SCIRT lncRNA restrains tumorigenesis by opposing transcriptional programs of tumor-initiating cells

Sladjana Zagorac[1,*], Alex de Giorgio[1,*], Aleksandra Dabrowska[1], Mark Kalisz[2], Nuria Casas-Vila[3], Paul Cathcart[1], Angela Yiu[1], Silvia Ottaviani[1], Neta Degani[4], Ylenia Lombardo[1,5], Alistair Tweedie[6], Tracy Nissan[6], Keith W. Vance[7], Igor Ulitsky[4], Justin Stebbing[1,†] and Leandro Castellano[1,6,†,#]

Affiliations

[1] Department of Surgery and Cancer, Division of Cancer, Imperial College London, Imperial Centre for Translational and Experimental Medicine (ICTEM), London, W12 0NN, UK.

[2] Epithelial Carcinogenesis Group, Centro Nacional de Investigaciones Oncológicas, 28029 Madrid, Spain.

[3] Institute of Molecular Biology (IMB), Ackermannweg 4, 55128, Mainz, Germany.

[4] Department of Biological Regulation, The Weizmann Institute of Science, Rehovot, Israel.

[5] Nature Communications, The Macmillan Campus, 4 Crinan Street, London, N1 9XW, UK.

[6] University of Sussex, School of life Sciences, John Maynard Smith Building, Falmer, Brighton, BN1 9QG, UK.

[7] Department of Biology and Biochemistry, University of Bath, Bath, UK.

*Joint first authors

[†]Joint senior authors

Running Title

SCIRT is a novel lncRNA that regulates cancer stemness

Corresponding author

[#]Leandro Castellano,
University of Sussex,
School of life Sciences,
John Maynard Smith Building,
Brighton, BN1 9QG, UK.
Email: L.Castellano@sussex.ac.uk
Tel: +44(0)12 7367 8382

Competing interests

JS and LC declare their competing interests at https://www.nature.com/onc/editors none of which are relevant here. AdG is an employee of BenevolentAI. YL is an editor at Nature Communications. No other authors declare a conflict.

Keywords: LncRNA, self-renewal, cell-cycle, epigenetics, PRC2, CpG islands, Breast Cancer

Abstract

In many tumors, cells transition reversibly between slow-proliferating tumor-initiating cells (TIC) and their differentiated, faster-growing progeny. Yet how transcriptional regulation of cell cycle and self-renewal genes is orchestrated during these conversions remains unclear. In this study, we show that as breast TIC form, a decrease in cell-cycle and increase in self-renewal gene expression is coregulated by SOX2 and EZH2, which colocalize at CpG islands. This pattern was negatively controlled by a novel long non-coding RNA (lncRNA) that we name SCIRT, which was markedly upregulated in tumorspheres but colocalized with and counteracted EZH2 and SOX2 during cell cycle and self-renewal regulation to restrain tumorigenesis. SCIRT specifically interacted with EZH2 to increase EZH2 affinity to FOXM1 without binding the latter. In this manner, SCIRT induced transcription at cell cycle gene promoters by recruiting FOXM1 through EZH2 to antagonize EZH2-mediated effects at target genes. Conversely, on stemness genes, FOXM1 was absent and SCIRT antagonized EZH2 and SOX2 activity, balancing towards repression. These data suggest that the interaction of a lncRNA with EZH2 can alter the affinity of EZH2 for its protein binding partners to regulate cancer cell state transitions.

Statement of Significance

Findings show that a novel lncRNA SCIRT counteracts breast tumorigenesis by opposing transcriptional networks associated with cell cycle and self-renewal.

Introduction

The clonal model of tumor growth has been revised by the discovery of heterogeneous, tissue-like organization of many cancer types (1,2). In this view, cancers form hierarchies of tumor-initiating cells (TICs), which give rise to differentiated cells with limited proliferative potential (3). TICs can self-renew, divide indefinitely and produce differentiated cells within the tumor mass, generating cell states that create intra-tumor heterogeneity (3). TICs are highly metastatic and being slow-proliferating are resistant to chemotherapy, two properties linked to treatment failure and relapse (4). These risks have sparked interest in the potential for TIC-targeting therapies. However, strategies for eliminating TIC populations could be complicated by the dynamic equilibrium that exists between TIC and non-TIC states (3), suggesting a need to target both compartments simultaneously.

The self-renewal capacity that drives the formation and expansion of TICs has parallels to that in embryonic stem cells, with the involvement of pluripotency-associated transcription factors (TFs) such as SOX2 and chromatin modifiers such as EZH2 (5,6). Interestingly, SOX2 and EZH2 also promote cancer cell plasticity in prostate cancer by inducing expression of neuroendocrine markers which promote metastasis and anti-androgen resistance (7). In addition, EZH2 can display oncogenic Polycomb-independent functions to regulate transcription through binding specific TFs (8,9).

Certain surface markers can prospectively isolate TIC populations from tumors or cell lines (10). However, the specificity of these markers for TIC populations, especially in triple-negative breast cancer (TNBC), is imperfect (11). As a result, retrospective approaches to enrich for TICs have been developed, which use three-dimensional (3D) culturing conditions and low plating densities to form clonal cultures of TIC-enriched tumorspheres (spheres)

(12,13). By culturing spheres, we have studied the regulatory transcriptional networks that drive TIC formation in breast tumors.

While transcription factors (TFs) controlling self-renewal of TICs have been partially characterized (14), factors that regulate plasticity between TIC and non-TIC compartments remain unknown. We hypothesized that factors driving TIC formation could be counteracted by negative feedback loops keeping cells in a poised state to facilitate easy cell state transitions, with consequences for cancer treatment.

Here we demonstrate that SOX2 and EZH2 directly repress cell-cycle gene transcription and activate self-renewal by recognizing CpG islands in breast cancer cells. We further show that a previously undescribed lncRNA counteracts these processes, affecting these regulatory networks by negative feedback. Further understanding the regulatory dynamics by which lncRNAs control cell plasticity may aid the development of novel therapies that may not only target the existing TICs but also prevent the dynamic conversion of TICs to non-TICs and *vice versa.*

Materials and Methods

Mammalian cell culture

Breast cancer cell lines MDA-MB-231, MCF-7, SKBR3, T47D, MDA-MB-468, BT549, MDA-MB-453 and BT474 were obtained from ATCC (Manassas, VA, USA). MDA-MB-231 and MCF-7 cells were grown in Dulbecco's Modified Eagle Medium (Sigma), SKBR3 in McCoy's 5a Medium Modified (Sigma), T47D and BT549 in RPMI-1640 Medium (Sigma), MDA-MB-468 and MDA-MB-453 in Leibovitz's L-15 Medium (Sigma) and BT474 in DMEM/F12 Medium (Gibco), supplemented with 10% fetal calf serum, 2 mM L-glutamine, 100 U ml−1 penicillin,

and 100 mg ml−1 streptomycin. Between thawing and the use in the described experiments, the cells were passaged no more than five times. Routinely, the state of cells was checked for cellular morphology and compared with images from ATCC website. All cell lines were monthly tested for mycoplasma (MycoAlert, Lonza) and were always found negative.

Sphere culture (tumorspheres)

MDA-MB-231, MCF-7, SKBR3, T47D, MDA-MB-468, BT549, MDA-MB-453 and BT474 cells were plated in single-cell suspension in ultra-low attachment plates (Corning, # CLS3471). Cells were grown in serum-free DMEM/F12 medium (Gibco) supplemented with B27 (1:50, Gibco), 20 ng ml−1 basic fibroblast grown factor (bFGF, Biolegend) and 20 ng ml−1 epidermal grow factor (EGF, Sigma). Tumorspheres were collected after 16 hours or after 5 days of sphere formation. For sphere formation assay breast cancer cells were plated in ultra-low attachment plates at a density of $2x10^3$ for MDA-MD-231 and $1x10^3$ for MCF-7 cells/well and formed spheres with a size larger than 75 µm were counted under the microscope. Percentage of sphere formation efficiency (SFE%) was calculated as a ratio between the number of formed spheres divided by the number of cells seeded, multiplied by 100.

Transfections

Silencer Select siRNAs were purchased from Ambion. Cells were transfected using Lipofectamine RNAiMAX (Invitrogen) following the manufacturer's recommended protocol. Unless otherwise specified 25nM of siRNAs were transfected for 48h or 96h. siRNA sequences or catalog number can be found in Supplementary Table 8, sheet #1.

RNA isolation and RT-qPCR assays

Total RNA from cultured cells was extracted using TRI Reagent (Sigma) following the manufacturer's instructions including DNase I treatment. For gene expression, cDNA was synthesized from 1 µg of purified DNase-treated RNA using RevertAid^{TM} M-MuLV reverse transcriptase and random hexamer primers (Thermo Scientific), according to the manufacturer's protocols. qRT-PCRs were performed on a StepOne™ Real-Time PCR System using Fast SYBR® Green Master Mix (both from Applied Biosystems).

We calculated the transcript copy number by using a previously published protocol (15,16). Briefly, RNA relative copy numbers were determined by RT-qPCR using standard curves and normalized to β-actin levels. The primer sequences used are reported in Supplementary Table 8, sheet #2.

Primary tumor preparation

Fresh primary samples were collected from Charing Cross hospital within 1 hour of operating and retrieved in DMEM medium on ice. Samples were cut into pieces <1mm using a scalpel, washed once in DMEM medium, then proteolytically digested for 1.5-2 hours in 5ml of medium containing proteolytic enzymes (hyaluronidase 100U/ml, collagenase 3000U/ml). Once cells had appropriately detached from the extracellular matrix (assessed by checking cells under a haemocytometer), they were centrifuged at 200 x g for 10 mins at room temperature, with supernatant carefully removed and cells resuspended in full medium. Tumor cells were then sorted using Magnet Activated Cell Sorting (MACS) to deplete extraneous cell types, beads specific for blood cells (Lineage Depletion Kit, Human, MACS) and fibroblasts (fibroblast depletion kit, human, MACS).

Protein extraction and western blotting

Cells were harvested in RIPA buffer (Sigma-Aldrich, St. Louis, MI) supplemented with a protease inhibitor cocktail (Roche Applied Science, Indianapolis, IN). Cell lysates were centrifuged at 14,000 rpm and the supernatant was collected. Protein lysates were quantified using a BCA Protein Assay Reagent kit (Pierce, Thermo Scientific, Waltham, MA). Fifty micrograms of lysates were resolved by SDS-PAGE and transferred to nitrocellulose membranes (Amersham, Piscataway, NJ). After blocking in 1X PBS containing 5% (w/v) milk and 0.1% Tween20 (v/v), membranes were incubated with the specific antibodies overnight at 4°C. Membranes were then washed three times with 1X PBS containing 0.1% Tween20 (v/v), incubated with horseradish peroxidase-conjugated goat anti-rabbit or goat anti-mouse antibody (Sigma-Aldrich), and washed again to remove unbound antibodies. Bound antibody complexes were detected with SuperSignal chemiluminescent substrate (GE Healthcare, Little Chalfont, UK). Antibodies used for WB can be found in Supplementary Table 8, sheet #6.

Statistical analysis

Results for continuous variables are presented as means ± standard deviation unless stated otherwise and significance was determined using the Mann-Whitney test using GraphPad Prisma 8 (GraphPad Software, La Jolla California USA) or R (https://www.r-project.org/). Expression values and statistical analysis for differential gene expression studies were performed with DESeq2 from the Bioconductor or by using GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/). Significance for overlapping genes was computed by using a Hypergeometric test. Significant enrichment for SCIRT binding on promoters was determined using CEAS (http://liulab.dfci.harvard.edu/CEAS/). Significant differential expression of SCIRT, FOXM1, EZH2 and SOX2 in tumor, normal and metastatic samples from TCGA and GTEx datasets as well as correlation analyzes were obtained by

applying two-sided Welch's t-test and Pearson correlation respectively and computed using R (https://www.r-project.org/). Logrank *p*-values for survival analyzes were computed with KMplot or cBioPortal.

Differences were considered significant when p-values or p-adjusted-values were < 0.05.

Accession number

The dataset supporting the conclusions of this article is available in the Gene Expression Omnibus SuperSeries (GSE136195).

Results

Gene expression change during sphere formation reflects the induction of self-renewal and reduction of cell-cycle

To evaluate how gene expression changes during sphere formation in aggressive breast cancer, we cultured TNBC MDA-MB-231 cells in adherent (adh) and sphere conditions for 16 hours (h) and 5 days (d) and performed RNA-seq (Figure 1a and Supplementary Table 1). We selected MDA-MB-231 (and later MCF-7 for further validation) because these cell lines have been previously reported as good models to study breast cancer stem cells in a study that compared cell lines with primary tissues (17). We chose an early time point (16h) to specifically detect early gene expression changes during sphere formation without confounding factors, such as alterations of the geometry within the spheres or limited access to nutrients within the cultures. A five day time point was chosen to identify genes common between nascent and mature spheres. After bioinformatic analysis, we identified 2559 and 2636 genes that were up-regulated in spheres grown for 16h and 5d respectively with 1482 up-regulated in common. Furthermore, we uncovered that 2344 and 2856 genes

were down-regulated in spheres grown for 16h and 5d respectively, whereas 1547 were down-regulated in common (Figure 1a and Supplementary Table 1, *p*-adjusted < 0.01). Pathways enriched for genes up- and down-regulated in spheres grown for 16h and 5d are shown in Supplementary Table 2. We validated top up- and down-regulated genes by RT-qPCR (Figure S1a). As expected, transcripts coding for proteins known to be involved in stem cell renewal, including KLF4, NOTCH1 and TGFB1, were significantly up-regulated after 16h or 5d of sphere growth (Figure S1b and Supplementary Table 1, *p*-adjusted < 0.01), supporting the biological relevance of our model system and analysis. By using the Enrichr tool (18), we found that genes up-regulated in sphere culture at both time-points were significantly enriched in PI3K-Akt and TGFβ signaling pathways (Figure S1c, *p*-adjusted < 0.01), compared to cells grown in adh conditions. Each pathway has been shown to induce breast cancer stem cell formation (19,20). Importantly, analysis of the ChEA database (21), using Enrichr, indicated that ZNF217, KDM2B, SOX2 were the top enriched TFs or chromatin-modifying enzymes that bind within the promoters of observed up-regulated genes in spheres (Figure S1c, *p*-adjusted < 0.01). Each one of these factors has been identified as a strong regulator of breast cancer stem cell self-renewal and tumorigenesis (5,22,23). In summary, these results confirm that genes induced during sphere growth are enriched for tumorigenic pathways and regulated by tumorigenic TFs.

Interestingly, transcripts significantly down-regulated at 16h and 5d of sphere growth were strongly enriched for cell-cycle promoting genes, and frequently contained binding sites of two master transcriptional regulators of cell-cycle, FOXM1 or E2F4, within their promoters (Figure S1d *p*-adjusted < 0.01). Overall, these effects indicate concomitant reductions of proliferative and increases of self-renewal gene expression in breast cancer spheres during the adh to sphere transition.

SCIRT lncRNA is up-regulated in spheres (16h and 5d) but counteracts stemness

Next, we hypothesized that lncRNAs could regulate transcriptional dynamics observed during sphere formation and breast cancer tumorigenesis. To explore undescribed lncRNAs that could be involved in this process, we searched for long non-coding transcripts that were up- or down-regulated in spheres after both 16h and 5d, and displaying a degree of cross-species conservation, using a previously established pipeline (24). Seven lncRNA candidates presented such characteristics (Figure 1b, *p*-adjusted < 0.01). We focused on RP5-1120P11.1 (also annotated as LOC101929705 or AL109615.3) because it was strikingly up-regulated in spheres (Figure 1b and Supplementary Table 1, *p*-adjusted < 0.01) but surprisingly, its depletion strongly induced sphere formation in both MDA-MB-231 and MCF-7 breast cancer cells (Figure 1c, d), suggesting that it may act through negative feedback. We named this lncRNA Stem Cell Inhibitory RNA Transcript (SCIRT).

Our RNA-seq data indicated several SCIRT isoforms including RP5-1120P11.1 (Figure S2a). To precisely map the TSS of SCIRT and to understand which one of these two isoforms is most expressed in breast cancer, we analyzed public Cap Analysis of Gene Expression sequencing (CAGE-seq) data from the ENCODE project (https://www.encodeproject.org/), performed in MCF-7 breast cancer cell lines (Figure. S2b). RP5-1120P11.1 and isoform n3 present a CAGE-seq signal overlapping with their common TSSs in MCF-7 cells, reinforcing the hypothesis that the transcription of SCIRT starts from this region. Within the SCIRT locus, another undescribed transcript is annotated on the opposite strand of SCIRT (C6orf223), however, C6orf223 does not appear to be expressed in our system (Figure S2a). Next, we observed that SCIRT was also up-regulated in spheres derived from primary breast cancer specimens (Figure 2a) and several other breast cancer cell lines (Figure S2c) except BT549, maybe due to genetic or epigenetic defects accumulated in these cell lines. Additionally, we observed

that SCIRT expression levels were heterogeneous in different breast cancer cell lines (Figure S2c) but were significantly higher in more aggressive basal-like primary tumors compared to luminal breast cancer subtypes in cells derived from primary tumors (Figure 2a).

Breast cancer cells in which SCIRT expression was depleted by two independent siRNAs (siRNA knockdown efficiency >90%, Figure S2d) showed increases in sphere formation over two passages (Figure 2b). In contrast, over-expression of SCIRT in MDA-MB-231 cells (Figure S2e) decreases sphere formation efficiency (Figure S2f). These results indicate that SCIRT RNA restrains breast cancer self-renewal capacity. Additionally, when MDA-MB-231 spheres with stable down-regulation of SCIRT (shSCIRT) were injected subcutaneously into the flanks of immunocompromised mice, tumor formation was enhanced, suggesting that SCIRT can also reduce breast cancer tumorigenesis *in vivo* (Figure 2c). SCIRT silencing also significantly increased directional cell migration and cell speed (Figure S3a, b), in aggregate indicating that SCIRT opposes breast cancer progression.

SCIRT is a chromatin-associated lncRNA that down-regulates self-renewal genes and induces cell-cycle genes

To understand the mechanism by which SCIRT opposes stemness we first looked at its subcellular localization (Figure 2d, e and Figure S3c, d). RNA FISH showed that SCIRT was exclusively located into the nucleus of MCF-7 breast cancer cells (Figure 2d) and that its signal was strongly reduced in cells treated with SCIRT siRNA (siSCIRT). Subcellular fractionation followed by RT-qPCR indicated that SCIRT mainly localizes in the chromatin-associated compartment of spheres derived from TNBC MDA-MB-231 (Figure 2e) and MDA-MB-468 (Figure S3c) cells. Additionally, analysis of RNA-seq from MCF-7 fractions (25) also showed chromatin localization of SCIRT in these cells (Figure S3b).

Next, to identify genes regulated by SCIRT, we performed RNA-seq from MDA-MB-231 spheres where we depleted SCIRT using two different siRNAs. SCIRT silencing (siSCIRT#1 and #2 overlap) increased the expression of 653 and decreased the expression of 768 genes (*p*-adjusted < 0.05, Wald Test; Figure S4a and Supplementary Table 3). Interestingly, SCIRT regulation of gene expression recapitulated pathway enrichment observed for genes modulated in spheres versus adh cells but in the opposite direction. Accordingly, genes induced by SCIRT (down-regulated upon SCIRT depletion) were strongly enriched in cell-cycle-related signaling and mitosis whereas genes repressed by SCIRT (up-regulated upon SCIRT silencing) were strongly enriched in stem cell expansion as well as neuronal functions (Figure 3a and Figure S4a-c). This indicates that SCIRT silencing increases sphere formation (Figure 1c, d and Figure 2b) because this lncRNA acts by repressing self-renewal and by activating cell-cycle gene expression, operating in negative feedback to fine-tune separate gene expression programs.

SCIRT globally binds to promoter or enhancer regions to increase expression of cell-cycle genes and decrease expression of self-renewal genes

Since SCIRT is an abundant chromatin-associated lncRNA (Figure 2e and Figure S3c, d) with expression levels comparable to TFs that are active in breast cancer, such as SOX2 (Supplementary Table 3), it may regulate gene expression by interacting with several specific chromatin loci. To evaluate this possibility, we performed Capture Hybridization Analysis of RNA targets (CHART) with DNA-sequencing (26) to identify chromatin regions bound by SCIRT. Pulldown with SCIRT probes showed strong enrichment (20-25-fold) of the SCIRT transcript compared to DNA oligonucleotides complementary to control LacZ sequence (from *Escherichia coli*, frequently used as control sequence for this kind of

experiment (26-28)) (Figure 3b), indicating that we enriched for specific endogenous regions of SCIRT. Moreover, SCIRT probes did not enrich for MALAT1 or GAPDH transcripts used as negative control RNAs (Figure 3b). Following CHART, we used massively parallel sequencing (CHART-seq) to identify all the chromatin regions that are associated with SCIRT. Similar to the previous experiment (Figure 3b), regions bound by SCIRT were depleted of LacZ peaks (Figure 3c and Figure S5a). Moreover, SCIRT chromatin binding regions show strong PhastCons signal (given by evolutionary conserved elements in a multiple alignment) that peaks at their center (Figure S5b), implying that SCIRT interacts with conserved *cis*-regulative regions around the genome. Conversely, LacZ peaks do not occur at conserved regions (Figure S5b). By calculating the fold-change of SCIRT signal versus input control, followed by metagene profiling, we observed that SCIRT peaks occur in the proximity of transcriptional start sites (TSSs) and are depleted at transcriptional termination sites (TTSs) (Figure S5c). Interestingly, the highest levels of SCIRT binding, considering the ratio of observed over expected values, were at promoters and CpG islands (Figure S5d; $p < 2.2e-16$, One-sided binomial test). Using the HOMER peak caller (29), we detected 15,999 significant peaks for SCIRT (false discovery rate (FDR) < 0.001), which were common between two independent experiments (Supplementary Table 4). We observed that 12,724 of these peaks were associated with a total of 7,130 genes, being present within 200 kilobases (Kbs) from their TSSs (Figure 3d). We found a significant overlap between the genes up- or down-regulated following SCIRT depletion and the genes associated with the SCIRT peaks (Figure 3d, $p < 2.2e-16$, Hypergeometric test, 1.9-fold higher than expected by chance for both up-regulated and down-regulated genes), suggesting the association of SCIRT with chromatin as having a functional role in the regulation of those genes. Twenty-three percent of SCIRT peaks were located within promoters of genes regulated by SCIRT, as

shown by co-localization with promoter-specific H3K4me3 marks (Figure S5e). Additionally, 43% of the peaks were located within enhancers, as shown by co-localization with enhancer-specific H3K4me1 histone modifications (Figure S5e) and 32% in undefined positions. In both promoters and enhancers, SCIRT peaked exactly at the center of the valley formed by the histone modifications peaks (Figure S5f) indicating that it may interact with proteins within these locations. This indicates that SCIRT can increase or repress gene transcription through association with promoters or enhancers. We also observed that transcripts up-regulated by siSCIRT were significantly enriched for genes involved in TGFβ and PIK3-Akt pathways and had pluripotency factors (KDM2B, SOX2, SOX9) enriched within their promoters (from ChEA), (Figure 3d, bottom-right). On the other hand, transcripts down-regulated by siSCIRT were significantly enriched for cell-cycle and DNA replication pathways and had promoters that were enriched for TFs that induce cell-cycle (FOXM1, E2F4, E2F7) (Figure 3d, bottom-left).

SCIRT interacts with EZH2 to antagonize its Polycomb-independent activity.

By analyzing public ChIP-seq databases (ChEA), we showed that KDM2B represents the most enriched factor interacting significantly with the promoters of genes up-regulated by siSCIRT and associated with SCIRT peaks (Figure 3d bottom-right). KDM2B maintains pluripotency of stem cells by recruiting PRC1 to CpG islands which co-localize with PRC2 (30). Several studies have suggested that nuclear lncRNAs can interact with PRC2 to promote transcriptional gene silencing (31,32). To evaluate whether SCIRT acts through a similar mechanism, we performed RNA immunoprecipitation (RIP)-RT-qPCR for SCIRT, using an antibody recognizing the PRC2 catalytic component EZH2 as it has been shown that RNA preferentially binds PRC2 proximally to its methyltransferase center (33). In line with our

hypothesis, SCIRT is associated with EZH2 in spheres formed from both MCF-7 and MDA-MB-231 cells (Figure S6a). We then used streptavidin-binding S1 aptamers fused to four SCIRT partially overlapping fragments (F1-F4) to pull down interacting proteins *in vitro* followed by immunoblotting for EZH2 protein, to evaluate the region of SCIRT where EZH2 binds. We observed EZH2 interacts preferentially with the 5' half of SCIRT (Figure S6b; F1 and F2) as predicted by catRAPID (http://service.tartaglialab.com/page/catrapid_group) (Figure S6c and d). Based on this algorithm, EZH2 interacts with SCIRT at position 226-277 or 626-677 (Figure S6d). Interestingly, G4hunter (http://bioinformatics.ibp.cz:8888/#/analyze/quadruplex) predicted a G-quadruplex structure at position 214-264 of SCIRT. Because it has been shown that EZH2 specifically binds G-quadruplex structures present on RNA transcripts (34), this suggests that EZH2 specifically binds SCIRT by recognizing a G4-quadruplex structure in its 5' region.

Next, to evaluate the functionality of the SCIRT-EZH2 interaction, we silenced EZH2 (siEZH2) and performed RNA-seq in spheres to assess whether EZH2 regulates the same transcripts modulated by SCIRT (Supplementary Table 3 and Supplementary Table 5). Intriguingly, a significant fraction of genes up-regulated by siEZH2 were also down-regulated by siSCIRT (Figure 4a, $p < 2.2e\text{-}16$, hypergeometric test; fold enrichment = 7.6) suggesting that SCIRT counteracts EZH2-mediated gene activation in breast TICs. In contrast, genes down-regulated by siEZH2 significantly overlap with genes up-regulated by siSCIRT (Figure 4b, $p = 1.01e\text{-}13$, hypergeometric test; fold enrichment = 4.1). These data suggest that SCIRT may interact directly with EZH2 to antagonize both its repressive and activating functions. Next, we wondered whether SCIRT would change the profile of H3K27me3, a histone mark associated with gene repression induced by the EZH2/PRC2 complex, on genes coregulated

by SCIRT and EZH2. To assess this, we performed ChIP-seq of H3K27me3 in cells growing in 3D conditions after siSCIRT, siEZH2 or control siRNA (siNC) treatment and assessed its levels across all genes as well as SCIRT controlled genes, including promoters (Figure 4c-e). Although as expected, H3K27me3 signal was elevated throughout the bodies of genes that are not expressed, but absent at active genes (Figure 4c), H3K27me3 global profile did not change upon either siEZH2 or siSCIRT treatment (Figure 4c). More importantly, H3K27me3 levels were low and did not change upon siSCIRT or siEZH2 treatment, either for SCIRT-regulated genes or their promoters. (Figure 4d, e). Lack of change in H3K27me3 global levels following transient EZH2 inhibition is in line with previous observations (35) and it is likely due to compensation from EZH1 (36). Based on these results, we propose that gene expression change observed following siEZH2 and siSCIRT is likely to be Polycomb-independent.

SCIRT interacts with chromatin loci and acts through FOXM1, EZH2 and SOX2

We observed that genes directly activated by SCIRT were enriched for cell-cycle and mitotic genes mostly controlled by FOXM1 (Figure 3d, bottom-left) and those down-regulated by SCIRT were enriched for self-renewal, TGFβ and PIK3-Akt signaling, transcriptionally controlled by SOX2 or other pluripotency controlling factors (Figure 3d, bottom-right). As it has been shown in adh breast cancer cells (37), we observed that FOXM1 also activates the expression of mitotic genes in cells growing in spheres (Figure S6e). This led us to hypothesize that SCIRT could increase transcription of selected cell-cycle genes by recruiting FOXM1 to their promoters or enhancers, counteracting the Polycomb-independent repression of these genes exerted by EZH2. Accordingly, depletion of EZH2 (siEZH2) in spheres led to an increase in the expression of cell-cycle genes that are both targets of SCIRT

and FOXM1 (Figure S6f). At the same time, SCIRT could decrease transcription of self-renewal genes by forming a complex with EZH2 and SOX2 and counteracting their activating effect on those promoters (Figure S6g, h). In support of these hypotheses, FOXM1, EZH2 and SOX2 were all necessary for the effect of SCIRT on sphere formation for both MDA-MB-231 (Figure 4f) and MCF-7 cells (Figure S6i), as silencing of any of these factors could prevent increases in sphere growth following down-regulation of SCIRT (Figure 4f and Figure S6i). Effects seen on sphere formation appear additive for EZH2 and SOX2, suggesting they may act independently (Figure 4f). Additionally, we observed that the copy number of SCIRT transcripts per sample was similar to the copy number of EZH2 or SOX2 mRNAs (Figure S6j: i, ii and iii), indicating that they reach stoichiometric ratios and that SCIRT is an abundant lncRNA able to interact with several genomic regions together with these factors. To assess this effect at the gene level, we performed ChIP-seq for EZH2 and SOX2 in spheres, re-analyzing publicly available ChIP-seq data for FOXM1 (GSE40762) as well as H3K4me1 and H3K4me3 (GSE124379) from MDA-MB-231 cells and integrating these data with our SCIRT-CHART-seq and our ChIP-seq from H3K27me3. K-means clustering of these peaks indicated that SCIRT, FOXM1, EZH2 and SOX2 mostly colocalize close to TSSs of a fraction of genes regulated by SCIRT (Figure 5a). However, SOX2 and EZH2 colocalized with SCIRT at promoters of genes up-regulated upon siSCIRT treatment, as anticipated (Figure 5b, c and Figure S6a), but surprisingly also at promoters of cell-cycle genes down-regulated upon SCIRT knockdown together with FOXM1 (Figure 5b, d and Figure S6b). Intriguingly, this indicates that in addition to activating genes involved in self-renewal (5) (Figure S6g), SOX2 can also regulate the transcription of cell-cycle genes that are targets of FOXM1 and SCIRT during sphere formation. In line with this, depleting SOX2 using siRNAs (siSOX2) increased the expression of a set of genes involved in cell-cycle (Figure 5e), which are also increased

by siEZH2 (Figure S6f) but reduced by both siSCIRT and siFOXM1 (Figure 6f, Figure S6e and Supplementary Table 3). As expected, genes increased by siSCIRT (Supplementary Table 3) were down-regulated by siSOX2 and siEZH2 (Figure S6g, h). In aggregate, these data suggest that SCIRT, EZH2 and SOX2 colocalize at promoters of their target genes but that SCIRT antagonizes EZH2 and SOX2 in the transcriptional regulation of those genes.

In embryonic stem cells, the PRC2 complex and consequently H3K27me3, colocalize at promoters of developmental regulators with the pluripotency TFs SOX2, NANOG and OCT4 (38). Importantly, in mammals, PRC2 is enriched at genomic regions with high GC content, which are enriched with CpG islands (39). By applying K-means clustering in CHART-seq and ChIP-seq peaks at promoters of SCIRT regulated genes we showed that EZH2 and SOX2 colocalize next to SCIRT peaks and exactly on CpG islands (CGIs) (Figure S7a). However, these two proteins not only colocalized within promoters of genes regulated by SCIRT or CGIs but much more broadly at multiple locations with high GC content, (Figure S7b and Figure S8a, b). Accordingly, SOX2 and EZH2 global binding profiles showed a striking positive correlation of R^2=0.91 (Spearman's rank correlation, Figure S7b), and both TFs showed a strong positive correlation with GC% (Figure S8b; EZH2~GC% = 0.86; SOX2~GC% = 0.77).

SCIRT increases cell cycle gene expression by binding EZH2 to induce EZH2-FOXM1 interaction within promoters

By specifically looking at promoters of genes regulated by SCIRT, and by sorting regions for up- and down-regulated genes, we found that FOXM1 mostly colocalized with SCIRT on promoters of genes down-regulated by siSCIRT, that are enriched for cell-cycle and have a CHR motif (Figure 5b). Importantly, promoters bound to FOXM1 and SCIRT were strongly depleted of H3K27me3 signal (Figure 5b and Figure S8), indicating that the

SCIRT/FOXM1/EZH2 acts on gene transcription independently of PRC2 activity. Further supporting this, the promoter of CCNA2, a direct target of both FOXM1 and SCIRT (Figure S8b and Supplementary Table 3), was enriched for EZH2 but not by SUZ12 binding (Figure S9a). These data also indicate that in spheres, EZH2 activity is Polycomb-independent, regulating a subset of genes that are directly controlled by SCIRT/FOXM1/EZH2.

Interestingly, RIP-RT-qPCR for SCIRT after using an antibody against FOXM1 indicates that FOXM1 also interacts with SCIRT (Figure S9b), but, suppression of FOXM1 did not reduce EZH2 binding with SCIRT (Figure 6a). When we performed RIP after UV crosslinking (X_RIP), which only detects direct RNA-protein interactions, we found that only EZH2, not FOXM1, was able to form an RNA-protein complex with SCIRT (Figure S9c, d) indicating that FOXM1 binds to SCIRT indirectly. Next, by using co-IP, we found that FOXM1 physically interacted with EZH2 in spheres formed by MCF-7 or MDA-MB-231 (Figure 6b), but depletion of SCIRT completely disrupted this interaction (Figure 6c left and right panels).

Mechanistically, although siSCIRT treatment for 48h did not affect FOXM1 protein levels (Figure 6d), it reduced FOXM1 (Figure 6e) as well as Pol II and H3K27ac levels at promoters of cell-cycle genes (Figure S10a, b). The effect of siSCIRT on FOXM1 recruitment at cell-cycle gene promoters was stronger in MDA-MB-231 than MCF-7 cells (Figure 6e). This is likely due to the higher expression levels of SCIRT in MDA-MB-231 compared to MCF-7. This suggested that cell-cycle genes are repressed by EZH2 (Figure S6f), but that SCIRT activates their transcription by recruiting FOXM1, Pol II and histone acetyltransferases to promoters of cell-cycle genes to counteract the slow proliferation of spheres, and induce differentiation. In aggregate, we revealed that SCIRT interacts with EZH2 to promote EZH2-FOXM1 protein-protein interactions at promoters of cell-cycle genes to constrain the transcriptional

repression exerted by EZH2 on these genes. This fine-tunes their expression during TIC formation.

Because these genes are repressed during TIC formation whereas SCIRT is up-regulated, SCIRT acts in a negative feedback loop to attenuate this transcriptional response.

Single-cell RNA-seq analysis shows SCIRT is mostly present in primary breast cancer cells that express pluripotent TFs

Our results point to the role of SCIRT in counteracting stemness of TICs. It is thought that TICs are aggressive self-renewing tumor cells that are controlled by pluripotent TFs (40). To evaluate whether SCIRT is co-expressed with pluripotency TFs in breast cancer cells we re-analyzed a public single-cell RNA-seq (scRNA-seq) study (GSE75688) performed in 515 cells isolated from 11 primary human breast cancer cells and representing all the 4 breast cancer subtypes (Luminal A, Luminal B, HER2+ and TNBC) (41) (Figure S11). We identified two different HER2+ subtypes that cluster with two different patients (cluster 1 and 6, Figure S11a) and cluster 6 had SOX2 as discriminative marker (Figure S11b, c and Supplementary Table 6, *p*-adjusted < 0.01). Interestingly, the HER2+ breast cancer subtype represented by cluster 6 (Figure S11a), contained the highest proportion of cells expressing SCIRT and SCIRT was not expressed in stromal cells (Figure S11b, c). In this tumor, cells expressing SCIRT were fewer (20%) but had a similar average expression levels than SOX2 (Figure S11c), indicating that both can reach stoichiometric ratio in primary tumor cells as well as in cell lines. These data indicated that this HER2+ subtype that has cells with high levels of SCIRT is less differentiated than the second (cluster 1) and probably more aggressive. In line with this hypothesis, this HER2+ subtype had also a high expression of additional pluripotent TFs, such as KLF4, EZH2 as well as ZNF217 which are strongly involved in breast cancer stemness

(22) (Figure S11b, c) in addition to SOX2. Finally, specific gene markers that characterize this tumor (Supplementary Table 6, *p*-adjusted < 0.01) were mostly regulated by ZNF217, SOX2 and FOXM1 transcription factors (Figure S11d) confirming that SCIRT is more expressed in less differentiated/more aggressive tumors. POU5F1 that encodes for OCT4 was only detected in a very small number of cells (Figure S11b,c), despite its documented role in breast cancer stemness and tumorigenesis (42), probably because its expression levels are mostly below the limit of detection of this scRNA-seq experiment.

SCIRT is up-regulated in breast cancer specimens but its high expression is associated with good prognosis

Next, we evaluated the clinical significance of this SCIRT-controlled transcriptional regulatory network. We first measured SCIRT, FOXM1, EZH2 and SOX2 expression from RNA-seq data obtained from 1391 specimens from the TCGA and GTEx cohorts and related expression levels of these genes to available TCGA clinical data. Similar to FOXM1 and EZH2, SCIRT was significantly more expressed in breast cancers compared to normal tissues (Welch's t-test, $p < 2.2e-16$) (Figure S12 a-c) and showed the highest expression levels in the basal and Her2+ subtypes which represent more aggressive tumor types (Figure S12d-f). We observed similar trends in cells derived from primary tumors (Figure 2a and Figure S11b, c). However, differences detected for both FOXM1 and EZH2 were generally more dramatic and less variable than those observed for SCIRT (Figure S12a-f). Surprisingly, SOX2 levels showed a different pattern of expression in breast cancers compared to SCIRT, FOXM1 and EZH2 (Figure S12 g), suggesting that the role of SOX2 in SCIRT/EZH2/FOXM1 regulatory network is context-specific and can occur only in rarer TICs. In general, SCIRT levels correlated positively with both EZH2 and FOXM1 (Figure S13a, b) but did not correlate with

SOX2 in either cancer or normal samples (Figure S13c). However, FOXM1 correlated better with EZH2 than SOX2 (Figure S13d, e).

Strikingly, while high expression levels of FOXM1 and EZH2 were associated with worse disease-free survival (DFS), high expression levels of SCIRT were associated with better DFS (Figure S14a). These results are in line with our findings: that SCIRT is co-expressed and co-regulates gene expression with oncogenic EZH2 and FOXM1 but, in the opposite direction, restraining tumorigenesis if up-regulated in tumorigenic cells.

Next, anticipating that genes that are repressed by SCIRT (up-regulated by siSCIRT) are oncogenic, we selected the top 100 genes with greatest up-regulation upon depletion of SCIRT by 2 independent siRNAs (Supplementary Table 7) and evaluated their expression levels, clinical relevance and mutational landscape in both the TCGA (43) and the METABRIC breast cancer datasets (44) interrogating a total of 3334 breast cancer samples (Figure S14b). These genes were up-regulated in almost all breast cancer specimens (Figure S14c), probably due to genome amplification at the genomic loci containing these genes (Figure S14d). Finally, breast cancer samples presenting amplification in these loci (Figure S14e) and samples with high expression of these genes (Figure S14f) presented worse overall survival.

In summary, we describe a novel lncRNA named SCIRT to be up-regulated in TICs in breast cancer but to counteract tumorigenesis despite its increased expression. This suggests that re-expression of SCIRT or inhibition of genes that are strongly repressed by SCIRT (Figure S14b) in patients, could represent a useful therapeutic approach to tackle the dynamic process of TIC self-renewal and differentiation in breast cancer. We also find that genes repressed by SCIRT are frequently amplified in breast cancer. Taken together this gene set

can be used as a novel signature to stratify the disease with diagnostic and prognostic implications.

Discussion

Using an lncRNA that we describe for the first time, SCIRT, we expand our understanding of the transcriptional regulation of stemness and proliferative expression programs active in breast TICs. EZH2 and SOX2 are transcription factors able to increase self-renewal capacity of both embryonic stem cells and cancer stem cells (6,45,46). By using a breast cancer tumorsphere formation assay, which enriches for slow-proliferating cancer stem cells or TICs, we observed that these two factors widely colocalize within CG-rich chromatin regions, including CpG islands located within promoters and when they interact with specific promoters, they repress cell-cycle gene expression and activate self-renewal and neuronal gene programs in breast cancer cells that grow in 3D conditions, indicating that SOX2 and EZH2 are also crucial factors that instruct TICs to proliferate slowly.

We further observed that this transcriptional program regulated by SOX2 and EZH2 in breast tumor does not occur unopposed but it is counteracted by SCIRT, strongly up-regulated during tumorsphere formation, but unexpectedly, functioning by counteracting EZH2 and SOX2's effects on the transcription of many self-renewal and cell cycle genes. By acting in this negative feedback loop, SCIRT increases cell-cycle and represses self-renewal transcriptional programs of these cells (Figure S15). Our data are in line with the hypothesis that SCIRT acts as a regulator that only reduces transcription of genes involved in tumorigenesis without fully repressing their activity. In this manner SCIRT tends to be more expressed in tumors than normal cells, but when it is expressed is associated with a more favorable prognosis.

We show for the first time that during sphere growth EZH2 binds to cell-cycle gene promoters to act as a co-repressor with its binding partner FOXM1, restraining its effects on transcription. SCIRT directly interacts with EZH2 to increase EZH2-FOXM1 interaction and recruit more FOXM1 to promoters to fine-tune this cell-cycle transcriptional program (Figure S15). Previously, it was reported that FOXM1 and EZH2 interact during hypoxia in breast cancer to activate MMP2 and MMP7 transcription (47), suggesting that SCIRT may mediate the interaction between FOXM1 and EZH2 and their transcriptional effects during hypoxia as well. However, to the best of our knowledge, the importance of the EZH2-FOXM1 antagonistic interaction for the regulation of cell-cycle gene transcription has not been described elsewhere. Our results suggest that this effect mediated by SCIRT is specifically related to an EZH2-FOXM1 interaction which seems to be independent of the PRC2 complex. When selected lncRNAs interact with EZH2, it appears they may change EZH2's affinity for its protein binding partners. It would be interesting to evaluate whether the binding of selected lncRNAs to EZH2 also changes PRC2 complex composition and activity rather than its recruitment to chromatin as widely described, or whether this action may be mediated by SCIRT alone in other cellular contexts.

Unexpectedly, we also found that SOX2, a TF that promotes self-renewal gene expression in TICs (5), also colocalized with EZH2 on promoters of cell-cycle genes to repress their transcription. Yet transcription from these promoters occurs during sphere growth, due to the dominant effects of SCIRT and FOXM1. Supporting this novel role for SOX2 in cell-cycle regulation, alongside its established function in TIC self-renewal, SOX2 has been reported to repress pro-proliferative cell-cycle genes in cortical progenitors to maintain their slow proliferative state (48).

Although we demonstrate that the interaction between FOXM1 and EZH2 is modulated by SCIRT, FOXM1 is mostly absent from the promoters of self-renewal genes up-regulated upon SCIRT depletion. We suggest that FOXM1 is not present within these genomic positions despite the presence of SCIRT and EZH2, because these genomic sites do not have a CHR motif (Figure 5b) which is the FOXM1 consensus sites essential for its DNA binding (49). Additionally or alternatively, a higher level of GC content or (consequently) high levels of EZH2/SOX2 interactions within these promoters, may prevent FOXM1's DNA binding. In support of this latter view, we observed FOXM1 binding site peaks in promoters of genes down-regulated by SCIRT that are adjacent to EZH2 and SOX2 binding sites (Figure S7a and Figure S8a) indicating that FOXM1 cannot interact with DNA in regions occupied by EZH2 and SOX2, but resides next to them. As reported by others in embryonic stem cells, we did not find a physical interaction between EZH2-SOX2 in breast cancer, suggesting that their co-localization depends on DNA rather than protein-protein interaction (50).

In aggregate, based on these data we propose that SCIRT acts as a tumor suppressor in breast cancer, despite appearing to be more expressed in TICs compared to their more differentiated counterparts and in tumors compared to healthy control tissues. Indeed, despite being strongly co-expressed with FOXM1 and EZH2, high expression of FOXM1 and EZH2 in clinical samples predicts poor DFS, but DFS is better in patients with high levels of SCIRT supporting the idea that SCIRT is induced in tumorigenic cells but counteracts their aggressive properties (Figure 2 and Figure S2e, f).

Future cancer treatments may include TIC and differentiated cell-targeting components. Inducing SCIRT or SCIRT-like factors that promote differentiation towards chemotherapy-vulnerable cell states, could enhance the success of such future approaches.

Acknowledgments

The authors thank Action Against Cancer (AAC) (grant numbers: P75997_WSCC and PF9671). The Searle Memorial Trust, Charles and Diane Herlinger, Alessandro Dusi, Julian and Cat O' Dell, Sofiya Machulskaya and Jackie McCarthy for funding this study. This work used the computing resources of the UK MEDical BIOinformatics partnership—aggregation, integration, visualization, and analysis of large, complex data (UK MED-BIO) which is supported by the Medical Research Council. This study was supported by the Sussex University and the Imperial BRC and ECMC.

Author contributions

LC conceived the project, supervised and designed research and performed the bioinformatic analyzes. LC, SZ, and AdG wrote the manuscript with inputs from MK, SO, ND, IU and JS. SZ and AdG performed most of the experimental work. IU and ND performed the initial RNA-seq analysis and prioritize important lncRNAs for further study. TN designed and supervised the FISH experiment. AD, MK, NC-V, PC, AY, YL, TN, AT and SO performed experiments. KWV supervised the CHART-seq study. JS co-supervised research and provided necessary reagents. All authors approved the final version submitted.

References

1. Torres CM, Biran A, Burney MJ, Patel H, Henser-Brownhill T, Cohen AS*, et al.* The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. Science **2016**;353
2. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J*, et al.* Tumour evolution inferred by single-cell sequencing. Nature **2011**;472:90-4

3. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. Nature **2013**;501:328-37
4. Lytle NK, Barber AG, Reya T. Stem cell fate in cancer growth, progression and therapy resistance. Nat Rev Cancer **2018**;18:669-80
5. Domenici G, Aurrekoetxea-Rodriguez I, Simoes BM, Rabano M, Lee SY, Millan JS*, et al.* A Sox2-Sox9 signalling axis maintains human breast luminal progenitor and breast cancer stem cells. Oncogene **2019**
6. Chang CJ, Yang JY, Xia W, Chen CT, Xie X, Chao CH*, et al.* EZH2 promotes expansion of breast tumor initiating cells through activation of RAF1-beta-catenin signaling. Cancer Cell **2011**;19:86-100
7. Ku SY, Rosario S, Wang Y, Mu P, Seshadri M, Goodrich ZW*, et al.* Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. Science **2017**;355:78-83
8. Lee ST, Li Z, Wu Z, Aau M, Guan P, Karuturi RK*, et al.* Context-specific regulation of NF-kappaB target gene expression by EZH2 in breast cancers. Mol Cell **2011**;43:798-810
9. Xu K, Wu ZJ, Groner AC, He HH, Cai C, Lis RT*, et al.* EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. Science **2012**;338:1465-9
10. Dontu G, Al-Hajj M, Abdallah WM, Clarke MF, Wicha MS. Stem cells in normal breast development and breast cancer. Cell Prolif **2003**;36 Suppl 1:59-72
11. Leth-Larsen R, Terp MG, Christensen AG, Elias D, Kuhlwein T, Jensen ON*, et al.* Functional heterogeneity within the CD44 high human breast cancer stem cell-like compartment reveals a gene signature predictive of distant metastasis. Mol Med **2012**;18:1109-21
12. Ponti D, Costa A, Zaffaroni N, Pratesi G, Petrangolini G, Coradini D*, et al.* Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. Cancer Res **2005**;65:5506-11
13. Lombardo Y, de Giorgio A, Coombes CR, Stebbing J, Castellano L. Mammosphere formation assay from human breast cancer tissues and cell lines. J Vis Exp **2015**
14. Mathieu J, Zhang Z, Zhou W, Wang AJ, Heddleston JM, Pinna CM*, et al.* HIF induces human embryonic stem cell markers in cancer cells. Cancer Res **2011**;71:4640-52
15. Zagorac S, Alcala S, Fernandez Bayon G, Bou Kheir T, Schoenhals M, Gonzalez-Neira A*, et al.* DNMT1 Inhibition Reprograms Pancreatic Cancer Stem Cells via Upregulation of the miR-17-92 Cluster. Cancer Res **2016**;76:4546-58
16. D'Errico G, Alonso-Nocelo M, Vallespinos M, Hermann PC, Alcala S, Garcia CP*, et al.* Tumor-associated macrophage-secreted 14-3-3zeta signals via AXL to promote pancreatic cancer chemoresistance. Oncogene **2019**;38:5469-85
17. Wang R, Lv Q, Meng W, Tan Q, Zhang S, Mo X*, et al.* Comparison of mammosphere formation from breast cancer cell lines and primary breast tumors. J Thorac Dis **2014**;6:829-37
18. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV*, et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics **2013**;14:128
19. Bruna A, Greenwood W, Le Quesne J, Teschendorff A, Miranda-Saavedra D, Rueda OM*, et al.* TGFbeta induces the formation of tumour-initiating cells in claudinlow breast cancer. Nat Commun **2012**;3:1055

20. Singh JK, Farnie G, Bundred NJ, Simoes BM, Shergill A, Landberg G*, et al.* Targeting CXCR1/2 significantly reduces breast cancer stem cell activity and increases the efficacy of inhibiting HER2 via HER2-dependent and -independent mechanisms. Clin Cancer Res **2013**;19:643-56
21. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics **2010**;26:2438-44
22. Vendrell JA, Thollet A, Nguyen NT, Ghayad SE, Vinot S, Bieche I*, et al.* ZNF217 is a marker of poor prognosis in breast cancer that drives epithelial-mesenchymal transition and invasion. Cancer Res **2012**;72:3593-606
23. Kottakis F, Foltopoulou P, Sanidas I, Keller P, Wronski A, Dake BT*, et al.* NDY1/KDM2B functions as a master regulator of polycomb complexes and controls self-renewal of breast cancer stem cells. Cancer Res **2014**;74:3935-46
24. Hezroni H, Koppstein D, Schwartz MG, Avrutin A, Bartel DP, Ulitsky I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep **2015**;11:1110-22
25. Sun M, Gadad SS, Kim DS, Kraus WL. Discovery, Annotation, and Functional Analysis of Long Noncoding RNAs Controlling Cell-Cycle Gene Expression and Proliferation in Breast Cancer Cells. Mol Cell **2015**;59:698-711
26. Vance KW, Sansom SN, Lee S, Chalei V, Kong L, Cooper SE*, et al.* The long non-coding RNA Paupar regulates the expression of both local and distal genes. EMBO J **2014**;33:296-311
27. Grossi E, Raimondi I, Goni E, Gonzalez J, Marchese FP, Chapaprieta V*, et al.* A lncRNA-SWI/SNF complex crosstalk controls transcriptional activation at specific promoter regions. Nat Commun **2020**;11:936
28. Zhang Q, Chao TC, Patil VS, Qin Y, Tiwari SK, Chiou J*, et al.* The long noncoding RNA ROCKI regulates inflammatory gene expression. EMBO J **2019**;38
29. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P*, et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell **2010**;38:576-89
30. He J, Shen L, Wan M, Taranova O, Wu H, Zhang Y. Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat Cell Biol **2013**;15:373-84
31. Davidovich C, Cech TR. The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA **2015**;21:2007-22
32. Vance KW. Mapping Long Noncoding RNA Chromatin Occupancy Using Capture Hybridization Analysis of RNA Targets (CHART). Methods Mol Biol **2017**;1468:39-50
33. Zhang Q, McKenzie NJ, Warneford-Thomson R, Gail EH, Flanigan SF, Owen BM*, et al.* RNA exploits an exposed regulatory site to inhibit the enzymatic activity of PRC2. Nat Struct Mol Biol **2019**;26:237-47
34. Wrighton KH. G-tracts give PRC2 the boot. Nat Rev Mol Cell Biol **2019**;20:662
35. Kim J, Lee Y, Lu X, Song B, Fong KW, Cao Q*, et al.* Polycomb- and Methylation-Independent Roles of EZH2 as a Transcription Activator. Cell Rep **2018**;25:2808-20 e4
36. Shen X, Liu Y, Hsu YJ, Fujiwara Y, Kim J, Mao X*, et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell **2008**;32:491-502

37. Sanders DA, Ross-Innes CS, Beraldi D, Carroll JS, Balasubramanian S. Genome-wide mapping of FOXM1 binding reveals co-binding with estrogen receptor alpha in breast cancer cells. Genome Biol **2013**;14:R6
38. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM*, et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. Cell **2006**;125:301-13
39. Li H, Liefke R, Jiang J, Kurland JV, Tian W, Deng P*, et al.* Polycomb-like proteins link the PRC2 complex to CpG islands. Nature **2017**;549:287-91
40. Liu A, Yu X, Liu S. Pluripotency transcription factors and cancer stem cells: small genes make a big difference. Chin J Cancer **2013**;32:483-7
41. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT*, et al.* Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun **2017**;8:15081
42. Gwak JM, Kim M, Kim HJ, Jang MH, Park SY. Expression of embryonal stem cell transcription factors in breast cancer: Oct4 as an indicator for poor clinical outcome and tamoxifen resistance. Oncotarget **2017**;8:36305-18
43. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature **2012**;490:61-70
44. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ*, et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature **2012**;486:346-52
45. Domenici G, Aurrekoetxea-Rodriguez I, Simoes BM, Rabano M, Lee SY, Millan JS*, et al.* A Sox2-Sox9 signalling axis maintains human breast luminal progenitor and breast cancer stem cells. Oncogene **2019**;38:3151-69
46. Collinson A, Collier AJ, Morgan NP, Sienerth AR, Chandra T, Andrews S*, et al.* Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells. Cell Rep **2016**;17:2700-14
47. Mahara S, Lee PL, Feng M, Tergaonkar V, Chng WJ, Yu Q. HIFI-alpha activation underlies a functional switch in the paradoxical role of Ezh2/PRC2 in breast cancer. Proc Natl Acad Sci U S A **2016**;113:E3735-44
48. Hagey DW, Muhr J. Sox2 acts in a dose-dependent fashion to regulate proliferation of cortical progenitors. Cell Rep **2014**;9:1908-20
49. Chen X, Muller GA, Quaas M, Fischer M, Han N, Stutchbury B*, et al.* The forkhead transcription factor FOXM1 controls cell cycle-dependent gene expression through an atypical chromatin binding mechanism. Mol Cell Biol **2013**;33:227-36
50. Guo X, Wang Z, Lu C, Hong W, Wang G, Xu Y*, et al.* LincRNA-1614 coordinates Sox2/PRC2-mediated repression of developmental genes in pluripotency maintenance. J Mol Cell Biol **2018**;10:118-29

Figure 1 – RNA-seq from MDA-MB-231 breast cancer cells grown in 2D or 3D conditions identify the change in lncRNA expression. (**a**) Left. Representative images of MDA-MB-231 cells grown in adherent (adh) and tumorspheres (spheres) conditions. Right. Heatmap showing expression levels of significantly up- and down-regulated genes between adh, 16h

and 5d spheres (R1 =replicate #1; R2 = Replicate #2). (**b**) Log2 fold change of lncRNA candidates' expression between 16h or 5d MDA-MB-231 spheres compared to adh cells. (**c**) Percentage of sphere formation efficiency (SFE%) after down-regulation of lncRNA candidates in MDA-MB-231 and (**d**) MCF-7 cells. Results are from two independent experiments each performed in triplicate. n=6, * $p<0.05$, ** $p<0.01$. P-values are calculated using non-parametric Mann–Whitney U test.

Figure 2 – SCIRT is chromatin-enriched, up-regulated in spheres (16h and 5d) but counteracts stemness. (**a**) Boxplots showing fold change in expression of SCIRT in spheres compared to adh cells isolated from luminal A and basal breast cancer patient tumors. Data are from two independent tumors per subtype each plated in triplicates; n=6, ** $p<0.01$. (**b**) Barplots showing SFE% in the first and second generation after down-regulation of SCIRT with two independent siRNAs (siSCIRT#1 or siSCIRT#2) in breast cancer cell lines MDA-MB-231 and MCF-7. Results are from two independent experiments each performed in triplicate. n=6, * $p<0.05$, ** $p<0.01$, ns=not significant. (**c**) *In vivo* tumorigenicity assay using MDA-MB-231 cells stable knockdown for lncRNA SCIRT (shSCIRT) or control cells (shNC) injected subcutaneously in immunocompromised, nude mice. Tumor take frequency is shown in the upper panel. Images of resected tumors are shown in the middle panel. Tumor growth during 48 days of monitoring is present in the lower panel. (**d**) Images of RNA FISH experiments with probes for lncRNA SCIRT in MCF-7 cells treated with siSCIRT or siNC. (**e**) RT-qPCR values showing the expression of SCIRT, MALAT1 and GAPDH in different cellular fractions of MDA-MB-231 sphere derived cells. The values are from two independent experiments and are expressed as fold change to free nuclear fraction. P-values are calculated using non-parametric Mann–Whitney U test.

Figure 3 – SCIRT specifically interacts with cis-regulatory regions to down-regulate stemness genes and up-regulate cell cycle genes. (**a**) GSEA analysis for top enriched pathways for genes up- and down-regulated by siSCIRT. (**b**) Specific enrichment of SCIRT using oligonucleotides complementary to accessible regions of SCIRT compared to LacZ or mock control samples. (**c**) SCIRT binding within promoters of two genes significantly down-regulated by siSCIRT (PRR11 and NEK2) and two genes significantly up-regulated by siSCIRT (SOX4 and BSDC1). (**d**) Top. Venn diagram showing the overlap of genes containing SCIRT peaks within cis-regulatory regions (±200 kb from TSSs) of genes significantly up or down-regulated after siSCIRT. Bottom-left. Enrichment of pathways (top) and TF factor binding (ChEA) (bottom) analyzes performed by using Enrichr (https://amp.pharm.mssm.edu/Enrichr/) for genes bound by SCIRT and down-regulated after siSCIRT#1 and #2 treatment. Bottom-right. Enrichment of pathways (top) and TF factor binding (ChEA) (bottom) analyzes performed by using Enricher (https://amp.pharm.mssm.edu/Enrichr/) for genes bound by SCIRT and up-regulated after siSCIRT#1 and #2 treatment.

Figure 4 – SCIRT interacts with EZH2 to antagonize its Polycomb-independent activity. (**a**) Venn diagram showing the overlap of genes significantly down-regulated after SCIRT knockdown and up-regulated after EZH2 knockdown. Enrichr ChEA analysis shows FOXM1 and E2F4 TF are enriched and bind to genes from the overlap. (**b**) Venn diagram showing the overlap of genes significantly up-regulated after silencing SCIRT and genes significantly down-regulated after silencing EZH2. Enrichr ChEA analysis shows SOX2 TF is enriched and binds to genes from the overlap. (**c**) Representation of H3K27me3 occupancy from ChIP-seq

data of siNC, siEZH2 and siSCIRT treated MDA-MB-231 spheres at expressed and repressed genes. (**d**) Representation of H3K27me3 occupancy from ChIP-seq data of MDA-MB-231 spheres treated with siNC, siEZH2 and siSCIRT at genes down-regulated after silencing SCIRT. (**e**) Representation of H3K27me3 occupancy from ChIP-seq data of MDA-MB-231 spheres treated with siNC, siEZH2 and siSCIRT at genes up-regulated after silencing SCIRT. (**f**) Barplots show sphere formation efficiency after siRNA down-regulation of SCIRT, FOXM1 and SCIRT/FOXM1 co-repression; SCIRT, SOX2 and SCIRT/SOX2 co-repression; SCIRT, EZH2 and SCIRT/EZH2 co-repression in MDA-MB-231 cells. Spheres were counted at day 5 from siRNA transfection and the results are from two independent experiments each performed in duplicate. n=4, * $p<0.05$; ns=not significant. P-values were calculated using non-parametric Mann–Whitney U test.

Figure 5 – SCIRT colocalizes with EZH2, FOXM1 and SOX2 at promoters of regulated genes. (**a**) Left. Heatmap and K-mean clustering of genomic regions occupied by H3K4me1, H3K4me3, FOXM1, EZH2 and SOX2 within SCIRT peaks associated with SCIRT-regulated genes. These regions were not occupied by LacZ used as negative control. Right. Bar-plot depicting the number of genes from the top cluster in the heatmap, associated with promoters or distant sites (**b**) Heatmap generated from each replicate pairs from CHART-seq (LacZ and SCIRT) and ChIP-seq (SOX2, EZH2, H3K27me3) experiments and publicly available FOXM1 (GSE40762) ChIP-seq data. Enriched known TF motif in SCIRT peaks were obtained using HOMER (**c**). Average of normalized peak density of SCIRT, EZH2 and SOX2 around TSSs of genes up-regulated or (**d**) down-regulated by siSCIRT treatment. (**e**) Barplots showing expression changes of a subset of genes up-regulated by siSCIRT, upon siSOX2 (right treatment compared to siNC treatment. Data derive from two independent experiments

each performed in duplicate; n=4, * p<0.05. P-values are calculated using non-parametric Mann–Whitney U test.

Figure 6 – SCIRT interacts with FOXM1 and is important for FOXM1 recruitment on cell cycle genes. (**a**) RNA-immunoprecipitation (RIP) followed by RT-qPCR shows that SCIRT interacts with EZH2, but FOXM1 does not modulate this interaction. Data shown are from two independent experiments, n=2. U1 snRNA and IgG were used as negative controls. (**b**) Co-IP assay shows a physical interaction between EZH2 and FOXM1 in MDA-MB-231 spheres. (**c**) Co-IP assay obtained by immunoprecipitation of EZH2 (left panel) or FOXM1 (right panel) shows that SCIRT depletion disrupts the interaction between EZH2 and FOXM1 in MDA-MB-231 spheres. (**d**) Western blot showing that the level of FOXM1 protein remains unchanged in the cells where SCIRT is transiently down-regulated. (**e**) FOXM1 chromatin-immunoprecipitation (ChIP) followed by qPCR shows that FOXM1 binding on mitotic genes is reduced after SCIRT down-regulation in spheres. Data are from two independent experiments each performed in duplicate (MDA-MB-231, n=2; MCF-7, n=4). * p<0.05, ** p<0.01, *** p<0.001. P-values are calculated using t-test (**f**) Left panel: Venn diagram showing overlap between genes significantly down-regulated by SCIRT siRNA (siSCIRT) (p-adjusted-value < 0.05) with genes significantly down-regulated by FOXM1 (siFOXM1) (p-adjusted-value < 0.05) in MDA-MB-231 cells. Right panel: GO-term enriched analysis performed by using Enrichr of genes significantly down-regulated by siSCIRT (2 independent siRNAs) only (top), by both siSCIRT and siFOXM1 (middle) and siFOXM1 only (bottom).

Figure 1

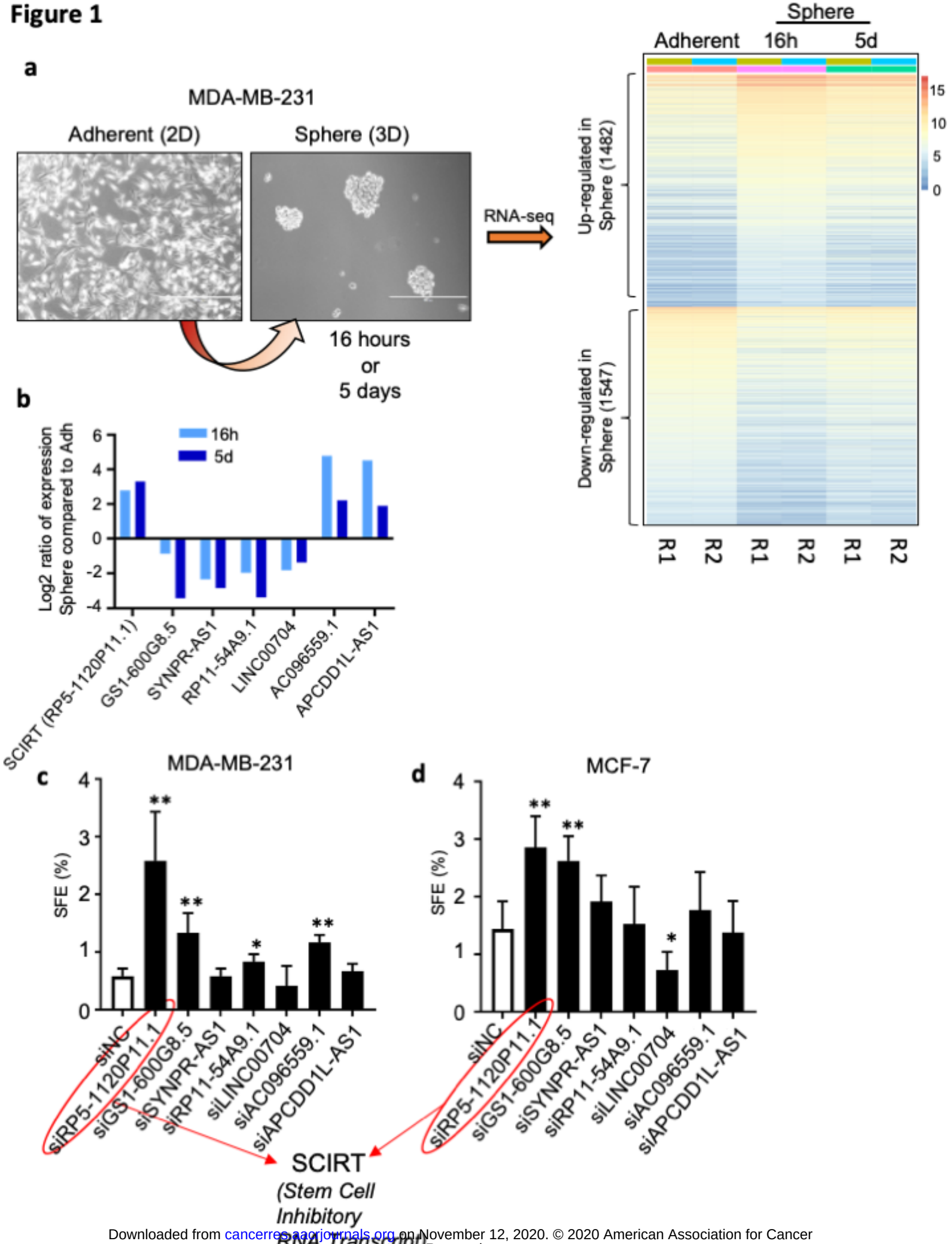


Figure 2

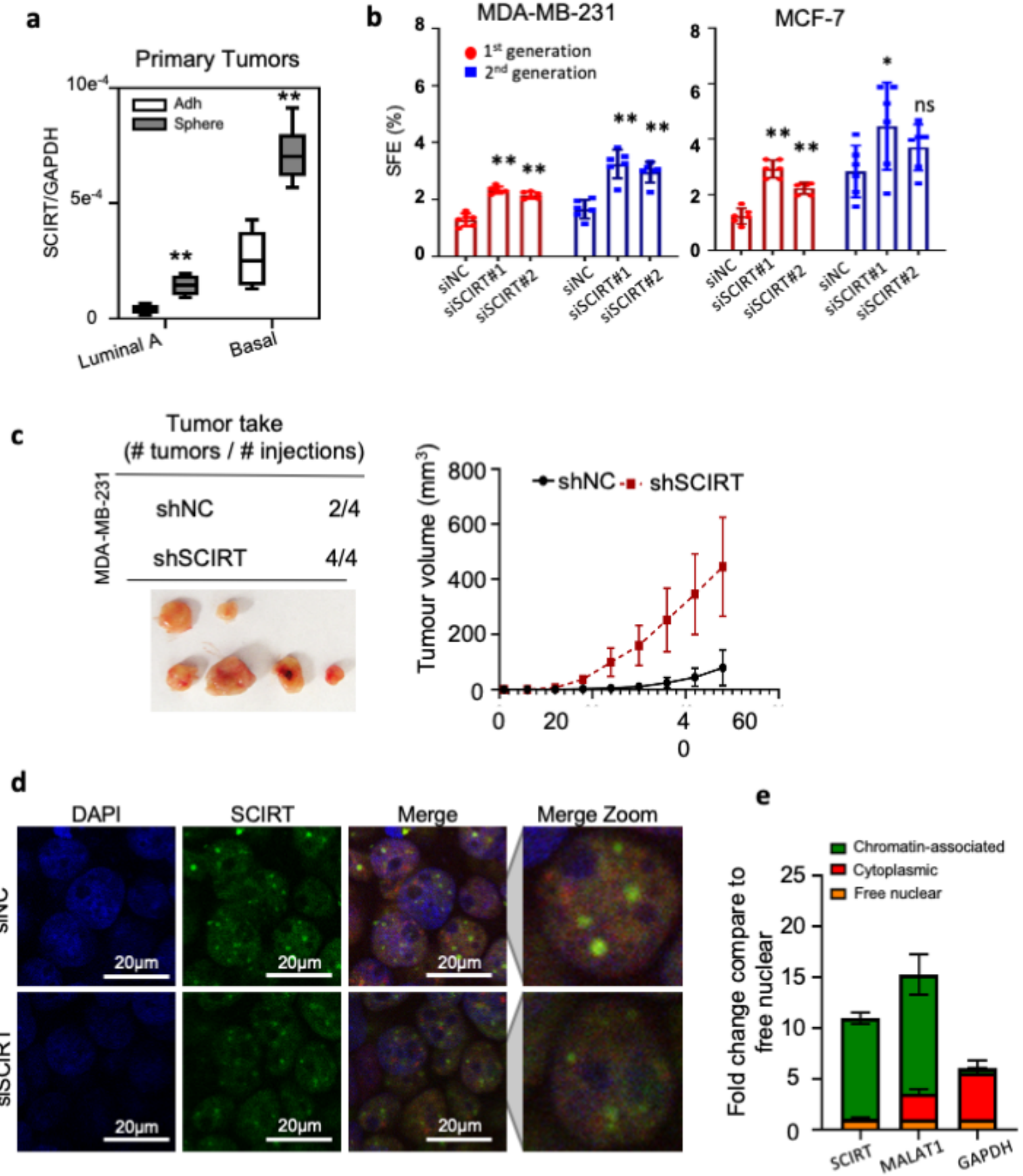
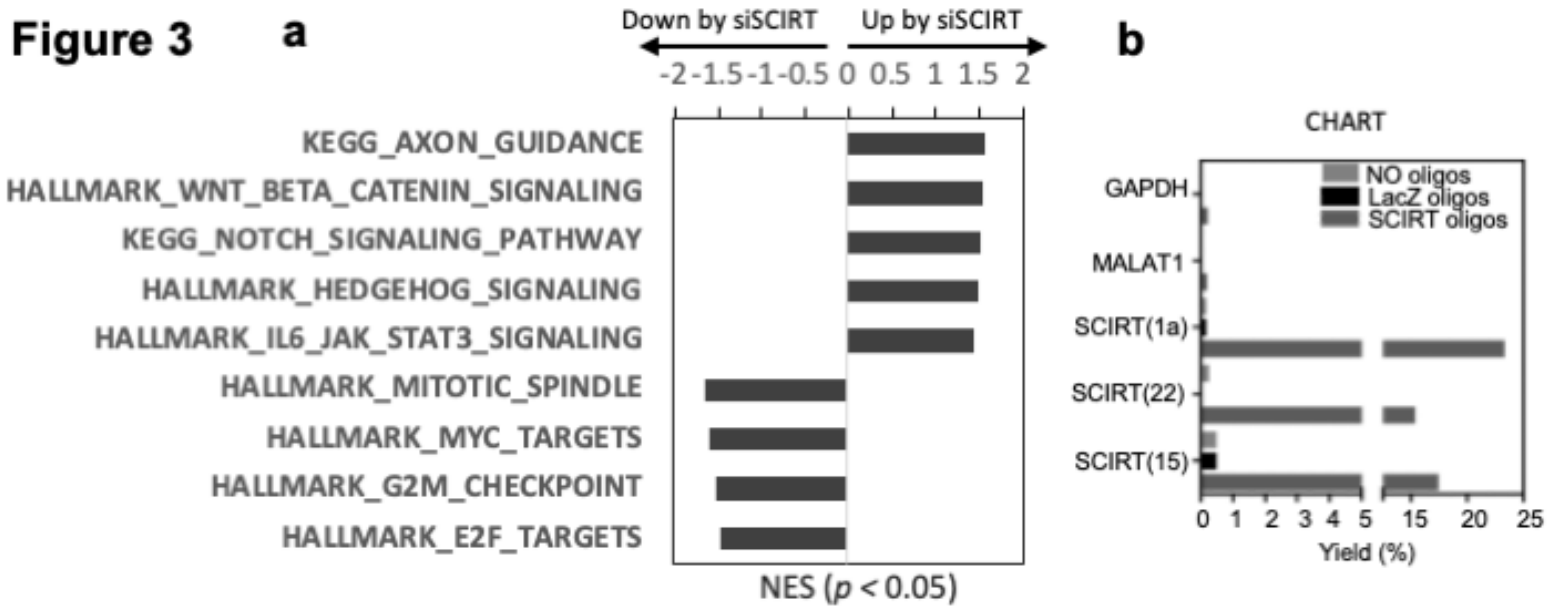
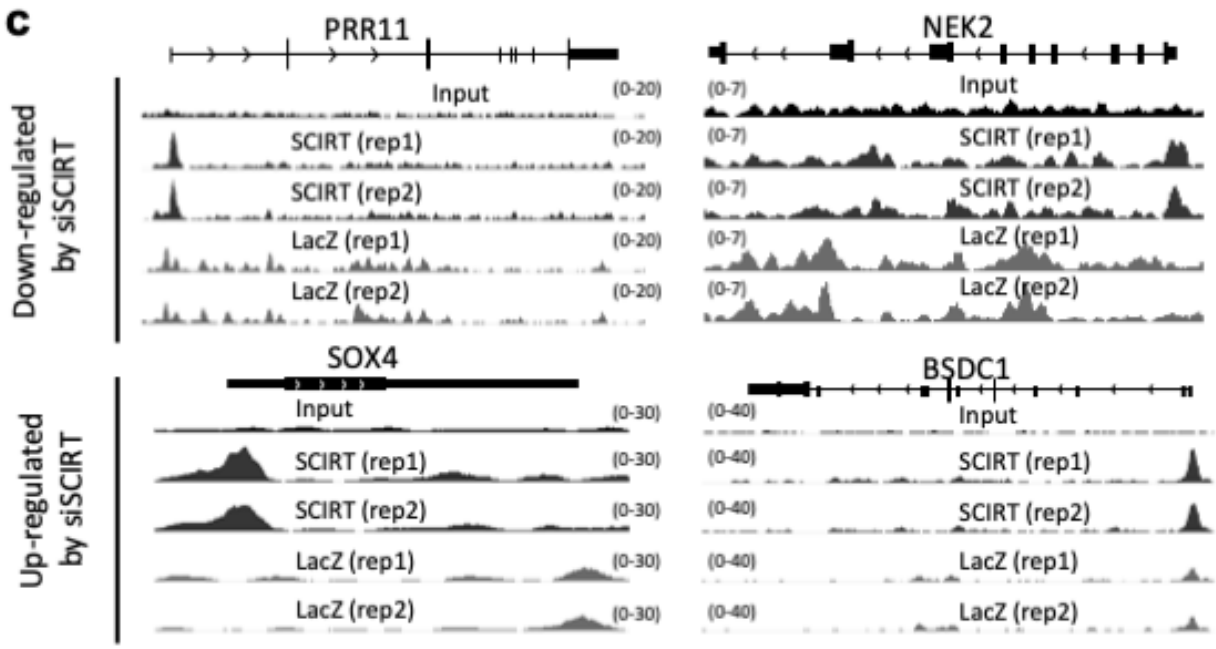
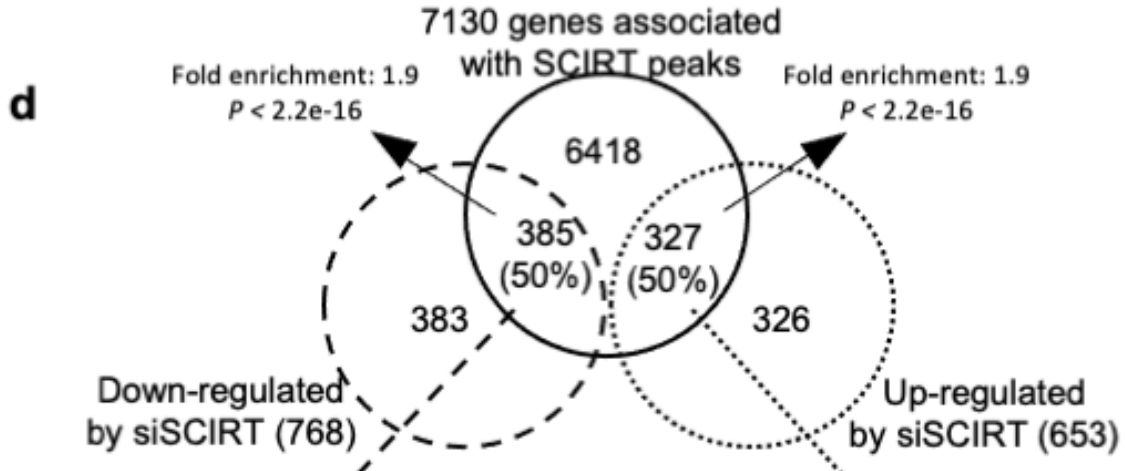


Figure 3







Pathway	Adj p-value
Cell cycle	1.6e-35
DNA replication	3.0e-31
TF ChIP-seq (ChEA)	
FOXM1 ChiP-seq Human	1.8e-35
E2F4 ChiP-seq Human	2.1e-23
E2F1 ChiP-seq Human	1.2e-9

Pathway	Adj p-value
TGF-beta signalling	4.2e-3
PIK3-Akt signalling	5.3e-3
TF ChIP-seq (ChEA)	
KDM2B ChiP-seq Human	4.4e-8
SOX9 ChIP-seq Human	3.5e-6
SOX2 ChIP-seq Human	8.3e-3

Figure 4

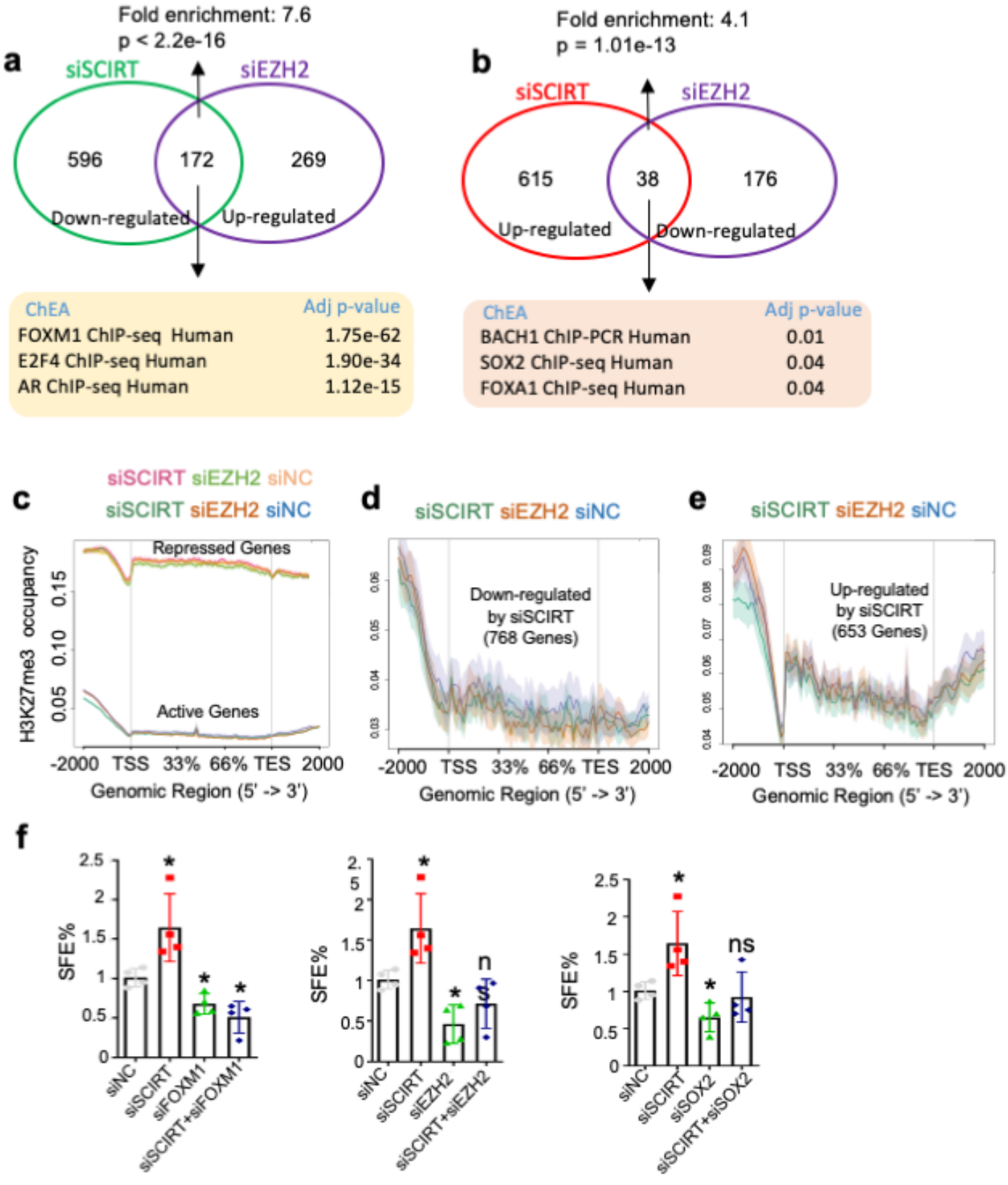


Figure 5

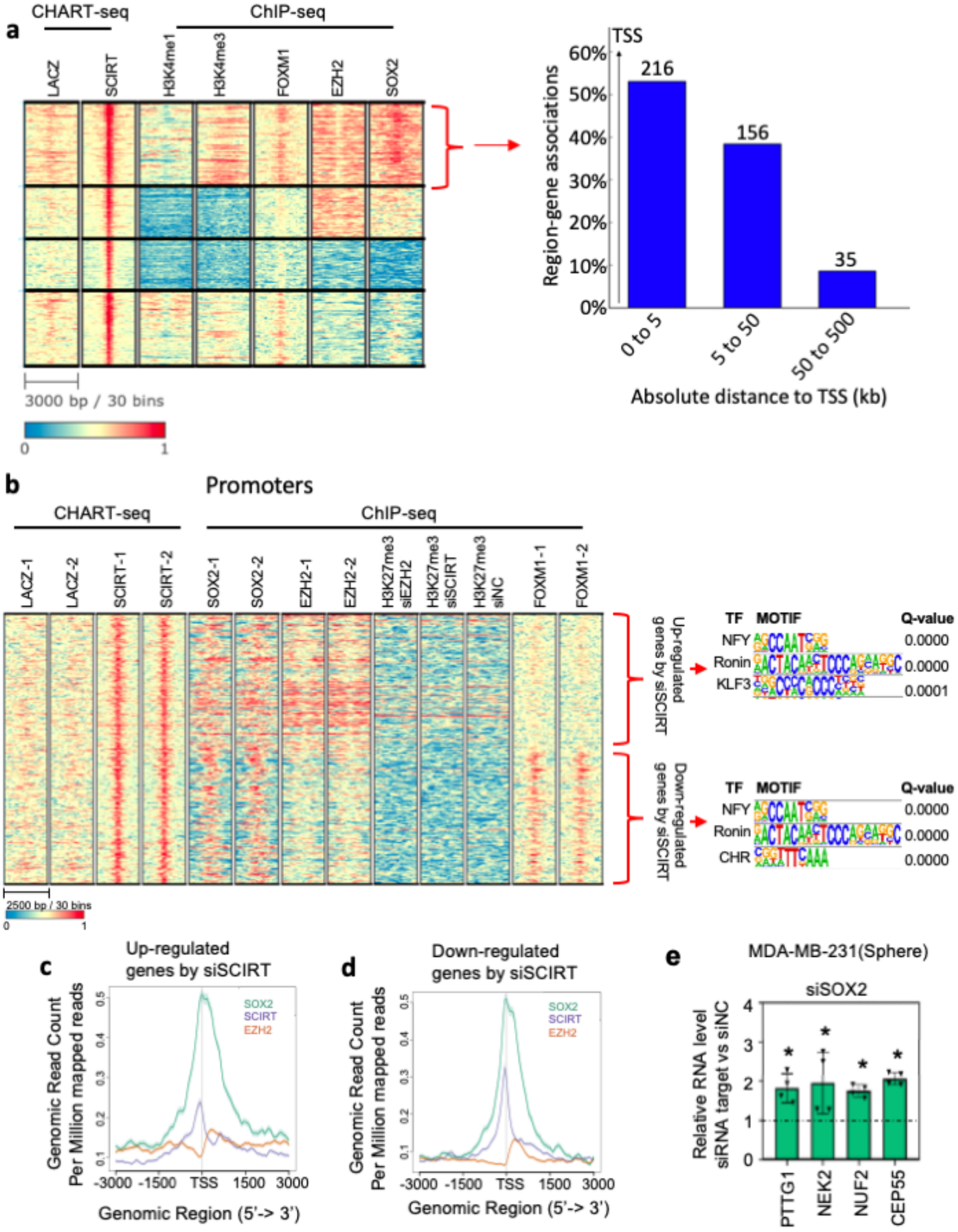
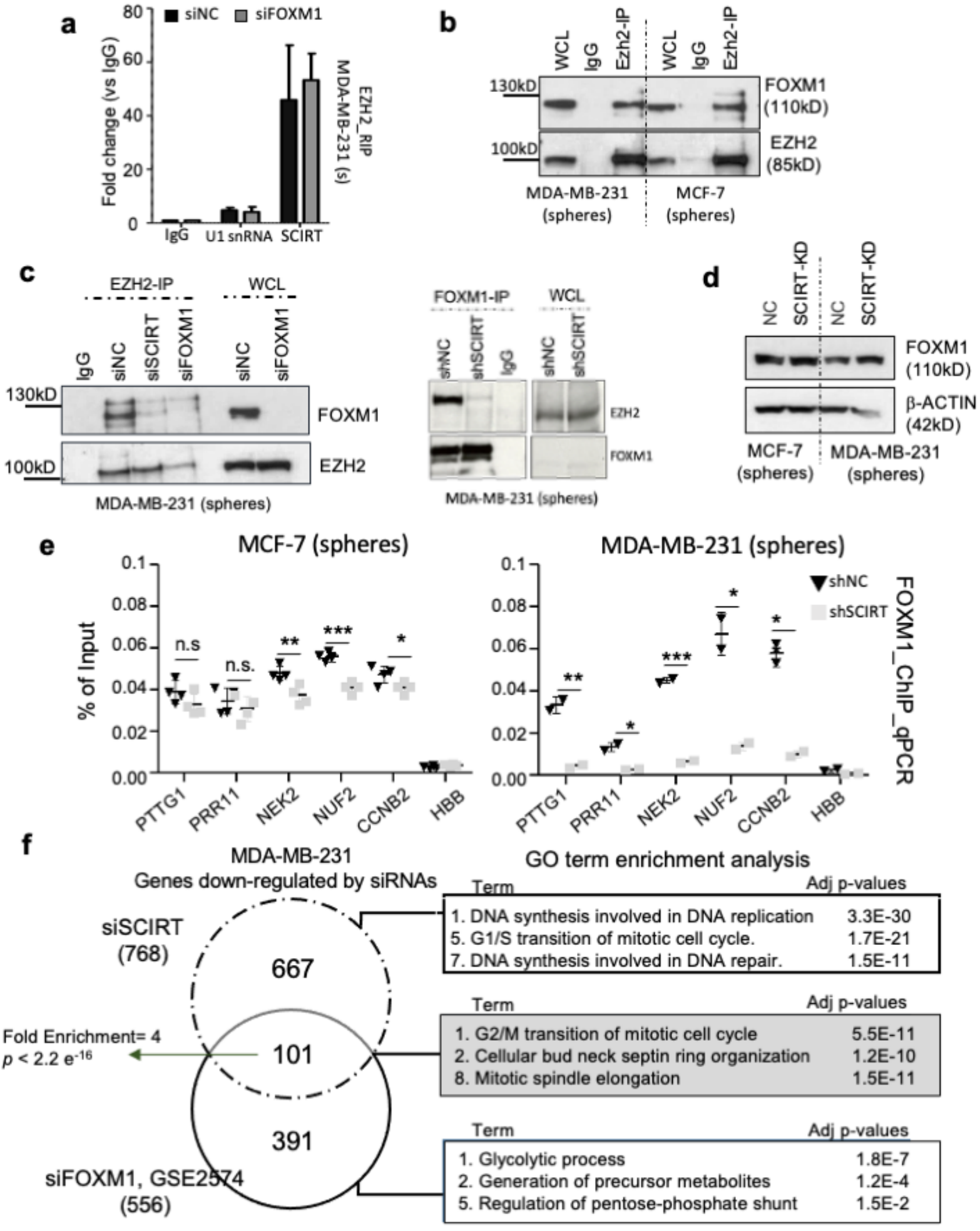


Figure 6



Cancer Research
The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

SCIRT lncRNA restrains tumorigenesis by opposing transcriptional programs of tumor-initiating cells

Sladjana Zagorac, Alex de Giorgio, Aleksandra Dabrowska, et al.

Cancer Res Published OnlineFirst November 10, 2020.