in a significant increase in ET-1 secretion by these cells (Figure 3H). Taken together these observations demonstrate that SOX4 directly induces EDN1 expression on the mRNA level resulting in increased expression of the secreted ET-1 peptide.

**SOX4 induced ET-1 expression promotes tumor-induced angiogenesis in vitro**

Since activation of SOX4 resulted in increased expression of ET-1, we next investigated whether this was able to induce pro-angiogenic effects on endothelial cells. To this end, SOX4 in ERSOX4 HMLE cells was activated for a period of eight hours, and subsequently the media was refreshed for overnight conditioning. The SOX4-activated conditioned media (SCM+) was subsequently tested for its
The functional effects of SCM+ on HMEC-1 cells were first investigated in a wound-healing assay (van Balkom et al., 2013). Relative to basal media alone, exposure of HMEC-1 cells to SCM+ was capable of increasing migration of HMEC-1 cells, whereas no significant increase was observed with SCM- or ER-control HMLE conditioned media (ERCM+) (Figure 4). Importantly, SCM+-induced migration could be inhibited by the dual endothelin receptor antagonist Bosentan (Clozel et al., 1994), demonstrating that the pro-migratory effect indeed required endothelin receptor activation (Figure 4B). SCM+ did not affect HMEC-1 cell proliferation, indicating that wound-closure is exclusively dependent on cell migration (Figure 4—figure supplement 1A). Similar to SCM+, treatment of HMEC-1 cells with a synthetic ET-1 peptide resulted in a significant increase in wound closure, although not to the same extent as complete media which contains additional pro-migratory and pro-angiogenic factors (Figure 4C). In order to assess whether the SCM+-mediated effects required transcriptional induction of EDN1 by SOX4, siRNA-mediated knockdown of EDN1 was performed in HMLE cells 24 hr prior to conditional SOX4 activation. EDN1 knockdown reduced the SCM+-induced migration in HMLE-1 cells (Figure 4D, Figure 4—figure supplement 1B). Similar to conditional activation in HMLE cells, stable expression of SOX4 in MDA-MB-231 cells also resulted in increased EDN1 expression (Figure 4—figure supplement 1C) and conditioned media with the capacity to increase HMEC-1 migration (Figure 4E).

To further evaluate the pro-angiogenic effects of ET-1 and SCM+, a matrigel angiogenesis assay was utilized. When plated in low-density on a matrigel surface, HMEC-1 cells rapidly form a meshed network, the length of which is an in vitro measure for their angiogenic potential. Relative to basal media alone, SCM+ treatment of HMEC-1 cells resulted in an increase in the relative network length, which could be inhibited by Bosentan, whereas no increase in network formation was observed in control conditions (Figure 4—figure supplement 1D–E). Similarly, incubation with the ET-1 peptide significantly increased the network length, an effect that could also be inhibited by treatment with Bosentan (Figure 4—figure supplement 1F).

In an analogous manner, the capacity of SCM+ and ET-1 to induce in vitro angiogenic sprouting was also investigated. Collagen-coated sephadex beads were incubated with HMEC-1 cells, allowing cell adhesion to the beads, after which cell-bound beads were embedded in matrigel. In the presence of pro-angiogenic stimuli, HMEC-1 cells form angiogenic sprouts that penetrate the matrigel layer. In this assay both ET-1 and SCM+ treatment of HMEC-1 cells resulted in a marked increase in sprouting, although less pronounced than full medium (Figure 4F–J). Relative to basal media, SCM+ potently induced sprouting in this assay by approximately 6-fold which could be inhibited by Bosentan (Figure 4F–G). Correspondingly, ET-1 induced sprouting by approximately 5-fold, which could also be completely inhibited by Bosentan (Figure 4H). Similar to the results obtained in the wound-closure assays, siRNA mediated depletion of EDN1 prior to conditional SOX4 activation resulted in an almost complete inhibition of SCM+-induced sprouting (Figure 4I). Additionally, conditioned media from SOX4 expressing MDA-MB-231 cells also resulted in increased HMEC-1 sprouting (Figure 4J). In accordance, shRNA-mediated knockdown of SOX4 in MDA-MB-231 cells resulted in a decrease of EDN1 expression (Figure 6A–B, Figure 6—figure supplement 1A–B) and a concomitant decrease in the ability to induce migration and sprouting of endothelial cells (Figure 4—figure supplement 1G–H).

Taken together, these findings show that SOX4 activation in HMLE and in MDA-MB-231 cells can bestow pro-angiogenic effects that are dependent on ET-1.

**SOX4 expression in breast epithelial cells induces angiogenesis in vivo in a zebrafish tumor-xenograft model**

To determine whether SOX4 may regulate angiogenesis in vivo, we initially made use of a zebrafish tumor-xenograft model (Figure 5A). In this model, fluorescently labeled tumor cells (red) are injected into the yolk sac of 1 day post-fertilization (dpf) zebrafish embryos in proximity to the subintestinal vein (SIV) structure. In the presence of tumors, alterations in the normal development of the SIV, such as protrusions extending from the structure, are considered to be representative of tumor-induced angiogenesis. Ectopic neovascularization originating from the subintestinal vessels (SIV) can...
**Figure 4.** SOX4 mediated induction of ET-1 expression promotes tumor-induced angiogenesis in vitro. (A) Representative images of wound-healing assays performed on HMEC-1 cells, treated for 5 hr with basal media, full growth media or conditioned media from HMLE-S4 cells with or without overnight treatment with 4-OHT (SCM + and SCM-, respectively). Data from three independent biological replicates (B) Quantification of wound closure relative to basal media for HMEC-1 cells treated with conditioned media; SCM-, SCM+, SCM+ with Bosentan, and conditioned media from ER-control

Figure 4 continued on next page
be quantified by fluorescence microscopy taking advantage of the presence of GFP-positive blood vessels characteristic of the Tg(fli:GFP) embryos used in these experiments.

ER-control and ERSOX4 HMLE cells were either pre-treated in vitro with 4-OHT for 24 hr or left untreated before injection into the yolk sac of zebrafish embryos (Figure 5B). Neovascularization was quantified by counting the number of blood vessels protruding from the SIV. A significant increase in sprouting from the SIV was observed in 4-OHT treated ERSOX4 HMLE cells compared to untreated ERSOX4 HMLE cells as well as both treated and untreated ER-control HMLE cells, indicating that conditional activation of SOX4 in HMLE cells can induce angiogenesis in vivo (Figure 5B–D). No significant changes were observed upon injection of wild-type HMLE cells (Figure 5—figure supplement 1A).

We next examined whether depletion of SOX4 could inhibit ectopic sprouting induced by pro-angiogenic MDA-MB-231 tumor cells. To this end, MDA-MB-231 cells stably expressing a shRNA targeting SOX4 (SOX4 KD) or a non-targeting shRNA were injected into zebrafish embryos. Quantification of neovascularization demonstrated that SOX4-depleted MDA-MB-231 cells (Figure 5—figure supplement 1B) were significantly impaired in their ability to induce ectopic sprouting from the SIV, as illustrated by a decrease in the number of sprouting vessels and an overall reduction in the percentage of fish with sprouting blood vessels (Figure 5E–G). In order to confirm that SOX4 pro-angiogenic effects in vivo are, in part, dependent of ET-1 expression, siRNA-mediated knockdown of EDN1 was performed in HMLE cells 24 hr prior to SOX4 conditional activation, after which cells were injected into zebrafish embryos. Similar to our previous observations in vitro, depletion of ET-1 in SOX4-activated cells results in the reduction of angiogenic potential, shown by the decreased capacity to induce ectopic sprouting from the SIV (Figure 5H–J). Taken together, these results indicate that, similar to our in vitro observations, SOX4 controls tumor-induced angiogenesis in vivo, in an ET-1 dependent manner.

**SOX4 is required for ET-1 expression, tumor-vascularization and metastasis in a xenograft mouse model of breast cancer**

To further study the role of SOX4 in tumor-induced angiogenesis in vivo, we made use of a xenograft mouse model of breast cancer, in which MDA-MB-231 cells are transplanted into the mammary fat-pad of immune-deficient mice. Luciferase expressing MDA-MB-231 cells were generated and lentivirally transduced with a SOX4 KD shRNA or a control-scrambled shRNA construct. Prior to injection, we validated the efficiency of SOX4 knockdown by qRT-PCR. This revealed a significant reduction in SOX4 expression in SOX4 KD MDA-MB-231 cells compared to scrambled control cells, which was accompanied by a significant decrease in EDN1 expression (Figure 6A–B). Following validation, cells were transplanted into the mammary fat-pad and we subsequently monitored tumor growth up to six weeks post-implantation. Primary tumor growth was unaffected by SOX4 depletion and both scrambled control and SOX4 KD mice were sacrificed in the same week (Figure 6C). Analysis of primary tumors by immunohistochemistry (IHC) for SOX4, ET-1 and CD31 (PECAM-1) expression showed that SOX4 KD knockdown was maintained in vivo (Figure 6D). Moreover, quantitative analysis of ET-1 expression revealed a significant reduction in ET-1 IHC staining in
Figure 5. SOX4 controls tumor-induced angiogenesis in vivo in a zebrafish tumor-xenograft model. (A) Schematic representation of the zebrafish tumor-xenograft model, with the subintestinal vessels indicated in green (SIV) and tumor cells in red. (B) Representative images of zebrafish injected with ER-control and ERSOX4 HMLE cells treated with 4-OHT. Arrows indicate newly formed blood vessels. (C) Quantification of the number of ectopic sprouts observed per fish injected with ER-control and ERSOX4 HMLE cells. Individual data points and data represented as mean ± SD of three
independent biological replicates (**p-value<0.01, ***p-value<0.001; ANOVA). (D) Quantification of the relative increase in ectopic sprouting in fish injected with 4-OHT treated relative to untreated ER-control and ERSOX4 cells. Data represented as mean ± SD of three independent biological replicates (* p-value<0.05, Student’s t-test). (E) Quantification of the number of ectopic sprouts in fish injected with SCR control and SOX4 KD MDA-MB-231 cells. Results obtained from three independent biological replicates (**p-value<0.01, Student’s t-test). (F) Representative images of zebrafish injected with SCR control or SOX4 KD MDA-MB-231 cells. Arrows indicate newly formed blood vessels (G) Quantification of the number of fish injected with MDA-MB-231 cells with or without ectopic sprouting. Results obtained from three independent biological replicates (**p-value<0.001, Student’s t-test). (H) Quantification of the number of ectopic sprouts observed per fish injected with 4-OHT-treated ERSOX4 HMLE cells exposed with siRNA control or siRNA targeting ET-1. Individual data points and data represented as mean ± SD (*p-value<0.05, ANOVA). (I) Quantification of the relative increase in ectopic sprouting in fish injected with 4-OHT treated ERSOX4 HMLE cells exposed with siRNA control or siRNA targeting ET-1. Data represented as mean ± SD of three independent biological replicates (**p-value<0.05, ***p-value<0.001; ANOVA). (J) Representative images of zebrafish injected with 4-OHT treated ERSOX4 HMLE cells exposed with siRNA control or siRNA targeting ET-1. Arrows indicate newly formed blood vessels.

DOI: https://doi.org/10.7554/eLife.27706.014

The following figure supplement is available for figure 5:

**Figure supplement 1.** Analysis of angiogenic potential of HMLE WT cells.

DOI: https://doi.org/10.7554/eLife.27706.014

SOX4 KD compared to scrambled control tumors (**Figure 6D–E**). We next assessed tumor vascularization by quantifying expression of the endothelial cell surface marker CD31. SOX4 KD tumors had significantly lower vascularization as illustrated by the reduction in total blood vessel area, blood vessel count and blood vessel size (**Figure 6F**). These findings demonstrate that SOX4 expression is required for tumor-associated ET-1 expression in vivo, and SOX4-depletion correlates with reduced tumor blood vessel density. Importantly, SOX4 depletion in this orthotopic xenograft model also significantly reduced the development of lung metastases (**Figure 6G**). To exclude the possibility that these results were simply caused by off-target effects of the shRNA, we performed an independent experiment using a second shRNA targeting SOX4 (**Figure 6—figure supplement 1**). This experiment confirmed our prior experiment, as mice transplanted with SOX4 shRNA2 MDA-MB-231 cells presented with a strong decrease in both metastatic outgrowth and vascularization. Here, there was also a modest effect on primary tumor growth, which may reflect better SOX4 knockdown in this experiment. Together, these findings demonstrate that SOX4 can control multiple aspects of breast tumor biology, and show for the first time that SOX4 may regulate tumor angiogenesis in vivo.

**SOX4 expression correlates with decreased survival, increased metastasis and elevated blood vessel density in patients with breast cancer**

To evaluate whether SOX4 expression correlates with clinical parameters in breast cancer, we performed IHC analysis of nuclear SOX4 expression levels in a tissue microarray (TMA) representing a cohort of 452 breast cancer patient tumor samples from both invasive ductal (IDC) and lobular carcinomas (ILC) (**Figure 7A–B**). Nuclear SOX4 expression was found to significantly correlate with tumor size, histological grade and mitotic index, all parameters of poor-prognosis (**Figure 7—source data 1, Figure 7—source data 2, Figure 7—source data 3, Figure 7—source data 4, Figure 7—source data 5**). Accordingly, high levels of SOX4 expression significantly correlated with decreased overall survival in breast cancer patients, and this was also observed when patients were subdivided into the major subtypes ILC and IDC (**Figure 7C–E**). Stratification of patients with breast cancer according to the ER-status revealed that patients with high levels of SOX4 have a significantly worse prognosis in the ER-positive tumor group, and a trend towards poor-prognosis in the ER-negative tumor group (**Figure 7—figure supplement 1A**).

We next evaluated the correlation between nuclear SOX4 intensity and the metastatic status for those cases in the TMA cohort where this information was available. This revealed that the percentage of tumors with high SOX4 expression was observed to be significantly higher in those patients who develop metastasis as compared to patients who do not present with metastatic disease (p=0.016; **Figure 7—source data 5**). Moreover, analysis of SOX4 expression in 23 matched primary-tumor metastasis patient samples revealed that SOX4 expression in the primary tumor reflects the expression level observed in the metastasis (p=0.039; **Figure 7—source data 5**). These observations indicate that SOX4 expression is a clear marker of poor-prognosis in breast cancer and underscore
Figure 6. Depletion of SOX4 in MDA-MB-231 cells reduces ET-1 expression and decreases tumor vascularization and metastasis in a xenograft mouse model of breast cancer. (A–B). qRT-PCR quantification of SOX4 and EDN1 expression in luciferase-marked SCR control and SOX4 KD MDA-MB-231 cells. Data represented as mean ± SD of three independent biological replicates (*p-value<0.05, Student’s t-test). (C) Growth curve of SOX4 KD and SCR control MDA-MB-231 tumors. Data represented as mean ± SD of three independent biological replicates. (D) IHC analysis of scrambled control and SOX4 KD MDA-MB-231 cell tumors. Data represented as mean ± SD of three independent biological replicates. (E–G) Hematoxylin & Eosin (H&E), SOX4, ET-1, and CD31 IHC images of scrambled control and SOX4 KD MDA-MB-231 cell tumors.
the important role of SOX4 in tumor-progression and metastasis observed in breast cancer model systems. Interestingly, no significant benefit to patient survival was observed for adjuvant radio- or chemotherapy-treated SOX4-high tumors as compared to untreated tumors, whereas a significant treatment response was observed for SOX4-low tumors (Figure 7—figure supplement 1C). These findings suggest that SOX4 may also contribute to therapy resistance.

To investigate whether SOX4 expression is also associated with increased angiogenesis in patients with human breast cancer, we examined CD34 expression (a marker for small and newly formed blood vessels) in 17 SOX4-low (SOX4LO) and 17 SOX4-high (SOX4HI) ductal carcinomas from our cohort (Figure 7F). Quantitative analysis of CD34-staining revealed that tumors expressing high levels of SOX4 had a significantly increased total blood vessel area and size, with an apparent but non-significant increase in the total number of blood vessels, thus suggesting that SOX4 can influence tumor-induced angiogenesis in human breast cancer (Figure 7G).

Taken together our integrative genome-wide analysis of the SOX4 transcriptional network has uncovered a novel role for SOX4 in tumor-induced angiogenesis in vitro and in vivo as well as in breast tumors, revealing a novel mechanism by which SOX4 may contribute to breast cancer metastasis (Figure 7H).

Discussion

Using a combination of in vitro experiments utilizing multiple epithelial-derived breast cell lines, in vivo animal models, and analysis of patient tumor samples, we show that SOX4-mediated transcriptional regulation can affect breast cancer progression by promoting tumor-induced angiogenesis. SOX4 activation promotes ET-1 expression and controls the angiogenic behavior of endothelial cells both in vitro and in vivo, thus affecting tumor-progression in a non-cell autonomous manner. Finally, we demonstrate that elevated SOX4 expression correlates with increased blood vessel density in patient tumors and predicts poor-prognosis.

Despite observations of increased SOX4 expression in a wide variety of tumors, effects can differ greatly between cancer-types resulting in cellular transformation, increased proliferation, EMT and metastasis, or even tumor-suppression (Rhodes et al., 2004; Vervoort et al., 2013a). The cell-type and context dependent effects underlying this differential SOX4 response are poorly understood. Our findings indicate that SOX4 acts as an opportunistic transcription factor directed by a pre-existing chromatin landscape, which upon recruitment to open chromatin further enhances the active chromatin-state. In addition to the epigenome, co-factors may determine SOX4 target gene selection as indicated by the co-occurrence of consensus-binding sites for AP-1, ETS-1 and SMAD3 in SOX4-bound sites. Cooperative binding appears to be a common theme in the SOX-family, as highlighted by SOX2 and OCT4 in embryonic stem cells, which results in the specific regulation of pluripotency genes including NANOG (Kamachi et al., 2000; Sarkar and Hochedlinger, 2013). These observations indicate that both the specific cell-type epigenome and the expression of co-factors are important in shaping the SOX4 transcriptional network, providing an explanation for the relatively minor degree of overlap in SOX4 target genes between cancer-types (Vervoort et al., 2013a). The context dependent nature of SOX4 transcriptional networks is of particular importance in heterogeneous diseases such as breast cancer. An important finding is that the SOX4 core-
Figure 7. Immunohistochemical analysis of SOX4 expression in patients with human breast cancer reveals a positive correlation with decreased survival, increased metastasis and elevated blood vessel density. (A–B) Representative images of SOX4 expression in invasive lobular carcinoma (ILC, classical and pleomorphic) and invasive ductal carcinoma (IDC, low-grade tumor and high-grade tumor) cases in the tissue microarray (TMA) cohort. (C–E) Kaplan-Meier curves for cumulative (cum) survival for patients with high or low nuclear SOX4 expression for the total cohort, ILC and IDC breast.
cancers. Indicated p-values were calculated using the Log-rank test. (F) Representative images of CD34 staining in SOX4<sup>HI</sup> and SOX4<sup>LO</sup> tumors as identified previously in TMA analysis. Arrows indicate blood vessels. (G) Quantitative analysis of CD34 staining in 17 SOX4<sup>HI</sup> and 17 SOX4<sup>LO</sup> ductal carcinomas. (Student t-test, *p-value<0.05). (H) Schematic model for the pro-angiogenic function of SOX4 during breast cancer progression. The SOX4 transcription factor binds to regions of active/open chromatin and subsequently acts as positive regulator of many genes involved in tumorigenesis, such as EDN1. Upregulation of ET-1 contributes to the increase of neovascularization, which ultimately may facilitate tumor cell intravasation and growth factor accessibility.

DOI: https://doi.org/10.7554/eLife.27706.017

The following source data and figure supplements are available for figure 7:

**Source data 1.** Correlation of nuclear SOX4 expression with clinicopathological and molecular features of invasive breast cancer.
DOI: https://doi.org/10.7554/eLife.27706.020

**Source data 2.** Correlation of nuclear SOX4 expression with clinicopathological features of IDC and ILC.
DOI: https://doi.org/10.7554/eLife.27706.021

**Source data 3.** Clinicopathological characteristics of 452 breast cancer patients studied for the expression of SOX4.
DOI: https://doi.org/10.7554/eLife.27706.022

**Source data 4.** Clinicopathological characteristics of 27 patients with metastatic breast cancer studied for the expression of SOX4.
DOI: https://doi.org/10.7554/eLife.27706.023

**Source data 5.** Correlation of nuclear SOX4 expression with metastasis formation.
DOI: https://doi.org/10.7554/eLife.27706.024

**Figure supplement 1.** SOX4 mRNA expression correlates with decreased patient survival in both ER positive and negative breast cancers.
DOI: https://doi.org/10.7554/eLife.27706.018

**Figure supplement 2.** Gene set enrichment analysis on RNA-seq data derived from conditional activation of SOX4 in HMLE cells.
DOI: https://doi.org/10.7554/eLife.27706.019
expression and blood vessel density. It has been previously reported that gene signatures associated with vascular activation in Claudin\textsuperscript{low} tumors were predictive for metastatic dissemination (Harrell \textit{et al.}, 2014). Our SOX4-regulated RNAseq data in HMLEs showed enrichment for one of these vascular signatures (Figure 7—figure supplement 2C) suggesting that SOX4 may drive tumor angiogenesis in Claudin\textsuperscript{low} tumors by regulating these genes in addition to \textit{EDN1}. This may contribute to the poor survival of patients with the SOX4 signature in patients with Claudin\textsuperscript{low} tumors.

Thus, our study may signify a potential connection between SOX4-mediated alterations in the tumor-microvasculature and the ability of cells to metastasize. Based on our current data however we cannot determine the degree to which tumor angiogenesis contributes to the pro-metastatic effects of SOX4, since it is not possible to exclude other processes such as EMT and invasiveness. While previous studies have reported that SOX4 can directly modulate tumor cell transcription, our study has demonstrated for the first time that SOX4 can impact tumorigenesis in a non-cell-autonomous fashion.

In agreement with a role in tumor-progression, ET-1 expression has been demonstrated to positively correlate with tumor-grade and the formation of distant metastasis in breast cancer (Wülffing \textit{et al.}, 2003). In addition to promoting angiogenesis, SOX4 induced ET-1 may also promote tumor-progression in an autocrine manner by inducing EMT or for example by affecting T-cell infiltration, both of which have demonstrated in ovarian cancer (Buckanovich \textit{et al.}, 2008; Rosano \textit{et al.}, 2005; Rosanò \textit{et al.}, 2013). SOX4-ET-1 induced autocrine and paracrine signaling may thus affect multiple tumor-promoting processes. It would be interesting to evaluate whether restoration of ET-1 expression is sufficient to restore the metastatic capacity of SOX4-depleted MDA-MB-231 cells in vivo. Since SOX4 does affect multiple processes that affect tumor progression we envision that ET-1 depletion would strongly reduce metastasis formation, but would most likely not abolish metastatic outgrowth completely.

In the last decade a number of anti-angiogenic therapies, mostly targeting the VEGF-receptor pathway, have been tested in clinical trials with limited success due to unwanted side-effects (Jain, 2014). These include the inhibition of effective delivery of anti-cancer therapeutics, but also the induction of pro-metastatic processes such as EMT, activated by nutrient deprivation and hypoxia (Jain, 2014). As a driver of both EMT and tumor-induced angiogenesis, SOX4 is an interesting therapeutic target in breast cancer, allowing for the simultaneous inhibition of multiple biological processes required for tumor metastasis. Alternatively, ETAR and ETBR receptor inhibitors such as Bosentan are currently used to treat pulmonary vascular hypertension, and based on our results could potentially be repurposed to specifically treat aggressive SOX4\textsuperscript{HI} breast tumors (Rosanò \textit{et al.}, 2013). Although success of ET-1 receptor antagonists in the treatment of advanced cancers, including glioma and melanoma, has been limited, promising results were obtained in a number of ovarian cancer patients when combined with standard chemotherapeutics (Rosanò \textit{et al.}, 2013). Based on our findings, patients with SOX4\textsuperscript{HI} tumors are more likely to profit from treatment with ET-1 receptor antagonists, thus arguing for the specific monitoring of this subgroup in future clinical studies. A caveat here is that considering the context-dependent nature of SOX4 transcriptional activity it is possible that \textit{EDN1} regulation is also dependent on the expression of co-factors and on the presence of open chromatin. Despite our finding that \textit{EDN1} was one of the few genes commonly regulated by SOX4 in the different breast epithelial cell lines used in this study, future work will have to uncover whether this regulation is conserved in all breast cancer cells and subtypes.

Taken together, our global-transcriptional analysis provides mechanistic insight into the contextual nature of SOX4 responses and uncovers a novel role for SOX4 in tumor-induced angiogenesis by direct regulation of ET-1 expression. The identification of this novel SOX4 controlled pathway underscores the validity of our approach and highlights the value of our dataset as a resource to uncover further novel functions for SOX4 in cancer and potentially to understand its role in developmental biology.
Materials and methods

Cell culture
Human Microvascular Endothelial Cells (HMEC-1) were obtained from the Center for Disease Control and Prevention, Atlanta USA (Ades et al., 1992) and maintained in MCDB 131 medium containing 10% FCS, 10 mM L-Glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin (All from Life Technologies, USA), 50 nM hydrocortisone (Sigma-Aldrich, USA), and 10 ng/ml rhEGF (Invitrogen, USA) up to passage 27 at 37°C, 5% CO₂ and ambient oxygen level (20%). Immortalized human mammary epithelial cells (HMEL) were kindly provided by Dr. R.A. Weinberg and were cultured in human mammary epithelial growth media (cc-3150, Lonza) as described previously (Vervoort et al., 2013a). MDA-MB-231 cells and HCC1954 cells were purchased from ATCC and cultured as described previously (Bruna et al., 2012). All cells were regularly checked for mycoplasma during the course of this study and only mycoplasma-negative cells have been used for the experiments described in this study.

Lentiviral transduction
Lentiviral transduction of MDA-MB-231 and HMEL cells was performed as described previously (Vervoort et al., 2013a). Lentiviral constructs used were Control vector expressing shRNA control [(SHC002); Sigma-Aldrich, Missouri, USA] and shRNA targeting SOX4 (TRCN0000018214, Sigma) for SOX4 KD1 used in Figure 6, and TRCN0000274152 for SOX4 KD2 used in Figure 6—figure supplement 1. Knockdown of EDN1 in HMEL cells was performed using 25 nM human EDN1 siRNA (Thermo Scientific, ON-TARGET plus SMARTpool, L-016692-00-0010, Waltham, MA, USA) with Lipofectamine RNAiMAX reagents (Invitrogen), 24 hr before 4-OHT treatment.

Statistics
Statistical analysis was performed using IBM SPSS Statistics version 21 (SPSS Inc., Chicago, IL, USA) for all patient data. Associations between categorical variables were examined using the Pearson's Chi-square test or Fisher's exact test when required. Survival analyses were performed using Kaplan-Meier and differences were analyzed using Log rank test. P-values <0.05 were considered to be statistically significant. Additional statistical analyses were performed using GraphPad software (v6), and statistical analyses used are indicated in the figure legends. Statistical analysis of the in vitro data was performed using the analysis of variance (ANOVA) with a Post Hoc Dunnett's multiple comparisons test, unless indicated differently. Gene signature analysis was performed by calculating Spearman-rank correlations between centroids of the SOX4 signature and patient gene expression profiles. ChIP-seq is represented as boxplots with 5–95% whiskers. Statistical comparison was performed with a two-tailed Mann-Whitney U test (**p<0.0001) for unpaired peak-sets and a two-tailed Wilcoxon-signed-rank test for paired peak-sets (****p<0.0001).

Survival analysis
The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) cohort includes DNA copy number, whole gene expression and breast cancer-specific survival of 1980 patients (data is freely available and can be found on cBioPortal; accession number EGAS000000000083) (Curtis et al., 2012). Claudin low subtyping was performed using transcriptomic data and the predictor developed by Prat et al. (Prat et al., 2010). Signature analysis was performed using the genes that were more than 1.5 log2 fold change differentially expressed between 4-OHT treated and untreated ERαSOX4 HMEL in the transcriptomic characterization and bound in SOX4 ChIP-seq analysis of DOX treated and untreated HMEL-S4. Samples were assigned a Spearman correlation-based score calculated between the gene expression data of each tumour and each signature. Survival was analysed using Kaplan-Meier curves and significance assessed by log-rank test. Data were analysed using R 3.2.2 and a two-sided p<0.05 was considered significant.

Patients
The study population was derived from the archives of the Departments of Pathology of the University Medical Center Utrecht, Utrecht, and the Radboud University Nijmegen Medical Center,
Nijmegen, The Netherlands. These comprised 452 cases of invasive breast cancer. Histological grade was assessed according to the modified Bloom and Richardson score (Elston and Ellis, 1991), and the mitotic activity index (MAI) was assessed as before (van der Groep et al., 2006). Other clinicopathological characteristics are shown in Table S3-7. From representative donor paraffin blocks of the primary tumors, tissue microarrays were constructed as described by Vermeulen et al. (Vermeulen et al., 2012b). The use of anonymous or coded leftover material for scientific purposes is part of the standard treatment contract with patients in The Netherlands (van Diest, 2002). Ethical approval was therefore not required. Overall survival and treatment data were obtained from the Comprehensive Cancer Center of The Netherlands (Integraal Kankercentrum Nederland). Survival data were available of 295 out of 452 breast cancer cases, with a follow up of 72 months for the ductal and 120 months for the lobular breast cancer cases.

**Immunohistochemistry**

Immunohistochemistry was carried out on 4 μm thick sections. Data on HER2, progesterone receptor (PR), estrogen receptor α (ERα), and E-cadherin were derived from Vermeulen et al. (Vermeulen et al., 2012a; Vermeulen et al., 2012b). After deparaffination and rehydration, endogenous peroxidase activity was blocked for 15 min in a buffer solution pH5.8 containing 0.3% hydrogen peroxide. After antigen retrieval, that is boiling for 20 min in 10 mM citrate pH6.0, a cooling period of 30 min preceded the primary antibody incubation. Primary antibodies against SOX4 (HPA029901, Sigma Aldrich) 1:50 were diluted in PBS containing 1% BSA and incubated overnight at 4°C. The signal was amplified using the Novolink kit (Leica, Rijswijk, The Netherlands) and developed with diaminobenzidine, followed by counterstaining with haematoxylin, dehydation in alcohol, and mounting. Appropriate negative and positive controls were used throughout.

All scoring was done blinded to patient characteristics and results of other stainings by three independent observers (SVJ, JFV, PJvD). E-cadherin and HER2 stainings were scored using the DAKO/HER2 scoring system for membrane staining. Membranous scores 1+, 2+, and 3 + were considered positive. For HER2 only a score of 3 + was considered positive. ERα and PR were scored by estimating the percentage of positive tumor cells, considering cancers with more than 10% positive tumor nuclei as positive. Nuclear and cytoplasmic SOX4 staining intensity was scored as 0–3+, considering samples with 3 + staining intensity as positive. The Perou/Sorlie molecular classification was simulated by ER/PR/HER2 as before (Kornegoor et al., 2012).

The CD34 staining (Immunotech, QBEND10) was performed in a 1:800 dilution on a Ventana autostainer. For xenograft tumors in Figure 6 immunohistochemistry was carried out on 4 μm thick sections and antigen retrieval was performed as described for patient material. Primary antibodies against SOX4 (HPA029901, Sigma Aldrich) CD31 (PECAM-1 Antibody (M-20), sc-1506) and ET-1 (N-8, SantaCruz, sc-21625) 1:50 were diluted in PBS containing 1% BSA and incubated overnight at 4°C. The signal was amplified using poly-HRP (Leica, Rijswijk, The Netherlands) and developed with diaminobenzidine, followed by counterstaining with haematoxylin, dehydration in alcohol, and mounting. For xenograft tumors in Figure 6—figure supplement 1 immunohistochemistry was carried out on 4 μm thick sections and antigen retrieval was performed as described for patient material. Primary antibodies against SOX4 (HPA029901, Sigma Aldrich) Endomucin (ab106100 abcam) and ET-1 (N-8, SantaCruz, sc-21625) were diluted in PBS containing 1% BSA and incubated overnight at 4°C. Note that the Endomucin antibody was used to identify blood vessels as the PECAM-1 antibody was commercially not available. The signal was amplified using the Novolink kit (Leica, Rijswijk, The Netherlands) and developed with diaminobenzidine, followed by counterstaining with haematoxylin, dehydration in alcohol, and mounting.

To quantify blood vessels in tumors, ImageJ software was used. To quantify CD34 (patient tumor material) and CD31 (murine model) 3,3′-Diaminobenzidine (DAB) staining or Endomucin staining. Positive signal was isolated using the ImageJ colour deconvolution plug-in, followed by a fixed threshold background subtraction and automatic particle analysis, using a minimum particle size of 50 pixels.

**Migration assay**

200,000 cells per well were plated 24 hr in advance in 24-well plate wells to form a confluent monolayer. Migration was assessed by making a single straight scratch in the HMEC-1 monolayer using a
20–200 µl pipette tip. Wells were washed with PBS to remove loose cells, and 0.5 ml test medium (basal-, full-, or conditioned medium) was added, followed by a 5 hr incubation at 37°C, 5% CO₂ and ambient oxygen level (20%). Images were recorded at t = 0 and t = 5 hr at similar locations, after which the scratch surface was determined using ImageJ software to determine scratch closure.

**Matrigel angiogenesis assay**

The effect of conditioned medium on angiogenic activity of HMEC-1 cells was determined by seeding 2,000 cells (in 10 µl basal medium) on to 10 µl solidified Growth Factor Reduced Matrigel (Becton Dickinson, United Kingdom) in a µ-Slide Angiogenesis (Ibidi, Germany) in the presence of 50 µl test medium (basal-, full-, or conditioned medium). After an 18 hr incubation at 37°C, images were recorded using an Olympus BX53 microscope with a DP71 digital camera, and analyzed for total network length using the ImageJ Angiogenesis Analyzer plugin.

**Measurement of angiogenic sprouting**

Angiogenic sprouting stimulation of conditioned medium was measured by seeding HMEC-1 cells on to Cytodex3 microcarrier beads (Sigma-Aldrich, USA) and embedding them in a 1:1:4 mixture of basal medium, test medium, and Growth Factor Reduced Matrigel (Becton Dickinson, USA). After solidification of the gels, 0.5 ml basal medium was added on top of the gels and incubated for 48 hr at 37°C, 5% CO₂ and ambient oxygen level (20%). Images were recorded using an Olympus CKX41 microscope with an SC30 digital camera. Sprout lengths were determined using ImageJ software.

**Measurement of proliferation**

100,000 HMEC-1 cells were seeded in full HMEC-1 medium in 24-well plate wells 24 hr before the assay. 0.4% BSA was added to all test media except full HMEC-1 medium. Test media was added to the wells, and the cells were incubated for 48 hr at 37°C, 5% CO₂ and 20% O₂. Then, cells were washed with PBS and isolated using 0.5 ml 0.25% Trypsin-EDTA (Gibco, Invitrogen, USA) for 5 min at 37°C, flushed loose with 0.5 ml full HMEC-1 medium, and counted using a TC20 automated cell counter (Bio-Rad, USA).

**Chromatin Immunoprecipitation**

Prior to chromatin immunoprecipitation (ChIP) and sequencing HMLE-S4 cells were grown in the absence or presence of 1 µg/ml of doxycycline for 16 hr, ChIP was performed as described previously (van Boxtel et al., 2013) using the following antibodies: 10 µg of SOX4 antibody (CS-129–100, Diagenode), 5 µg H3K4me3, H3K27ac, (ab8580, ab4729, Abcam), 5 µg anti-RBP1 (PB-7C2) antibody (Euromedex), and 5 µg H3K27me3 (39157, Active Motif). Truseq nano DNA sample preparation kit (Illumina) was used for End-repair, A-tailing, and ligation of sequence adaptors. Samples were subsequently PCR amplified after which the libraries were size selected to remove adapter dimers and select fragments in the 200–500 bp range. Barcoded libraries were sequenced on Illumina NextSeq500 sequencer (75 bp, single-end, Utrecht DNA sequencing facility).

**RNA sequencing**

Total RNA was extracted from ER and ERSOX4 HMLE cells treated with 100 nM 4-OHT for 8 hr, using the RNAeasy kit (Qiagen, Germany). Sample preparation was performed using TruSeq stranded total RNA with ribo-zero globin sample preparation kit (Illumina, USA) and samples were sequenced 75 bp single-end on Illumina NextSeq500 (Utrecht DNA sequencing facility).

**NGS data analysis**

BaseSpace software (Illumina) was used for sample demultiplexing, only reads with quality score of Q > 30 were subsequently used for further analysis. Differential gene expression analysis was performed using the Tophat2-cufflinks pipeline as described previously (Trapnell et al., 2012). Briefly, reads were mapped to the human reference genome (hg19) using TopHat v2.0.9 and using the reference gene annotation (hg19) as a guide transcripts were assembled using CuffLinks v2.2.1. Cuffdiff was used for differential gene analysis. Prior to differential gene analysis reads were quartile
normalized using the –library-norm quartile option and reads mapping to rRNA and tRNA were masked from the analysis (-M command). CummeRbund was used for quality assessment and figure generation. Cluster 3.0 and Java TreeView software version 1.1.6 were used for clustering and visualization of heatmaps. The DAVID gene ontology database (http://david.abcc.ncifcrf.gov/) and the ClueGO module in cytoscape were used for functional annotation of differentially expressed genes (Huang et al., 2009; Bindea et al., 2009).

ChIP-seq reads were mapped to the reference genome (hg19) with Bowtie 2.1.0 (Langmead and Salzberg, 2012) using default settings. Samtools version 0.1.19 was used for file conversions (SAM to BAM) and peaks were called using Cisgenome 2.0 using the input as a control (–e 150 -maxgap 200 –minlen 200) (Jiang et al., 2010). Mapped reads were extended according to the average fragment size and converted to TDF files with igvtools-2.3.36 and were visualized with IGV-2.3.34 (Thorvaldsdóttir et al., 2013). General manipulation of bed-files was performed using BEDtools v2.17.0 (Quinlan and Hall, 2010). Quantile normalization and quantification of ChIP-seq reads was performed as described previously (Eijkelenboom et al., 2013). HOMER software was used for motif discovery, peak annotation and the generation of histograms (Heinz et al., 2010).

Analysis of gene expression signatures
The log fold change values for each of the different gene signatures were used to create a centroid which was then correlated to the z-score transformed centroid for the corresponding genes for each patient sample in METABRIC using the Spearman correlation as described in Bruna et al. (Curtis et al., 2012; Bruna et al., 2012). The 25th, median or 75th percentile or median Spearman correlation value was used as a cut-off for the Kaplan-Meier analysis of breast cancer-specific death and survival for subgroups defined by ER status.

Zebrafish embryo maintenance, cell injection and imaging
Zebrafish maintenance and procedures were performed in the Faculty of Natural Sciences at Imperial College London (UK), in accordance to the UK Home office regulations (ASPA 1986). Zebrafish adult specimen were kept in a self-recirculating aquarium at an average temperature of 28°C with a 14 hr light 10 hr dark cycle. Adult specimens were fed twice a day on a diet of Hikari micropellets (Kyorin) and brine shrimp.

Zebrafish in vivo angiogenesis assays were performed on Tg(fli:GFP) embryos (Lawson and Weinstein, 2002). To facilitate visualization post-implantation, cell lines lacking intrinsic fluorescent signalling were pre-labeled with CM-DiI (Invitrogen, UK) according to the manufacturer’s instructions. For injections, labeled cells were harvested to obtain a pellet without medium and then resuspended in 10 µl PBS. 1 day post-fertilization (1dpf) embryos were manually dechorionated and then anesthetized in a solution containing 0.003% tricane (Sigma, UK) for 10 min. Injections were performed using a Narishige microinjector with a 12 mm gage borosilicate pipette. During the procedure, embryos were kept on an injection mould composed of 3% agar in a solution of pre-warmed PBS (+Mg + Ca) with the addition of 0.003% tricane (Sigma, UK). Approximately, 150–200 cells were injected in the yolk sac of each embryo, in proximity of the sub-intestinal vessel complex. For individual experiments, approximately 25 embryos were injected per cell line. Injections were completed within 1 hr, following which the embryos were maintained in a solution of system water (chlorine deprived tap and distilled water mixture) with the addition of 0.0003% (v/v) methylene blue (antifungal) and 30 µg/ml N-phenylthiourea (Sigma) (PTU, averts melanin formation), and kept at a constant temperature of 28°C. Live imaging was performed on individual embryos 2 days post-injection (dpi) under a wide-field fluorescent microscope (Olympus CKX41). Fish were anesthetized with the addition of 0.05% tricaine to their water, to prevent movement during the live imaging practice. All pictures were taken with Q Capture-Pro (QImaging) apparatus, with any alteration to the original picture performed via the use of Image J.

Orthotopic tumor xenograft model
Mammary fat-pad transplantations in female RAG2−/−;IL-2Rγ−/− immunodecient mice using approximately 1 × 106 luciferase-expressing MDA-MB-231 cells and bioluminescence imaging were performed as described previously (Ivanova et al., 2013). Tumour growth was measured weekly using
a digital caliper (VWR, Radnor, PA, USA) and mice were sacrificed when tumor volume exceeded 1000 mm$^2$ or visible metastases were detected by BLI. All animal experiments were approved by the Utrecht University Animal Experimental Committee (DEC-Utrecht no. 2013.III.02.020). Immediately after resection, the tumors were fixed in neutral buffered formalin and paraffin embedded. Immunohistochemistry was performed on 4 μm thick sequential sections, as described above. Macroscopic lung metastases were counted, ET-1 and CD31 staining was quantified using ImageJ as described above.

qPCR
For qPCR analysis, mRNA was isolated using the RNAeasy kit (Qiagen, Germany) according to the manufacturers’ protocol after which cDNA was generated using the iScript cDNA synthesis kit (Bio-Rad). Subsequently, real-time quantification was performed on a MyiQ Single-Color Real-Time PCR Detection System (Bio-Rad, USA) using SYBR Green Supermix (Bio-Rad, USA) for cDNA amplification according to the manufacturer’s protocol.

FW-SOX4 ACCGCACGCCAAGCTCATCC FW-LMO4 AAAAGCAGACCATGGTGAATCC
RV-SOX4 GTCCGGGCCTTGTACAGCGA RV-LMO4 CTGTGCAGCAACAACATCCAC
FW-B2M CCAGCAGAGAATGGAAAGTC FW-SMARCC1 CTGTGCAGCAACAACATCCAC
RV-B2M CACAGCAGAGAATGGAAAGTC RV-SMARCC1 GGGGGATACATGCCGTTAGG
FW-EDN1 TTGAGATCTGAGGAACCCGC FW-EC1 TGCCACACACACTGCAAAT
RV-EDN1 GTCAGCGCTAAGCTGACTGGT FW-ECE1 GTCAGCGCTAAGCTGACTGGT
FW-MARCKSL1 GGCTACAGAGCCATCCACTC FW-MMP10 GACACAGTGGCTCATGACTCC
RV-MARCKSL1 TGTACCTCAAGGAGACGACG RV-MMP10 TGTACCTCAAGGAGACGACG
FW-DNB1 GGAGTTCTTCCAGGGTGTCG FW-TEAD2 CCGCTCGAGAGTGTGGACGT
RV-DNB1 GGTCTTCAGGCTGGTGTTGCG RV-TEAD2 CCGCTCGAGAGTGTGGACGT
FW-MARCKS TGCCCCAGTTCTTCAAGACCGC FW-FZD7 ACAGAGCGCCAGGGAGAAAGC
RV-MARCKS GCCATTCTCCAGTCCATGCCTGCA RV-FZD7 CTCTCCACAGGAGAAAGC

Luciferase assays
For the luciferase assay, HEK293T cells at 50% confluence were transfected in 24-well plates with 0.1 μg EDN-1 promoter luciferase reporter or mutated versions thereof, with 0.1 μg of pcDNA3 empty vector, pcDNA3- flag-SOX4, pcDNA3 flag-SOX4 DN (aminoacids 1–135), and 0.02 μg pRL-TK Renilla (Promega) as a transfection control. The cells were lysed in 50 μl passive lysis buffer 3 days post-transfection, the soluble fraction was subsequently assayed for luciferase activity with a Dual-Luciferase Reporter Assay System (Promega, USA).

Oligonucleotide Pull-Down Assay
Oligonucleotide pull-down assays were performed as described previously (Vervoort et al., 2013a). Briefly, cell extracts of HEK239T cells transfected with either an empty vector control pcDNA or flag-SOX4 pcDNA were incubated with biotinylated double-stranded oligonucleotide probes, which were precoupled to streptavidin beads. Binding sites matching the consensus SOX4 binding sequence (Vervoort et al., 2013a) or the identified ET-1 promoter regions were used. After incubation, with the biotinylated probes the samples were washed in low-salt buffer (150 mM NaCl) and subsequently analyzed by western blot using an anti-flag antibody.

Continued on next page
ELISA

HMLE-ERSOX4 cells and HMLE-ER cells were seeded at 50,000 cell/well in a 6-well plates and were subsequently allowed to adhere overnight. The cells were subsequently treated or left untreated for 8 hr with 100 nM 4-hydroxytamoxifen (4-OHT) after which the media was replaced with HMEC-1 bare media for overnight conditioning in the presence or absence of 4-OHT. The next day the media was collected and cell-debris was removed by using a 20 μM filter. Samples were subsequently diluted 10x and analyzed for ET-1 expression using the Endothelin-1 Quantikine ELISA Kit (DET100, R and D systems) according to the manufacturer’s protocol.

Acknowledgements

The authors would like to thank Dr. RA Weinberg for the kind gift of HMLE cells, Hisham Mohammed and Jason Carroll for assistance with RIME assays. We would also like to thank J. van Loosdregt, R van Boxtel and J Beekman for helpful discussions. OGdJ is supported by The Netherlands Institute for Regenerative Medicine (NIRM, grant No. FES0908), MCV is supported by The Netherlands Organization for Scientific Research (Vidi grant 016.096.359) SJV, MGR, CLF and ARL are funded by the Dutch Cancer Foundation (UU-2013–5801 and UU-2015–7838). EFWL and LB are supported by grants from Breast Cancer Campaign (2012NovemberPhD016) and Cancer Research UK (A12011). Eric W-F Lam’s work is supported by MRC (MR/N012097/1), CRUK (C37/A12011;C37/A18784), Breast Cancer Now (2012MayPR070; 2012NovPhD016), the Cancer Research UK Imperial Centre, Imperial ECMC and NIHR Imperial BRC. Laura Bella was supported by Breast Cancer Now (2012NovPhD016).

Additional information

| Funding                     | Grant reference number | Author                  |
|-----------------------------|-------------------------|-------------------------|
| KWF Kankerbestrijding       | KWF UU 2013–5801        | Stephin J. Vervoort     |
|                             |                         | Ana Rita Lourenço       |
| Dutch Cancer Foundation     | UU-2013–5801            | Stephin J Vervoort      |
|                             |                         | M Guy Roukens           |
|                             |                         | Cynthia L Frederiks    |
|                             |                         | Ana Rita Lourenço       |
| The Netherlands Institute   | FES0908                 | Olivier G de Jong       |
| for Regenerative Medicine   |                         |                         |
| KWF Kankerbestrijding       | KWF UU 2015–7838        | Guy Roukens             |
|                             |                         | Cynthia L Frederiks    |
| Fundação para a Ciência e  |                         | Ana Rita Lourenço       |
| a Tecnologia                |                         | José L Sandoval         |
| Breast Cancer Campaign      | 2012NovemberPhD016      | Eric W-F Lam            |
|                             |                         | Laura Bella             |
| Cancer Research UK          | A12011                  | Eric W-F Lam            |
|                             |                         | Laura Bella             |
| Breast Cancer Now           | 2012NovPhD016           | Laura Bella             |
| Cancer Research UK          |                         | Alejandra Bruna         |
|                             |                         | Carlos Caldas           |
| Breast Cancer Campaign      |                         | Alejandra Bruna         |
|                             |                         | Carlos Caldas           |
The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions
Stephin J Vervoort, Cynthia L Frederiks, Conceptualization, Data curation, Formal analysis, Validation, Methodology, Writing—original draft; Olivier G de Jong, Jeroen F Vermeulen, Ana Rita Lourencô, Laura Bella, Ana Tufegdzic Vidakovic, José L Sandoval, Cathy Moelans, Miranda van Amersfoort, Margaret J Dallman, Edward Nieuwenhuis, Elsken van der Wall, Patrick Derksen, Paul van Diest, Michal Mokry, Conceptualization, Data curation, Formal analysis, Validation, Methodology; M Guy Roukens, Conceptualization, Data curation, Supervision, Funding acquisition, Methodology, Writing—review and editing; Alejandra Bruna, Carlos Caldas, Marianne C Verhaar, Eric W-F Lam, Conceptualization, Resources, Supervision, Investigation, Methodology, Writing—review and editing; Paul J Coffer, Conceptualization, Resources, Supervision, Funding acquisition, Investigation, Methodology, Writing—original draft, Project administration, Writing—review and editing

Author ORCIDs
Paul J Coffer http://orcid.org/0000-0001-8775-1939

Ethics
Human subjects: The use of anonymous or coded leftover material for scientific purposes is part of the standard treatment contract with patients in The Netherlands. Ethical approval was therefore not required.

Animal experimentation: All experiments were performed in accordance to international guidelines and approved by Experimental Animal Committee Utrecht (DEC-Utrecht, University Utrecht, Utrecht, The Netherlands).

Decision letter and Author response
Decision letter https://doi.org/10.7554/eLife.27706.035
Author response https://doi.org/10.7554/eLife.27706.036

Additional files
Supplementary files
• Transparent reporting form
DOI: https://doi.org/10.7554/eLife.27706.027

Data availability
All ChIPseq data and RNAseq data has been deposited to GEO (GSE104761).
The following datasets were generated:

| Author(s)               | Year | Dataset title                  | Dataset URL                          | Database and Identifier   |
|-------------------------|------|--------------------------------|--------------------------------------|---------------------------|
| SJ Vervoort, MG Roukens, PJ Coffer | 2018 | ChiP-seq HMLE vs HMLES4         | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104761 | Gene Expression Omnibus, GSE104761 |
| SJ Vervoort, MG Roukens, PJ Coffer | 2018 | ChiP-seq MDA-MB-231             | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104761 | Gene Expression Omnibus, GSE104761 |
| SJ Vervoort, MG Roukens, PJ Coffer | 2018 | ChiP-seq HC1954                 | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104761 | Gene Expression Omnibus, GSE104761 |
| SJ Vervoort, MG Roukens, PJ Coffer | 2018 | RNA-seq HMLE vs HMLES4          | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104761 | Gene Expression Omnibus, GSE104761 |

References

Ades EW, Candal FJ, Swerlick RA, George VG, Summers S, Bosse DC, Lawley TJ. 1992. HMEC-1: establishment of an immortalized human microvascular endothelial cell line. *Journal of Investigative Dermatology* **99**:683–690. DOI: https://doi.org/10.1111/1523-1747.ep12613748, PMID: 1361507

Bhattaram P, Penzo-Méndez A, Sock E, Colmenares C, Kaneko KJ, Vassilev A, Depampelis ML, Wegner M, Lefebvre V. 2010. Organogenesis relies on SoxC transcription factors for the survival of neural and mesenchymal progenitors. *Nature Communications* **1**:1–12. DOI: https://doi.org/10.1038/ncomms1008, PMID: 20596238

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**:1091–1093. DOI: https://doi.org/10.1093/bioinformatics/btp101, PMID: 19237447

Bos PD, Zhang XH, Nadal C, Shu W, Gomis RR, Nguyen DX, Minn AJ, van de Vijver MJ, Gerald WL, Foekens JA, Massagué J. 2009. Genes that mediate breast cancer metastasis to the brain. *Nature* **459**:1005–1009. DOI: https://doi.org/10.1038/nature08021, PMID: 19421193

Bruna A, Greenwood W, Le Quesne J, Teschendorff A, Miranda-Saavedra D, Rueda OM, Sandoval JL, Vidakovic AT, Saaâdi A, Pharoah P, Stingl J, Caldas C. 2012. TGFβ inhibits the formation of tumour-initiating cells in claudinlow breast cancer. *Nature Communications* **3**:1055. DOI: https://doi.org/10.1038/ncomms2039, PMID: 22968701

Buckanovich RJ, Facciabene A, Kim S, Benencia F, Saralos D, Balint K, Katsaros D, O’Brien-Jenkins A, Gimotty PA, Coukos G. 2008. Endothelin B receptor mediates the endothelial barrier to T cell homing to tumors and disables immune therapy. *Nature Medicine* **14**:28–36. DOI: https://doi.org/10.1038/nm1699, PMID: 18157142

Castillo SD, Matheu A, Mariani N, Carretero J, Lopez-Rios F, Lovell-Badge R, Sanchez-Cespedes M. 2012. Novel transcriptional targets of the SRY-HMG box transcription factor SOX4 link its expression to the development of small cell lung cancer. *Cancer Research* **72**:176–186. DOI: https://doi.org/10.1158/0008-5472.CAN-11-3506, PMID: 22084397

Clozel M, Breu V, Gray GA, Kalina B, Löffler BM, Burri K, Cassal JM, Hirth G, Müller M, Neidhart W. 1994. Pharmacological characterization of bosentan, a new potent orally active nonpeptide endothelin receptor antagonist. *The Journal of Pharmacology and Experimental Therapeutics* **270**:228–235. PMID: 8035319

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ho G, Haffari G, Bashashati A, Russell R, McKinney S, Langerod A, Green A, Provenzano E, Wishart G, et al. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**:346–352. DOI: https://doi.org/10.1038/nature10983, PMID: 22522925

Darnell JE. 2002. Transcription factors as targets for cancer therapy. *Nature Reviews Cancer* **2**:740–749. DOI: https://doi.org/10.1038/nrc906, PMID: 12360277

David CJ, Huang YH, Chen M, Su J, Zou Y, Bardeesy N, Iacobuzio-Donahue CA, Massagué J. 2016. TGF-β Tumor Suppression through a Lethal EMT. *Cell* **164**:1015–1030. DOI: https://doi.org/10.1016/j.cell.2016.01.009, PMID: 26898331

Eijkelenboom A, Mokry M, Smits LM, Nieuwenhuis EE, Burgering BM. 2013. FOXO3 selectively amplifies enhancer activity to establish target gene regulation. *Cell Reports* **5**:1664–1678. DOI: https://doi.org/10.1016/j.celrep.2013.11.031, PMID: 24360957

Elston CW, Ellis IO. 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* **19**:403–410. DOI: https://doi.org/10.1111/j.1365-2559.1991.tb00229.x, PMID: 1757079

Fan Y, Hseu YC, Lin YF, Hsing ST, Tai C, Uen YH, Lin KY. 2012. Clinical and prognostic association of transcription factor SOX4 in gastric cancer. *PLOS ONE* **7**:e52804. DOI: https://doi.org/10.1371/journal.pone.0052804, PMID: 23285187
Foronda M, Martínez P, Schoeftner S, Gómez-López G, Schneider R, Flores JM, Pisano DG, Blasco MA. 2014. Sox4 links tumor suppression to accelerated aging in mice by modulating stem cell activation. Cell Reports 8: 487–500. DOI: https://doi.org/10.1016/j.celrep.2014.06.031, PMID: 25043184

Harrell JC, Pfefferle AD, Zalies N, Prat A, Fan C, Khamtsov A, Olopade OI, Troester MA, Dudley AC, Perou CM. 2014. Endothelial-like properties of claudin-low breast cancer cells promote tumor vascular permeability and metastasis. Clinical & Experimental Metastasis 31:33–45. DOI: https://doi.org/10.1007/s10585-013-9607-4, PMID: 23975155

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular Cell 38:576–589. DOI: https://doi.org/10.1016/j.molcel.2010.05.004, PMID: 20513432

Hennessy BT, Gonzalez-Angulo AM, Stemke-Hale K, Gilcrease MZ, Krishnamurthy S, Lee JS, Fridlyand J, Sahin A, Agarwal R, Joy C, Liu W, Stivers D, Baggery K, Carey M, Lluch A, Monteagudo C, He X, Weigman V, Fan C, Calhoun J, et al. 2009. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. Cancer Research 69:4116–4124. DOI: https://doi.org/10.1158/0008-5472.CAN-09-3441, PMID: 19435916

Huang daW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4:44–57. DOI: https://doi.org/10.1038/nprot.2008.211, PMID: 19139156

Ivanova IA, Vermeulen JF, Ercan C, Houthuijzen JM, Saig FA, Vlug EJ, van der Wall E, van Diest PJ, Vooijs M, Derksen PW. 2013. FER kinase promotes breast cancer metastasis by regulating α6- and β1-integrin-dependent cell adhesion and anoikis resistance. Oncogene 32:5582–5592. DOI: https://doi.org/10.1038/onc.2013.277, PMID: 23873028

Jain RK. 2014. Antiangiogenesis strategies revisited: from starving tumors to alleviating hypoxia. Cancer Cell 26: 605–622. DOI: https://doi.org/10.1016/j.cccel.2014.06.006, PMID: 25517747

Jiang H, Wang F, Dyer NP, Wong WH. 2010. CisGenome Browser: a flexible tool for genomic data visualization. Bioinformatics 26:1781–1782. DOI: https://doi.org/10.1093/bioinformatics/btp286, PMID: 20513664

Kalluri R. 2009. EMT: when epithelial cells decide to become mesenchymal-like cells. Journal of Clinical Investigation 119:1417–1419. DOI: https://doi.org/10.1172/JCI39675, PMID: 19487817

Kamachi Y, Uchikawa M, Kondoh H. 2000. Pairing SOX off: with partners in the regulation of embryonic development. Trends in Genetics 16:182–187. DOI: https://doi.org/10.1016/S0168-9525(99)01955-1, PMID: 10729834

Kornegoor R, Verschuer-Maes AH, Buerger H, Hogenes MC, de Bruijn PC, Oudejans JJ, van der Groep P, Hinrichs B, van Diest PJ. 2012. Molecular subtyping of male breast cancer by immunohistochemistry. Modern Pathology 25:398–404. DOI: https://doi.org/10.1038/modpathol.2011.174, PMID: 22056953

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359. DOI: https://doi.org/10.1038/nmeth.1923, PMID: 22388286

Lawson ND, Weinstein BM. 2002. In vivo imaging of embryonic vascular development using transgenic zebrafish. Developmental Biology 248:307–318. DOI: https://doi.org/10.1006/dbio.2002.0711, PMID: 12167406

Liao YL, Sun YM, Chau FY, Chau YP, Lai TC, Wang JL, Horng JT, Hsiao M, Tsou AP. 2008. Identification of SOX4 target genes using phylogenetic footprinting-based prediction from expression microarrays suggests that overexpression of SOX4 potentiates metastasis in hepatocellular carcinoma. Oncogene 27:5578–5589. DOI: https://doi.org/10.1038/onc.2008.168, PMID: 18504433

Lin CM, Fang CL, Hseu YC, Chen CL, Wang JW, Hsu SL, Tu MD, Hung ST, Tai C, Uen YH, Lin KY. 2013. Clinical and Prognostic Implications of Transcription Factor SOX4 in Patients with Colon Cancer. PLOS ONE 8:e67128. DOI: https://doi.org/10.1371/journal.pone.0067128, PMID: 23826209

Lourenço AR, Coffer PJ. 2017. SOX4: joining the master regulators of epithelial-to-mesenchyme transition? Trends in Cancer 3:571–582. DOI: https://doi.org/10.1016/j.trecan.2017.06.002, PMID: 28780934

Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Oshen AB, Gerald WL, Massagué J. 2005. Genes that mediate breast cancer metastasis to lung. Nature 436:518–524. DOI: https://doi.org/10.1038/nature03799, PMID: 16049480

Mohammed H, D’Santos C, Serandour AA, Ali HR, Brown GD, Atkins A, Rueda OM, Holmes KA, Theodorou V, Robinson JL, Zwart W, Saadai A, Ross-Innes CS, Chin SF, Menon S, Stingl J, Palmieri C, Caldas C, Carroll JS. 2013. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. Cell Reports 3:342–349. DOI: https://doi.org/10.1016/j.celrep.2013.03.010, PMID: 23403292

Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X, Perou CM. 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Research 12:R68. DOI: https://doi.org/10.1186/bcr2635, PMID: 20813035

Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M, Perou CM. 2013. Characterization of cell lines derived from breast cancers and normal mammary mammary for the study of the intrinsic molecular subtypes. Breast Cancer Research and Treatment 142:237–255. DOI: https://doi.org/10.1007/s10549-013-2743-3, PMID: 24162158

Quinlan AR, Perou CM. 2011. Deconstructing the molecular portraits of breast cancer. Molecular Oncology 5:5–23. DOI: https://doi.org/10.1016/j.molonc.2010.11.003, PMID: 21147047

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. DOI: https://doi.org/10.1093/bioinformatics/btp033, PMID: 20110278
Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. 2004. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. PNAS 101:9309–9314. DOI: https://doi.org/10.1073/pnas.0401994101, PMID: 15184677

Rosano L, Spinaella F, Di Castro V, Nicotra MR, Dedhar S, de Herreros AG, Natali PG, Bagnato A. 2005. Endothelin-1 promotes epithelial-to-mesenchymal transition in human ovarian cancer cells. Cancer Research 65: 1149–1157. DOI: https://doi.org/10.1158/0008-5472.CAN-05-2123, PMID: 16357176

Rosano L, Spinaella F, Bagnato A. 2013. Endothelin 1 in cancer: biological implications and therapeutic opportunities. Nature Reviews Cancer 13:637–651. DOI: https://doi.org/10.1038/nrc3546, PMID: 23884378

Salani D, Tarabotelli G, Rosano L, Di Castro V, Borsotti P, Giavazzi R, Bagnato A. 2000. Endothelin-1 induces an angiogenic phenotype in cultured endothelial cells and stimulates neovascularization in vivo. The American Journal of Pathology 157:1703–1711. DOI: https://doi.org/10.1016/S0002-9440(99)44080-9, PMID: 11073829

Sarkar A, Hochedlinger K. 2013. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell 12:15–30. DOI: https://doi.org/10.1016/j.stem.2012.12.007, PMID: 23290134

Scharer CD, McCabe CD, Ali-Sayed M, Berger MF, Bulyk ML, Moreno CS. 2009. Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells. Cancer Research 69:709–717. DOI: https://doi.org/10.1158/0008-5472.CAN-08-3415, PMID: 19147588

Schilham MW, Oosterwegel MA, Moerper, Ya J, de Boer PA, van de Wetering M, Verbeek S, Lamers WH, Kruisbeek AM, Cumanu A, Clevers H. 1996. Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. Nature 380:711–714. DOI: https://doi.org/10.1038/380711a0, PMID: 8614465

Taube JH, Herschkowitz JJ, Komurov K, Zhou AY, Gong S, Yang J, Hartwell K, Onder TT, Gupta PB, Evans KW, Hollier BG, Ram PT, Lander ES, Rosen JM, Weinberg RA, Mani SA. 2010. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metastatic breast cancer subtypes. PNAS 107:15449–15454. DOI: https://doi.org/10.1073/pnas.1004900107, PMID: 20713713

Tavazoie SF, Alarcon C, Oskarsson T, Padua D, Wang Q, Bos PD, Gerald WL, Massague J. 2008. Endogenous human microRNAs that suppress breast cancer metastasis. Nature 451:147–152. DOI: https://doi.org/10.1038/nature06487, PMID: 18185580

Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14:178–192. DOI: https://doi.org/10.1093/bib/bbs017, PMID: 22517427

Tiwari N, Tiwari VK, Waldmeier L, Balwierz PJ, Arnold P, Pachkov M, Meyer-Schaller N, Schübsler D, van Nimwegen E, Christofori G. 2013. Sox4 is a master regulator of epithelial-mesenchymal transition by controlling Ezh2 expression and epigenetic reprogramming. Cancer Cell 23:768–783. DOI: https://doi.org/10.1016/j.ccr.2013.04.020, PMID: 23764001

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Racher L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7:562–578. DOI: https://doi.org/10.1038/nprot.2012.016, PMID: 22383036

van Balkom BW, de Jong OG, Smits M, Brummelman J, den Ouden K, de Bree PM, van Eijndhoven MA, Pegtel DM, Stoorvogel W, Würdinger T, Verhaar MC. 2013. Endothelial cells require miR-214 to secrete exosomes that suppress senescence and induce angiogenesis in human and mouse endothelial cells. Blood 121:3997–4006. DOI: https://doi.org/10.1182/blood-2013-02-479925, PMID: 23532734

van Boxtel R, Gomez-Puerto C, Mokry M, Eijkelenboom A, van der Vos KE, Nieuwenhuis EE, Burgering BM, Lam EW, Coffer PJ. 2013. FOXP1 acts through a negative feedback loop to suppress FOXO-induced apoptosis. Cell Death & Differentiation 20:1219–1229. DOI: https://doi.org/10.1038/cdd.2013.81, PMID: 23832113

den van der Groep P, Bouter A, van der Zanden R, Sicamara I, Monko FH, Gille JJ, van Kalen C, van der Wall E, Verheijen RH, van Diest PJ. 2006. Distinction between hereditary and sporadic breast cancer on the basis of clinicopathological data. Journal of Clinical Pathology 59:611–617. DOI: https://doi.org/10.1136/jcp.2005.032151, PMID: 16603649

van Diest PJ. 2002. No consent should be needed for using leftover body material for scientific purposes. The BMJ 325:648–651. DOI: https://doi.org/10.1136/bmj.325.7365.648, PMID: 12241820

Vermeulen JF, van Brussel AS, van der Groep P, Morsink FH, Bult P, van der Wall E, van Diest PJ. 2012a. Immunophenotyping invasive breast cancer: paving the road for molecular imaging. BMC Cancer 12:240. DOI: https://doi.org/10.1186/1471-2407-12-240, PMID: 22695343

Vermeulen JF, van de Ven RA, Ercan C, van der Groep P, van der Wall E, Bult P, Christgen M, Lehmann U, Daniel J, van Diest PJ, Derksen PW. 2012b. Nuclear Kaiso expression is associated with high grade and triple-negative invasive breast cancer. PLOS ONE 7:e37864. DOI: https://doi.org/10.1371/journal.pone.0037864, PMID: 22662240

Vervoort SJ, van Boxtel R, Coffer PJ. 2013a. The role of SRY-related HMG box transcription factor 4 (SOX4) in tumorigenesis and metastasis: friend or foe? Oncogene 32:3397–3409. DOI: https://doi.org/10.1038/onc.2012.506, PMID: 23246969

Vervoort SJ, Lourenço AR, van Boxtel R, Coffer PJ. 2013b. SOX4 mediates TGF-β-induced expression of mesenchymal markers during mammary cell epithelial to mesenchymal transition. PLOS ONE 8:e53238. DOI: https://doi.org/10.1371/journal.pone.0053238, PMID: 23301049

Wang L, Zhang J, Yang X, Chang YW, Qi M, Zhou Z, Zhang J, Han B. 2013. SOX4 is associated with poor prognosis in prostate cancer and promotes epithelial-mesenchymal transition in vitro. Prostate Cancer and Prostatic Diseases 16:301–307. DOI: https://doi.org/10.1038/pcan.2013.25, PMID: 23917306
Wilson M, Koopman P. 2002. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. *Current Opinion in Genetics & Development* **12**:441–446. DOI: https://doi.org/10.1016/S0959-437X(02)00323-4, PMID: 12100890

Wülfing P, Diallo R, Kersting C, Wülfing C, Poremba C, Rody A, Greb RR, Böcker W, Kiesel L. 2003. Expression of endothelin-1, endothelin-A, and endothelin-B receptor in human breast cancer and correlation with long-term follow-up. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* **9**:4125–4131. PMID: 14519635

Wülfing P, Kersting C, Tio J, Fischer RJ, Wülfing C, Poremba C, Diallo R, Böcker W, Kiesel L. 2004. Endothelin-1-, endothelin-A-, and endothelin-B-receptor expression is correlated with vascular endothelial growth factor expression and angiogenesis in breast cancer. *Clinical Cancer Research* **10**:2393–2400. DOI: https://doi.org/10.1158/1078-0432.CCR-03-0115, PMID: 15073116

Yeh YM, Chuang CM, Chao KC, Wang LH. 2013. MicroRNA-138 suppresses ovarian cancer cell invasion and metastasis by targeting SOX4 and HIF-1α. *International Journal of Cancer* **133**:867–878. DOI: https://doi.org/10.1002/ijc.28086, PMID: 23389731

Zhang J, Liang Q, Lei Y, Yao M, Li L, Gao X, Feng J, Zhang Y, Gao H, Liu DX, Lu J, Huang B. 2012. SOX4 induces epithelial-mesenchymal transition and contributes to breast cancer progression. *Cancer Research* **72**:4597–4608. DOI: https://doi.org/10.1158/0008-5472.CAN-12-1045, PMID: 22787120

Zhou Y, Wang X, Huang Y, Chen Y, Zhao G, Yao Q, Jin C, Huang Y, Liu X, Li G. 2015. Down-regulated SOX4 expression suppresses cell proliferation, metastasis and induces apoptosis in Xuanwei female lung cancer patients. *Journal of Cellular Biochemistry* **116**:1007–1018. DOI: https://doi.org/10.1002/jcb.25055, PMID: 25565486