Genome-wide analysis of endogenously expressed ZEB2 binding sites reveals inverse correlations between ZEB2 and GalNAc-transferase GALNT3 in human tumors

Pelin Balcik-Ercin, [1]

Metin Cetin, [1]

Irem Yalim-Camci, [1]

Gorkem Odabas, [1]

Nurettin Tokay, [1]

A. Emre Sayan, [2]

Tamer Yagci, [1]✉

Phone +90 (262) 605 2573
Email tyagci@gtu.edu.tr

[1] Department of Molecular Biology and Genetics, Laboratory of Molecular Oncology AQ1 , Gebze Technical University, C2-Building, 41400 Gebze-Kocaeli, Turkey

[2] Faculty of Medicine, Cancer Sciences, University of Southampton, Somers Building, Tremona Road, Southampton, SO16 6YD UK

Abstract

Background

ZEB2 is a transcriptional repressor that regulates epithelial-to-mesenchymal transition (EMT) through binding to bipartite E-box motifs in gene regulatory regions. Despite the abundant presence of E-boxes within the human genome and the multiplicity of pathophysiological processes regulated during ZEB2-induced EMT, only a small fraction of ZEB2 targets has been identified so far. Hence, we explored genome-wide ZEB2 binding by chromatin immunoprecipitation-sequencing (ChIP-seq) under endogenous ZEB2 expression conditions.

Methods

For ChIP-Seq we used an anti-ZEB2 monoclonal antibody, clone 6E5, in SNU398 hepatocellular carcinoma cells exhibiting a high endogenous ZEB2 expression. The ChIP-Seq targets were validated using ChIP-qPCR, whereas ZEB2-dependent expression of target genes was assessed by RT-qPCR and Western blotting in shRNA-mediated ZEB2 silenced SNU398 cells and doxycycline-induced ZEB2 overexpressing colorectal carcinoma DLD1 cells. Changes in target gene expression were also assessed using primary human tumor cDNA arrays in conjunction with RT-qPCR. Additional differential expression and correlation analyses were performed using exp*O* and Human Protein Atlas datasets.

Results

Over 500 ChIP-Seq positive genes were annotated, and intervals related to these genes were found to include the ZEB2 binding motif CACCTG according to TOMTOM motif analysis in the MEME Suite database. Assessment of ZEB2-dependent expression of target genes in ZEB2-silenced SNU398 cells and ZEB2-induced DLD1 cells revealed that the *GALNT3* gene serves as a ZEB2 target with the highest, but inversely correlated, expression level. Remarkably, *GALNT3* also exhibited the highest enrichment in the ChIP-qPCR validation assays. Through the analyses of primary tumor cDNA arrays and exp*O* datasets a significant differential expression and a significant inverse correlation between *ZEB2* and

GALNT3 expression were detected in most of the tumors. We also explored ZEB2 and GALNT3 protein expression using the Human Protein Atlas dataset and, again, observed an inverse correlation in all analyzed tumor types, except malignant melanoma. In contrast to a generally negative or weak ZEB2 expression, we found that most tumor tissues exhibited a strong or moderate GALNT3 expression.

Conclusions

Our observation that ZEB2 negatively regulates a GalNAc-transferase (GALNT3) that is involved in O-glycosylation adds another layer of complexity to the role of ZEB2 in cancer progression and metastasis. Proteins glycosylated by GALNT3 may be exploited as novel diagnostics and/or therapeutic targets.

Keywords

ZEB2
GALNT3
Antibody validation
ChIP-sequencing
Gene regulation
Tissue expression

Abbreviations

EMT	Epithelial-to-mesenchymal transition
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation-sequencing
ChIP-qPCR	Chromatin immunoprecipitation-quantitative polymerase chain reaction
HCC	Hepatocellular carcinoma
RT-qPCR	Real time- quantitative polymerase chain reaction

ZEB2	Zinc finger E-box binding homeobox 2
GALNT3	Polypeptide N-acetylgalactosaminyltransferase 3

Electronic supplementary material

The online version of this article (https://doi.org/10.1007/s13402-018-0375-7) contains supplementary material, which is available to authorized users.

1. Introduction

ZEB2, also known as SMAD Interacting Protein 1 (SIP1), is a member of the ZEB family of transcription factors and is encoded by the *ZFHX1B* gene on chromosome 2 [1]. The ZEB family proteins contain a SMAD binding domain, a homeodomain and two clusters of zinc fingers, one located at the amino (N) and one at the carboxyl (C) terminal end [2]. The N-terminal and C-terminal zinc finger clusters represent the most conserved sequences between ZEB1 and ZEB2, mediating their binding to the bipartite E2 box motif CACCT(N) [3]. Functionally, ZEB2 was initially found to downregulate the expression of E-cadherin by binding to E2 boxes within the promoter region of its encoding gene *CDH1* [4]. ZEB2 plays a role in TGF-β dependent gene repression via binding to the C-terminal Binding Protein (CtBP), but both wild type and CtBP mutant ZEB2 have been found to be capable of repressing E-Cadherin expression in a CtBP independent manner [5, 6]. Thus, E-cadherin repression occurs directly through binding of ZEB2 to the *CDH1* gene promoter, whereas zinc finger mutant ZEB2 does not repress E-cadherin expression [6, 7]. E-Cadherin downregulation is one of the characteristics of cells undergoing epithelial-to-mesenchymal transition (EMT), and as one of the EMT inducers, increased ZEB2 expression has been observed in E-Cadherin-negative human carcinoma cells [8, 9]. In non-epithelial tumors, ZEB2 and E-cadherin may be co-expressed. Caramel et al. [10] reported that ZEB1, but not ZEB2, is required for E-cadherin repression in human melanoma cells. Both *ZEB2* nonsense and frameshift mutations have been detected in patients with Mowat-Wilson syndrome, suggesting a role in neural development [11]. Another study has revealed a functional interaction between ZEB2 and the NuRD co-repressor complex, and the authors proposed that defective recruitment of NuRD by mutant ZEB2 may be the underlying cause of Mowat-Wilson syndrome [12]. Besides E-cadherin, induction of EMT by conditional ZEB2 expression in human epidermoid carcinoma A431 cells has been found to be accompanied by downregulation of epithelial junction proteins such as desmoplakin, plakophilin-2 and claudin 4 [13].

ZEB2 has also been highlighted as a negative regulator of hTERT expression upon stimulation with TGF-β [14]. Related to this, we previously found that exogenous ZEB2 expression may induce replicative senescence in hepatocellular carcinoma (HCC) cells [15]. ZEB2 expression induction in A431 cells has been found to cause mesenchymal conversion, along with G1 cell cycle arrest resulting from direct repression of cyclin D1 [16]. The regulation of ZEB2 expression by TGF-β has been found to be cell-context dependent, i.e., ZEB2 expression levels were found to be increased in TGF-β treated mouse mammary epithelial cells, whereas kidney mesangial cells treated with TGF-β showed a decreased ZEB2 expression through upregulation of a regulatory microRNA, miR192 [17, 18]. The miR200 family of microRNAs and miR205 have been found to be able to inhibit EMT by targeting both *ZEB2* and *ZEB1* transcripts [19, 20]. ZEB1 and ZEB2 can, in turn, repress microRNAs of the miR200 family, thereby creating a double negative feedback loop, allowing maintenance of epithelial or mesenchymal cellular phenotypes [21, 22]. In addition to microRNA-mediated repression, we previously observed ZEB2 downregulation in both HCC-derived cell lines and primary HCC tissues due to hypermethylation of three regulatory regions of the *ZEB2* gene [23]. The complexity of gene expression regulation outlined above was also reflected by differential ZEB2 expression patterns in solid tumors as detected by immunohistochemistry using a tissue microarray (TMA) platform [24]. Additionally, ZEB2 has been found to be barely detectable in bladder carcinoma cells in culture, but to be overtly expressed in primary bladder cancers for which it may act as an independent poor prognostic factor for survival and response to radiotherapy [7].

Despite its role in the aforementioned pathophysiological processes and given the frequent occurrence of E-boxes within the human genome, only a few ZEB2 target genes have been identified so far [25]. Previous chromatin immunoprecipitation (ChIP) studies focusing on ZEB2 were gene targeted (e.g. towards *CDH1*) and accomplished in exogenous ZEB2 expressing cell lines using tag antibodies for immunoprecipitation [8, 13, 16, 26, 27]. On the basis of these observations, we aimed to explore genome-wide ZEB2 binding sites in a high endogenous ZEB2 expressing HCC cell line, SNU398 [23]. A ChIP-sequencing (ChIP-seq) study was performed using a homemade anti-ZEB2 monoclonal antibody (6E5) revealing several targets associated with multiple cellular processes and pathways. From these, we selected *GALNT3*, a member of the GalNAc-transferases family of genes encoding Polypeptide N-acetylgalactosaminyltransferase 3 (GalNAc-T3), for further analysis based on its high enrichment by ChIP, and assessed its ZEB2-dependent expression in ZEB2 silenced SNU398 cells and in ZEB2 overexpressing colorectal carcinoma DLD1 cells [13]. We found that the

GALNT3 expression level was inversely correlated with that of *ZEB2.* A similar inverse pattern was observed in primary human tumor tissues. Our work reveals a new role for ZEB2 as glycosylation regulator in tumor cells.

2. Materials and methods

2.1. Antibodies and materials

The following antibodies were used: anti-ZEB2 (home-made monoclonal antibody, clone 6E5) [7], anti-ACTB mouse monoclonal antibody (A00702) from GenScript (Piscataway, NJ, USA), rabbit polyclonal anti-GALNT3 antibody (LS-C166355) from LifeSpan BioSciences (Seattle, WA, USA), rabbit polyclonal anti-MYC Tag antibody (ab9106) from Abcam (Cambridge, UK), and mouse monoclonal anti-CCND1 antibody (sc-20,044) and anti-E-cadherin antibody (sc-8426) from Santa Cruz Biotechnology (Dallas, TX, USA). GIPZ-shRNA lentiviral particles containing oligonucleotides targeting human ZEB2 (clone ID V2LHS-268,826 (Lot 1), V3LHS-373,825 (Lot 2), V2LHS-232,431 (Lot 3)), and a non-silencing Control (RHS4348) were purchased from Thermo Fisher Scientific (Waltman, MA, USA). TissueScan qPCR Cancer Survey cDNA arrays I (CSRT101) were purchased from OriGene Technologies (Rockville, MD, USA).

2.2. Cell lines and culture conditions

Hepatocellular carcinoma (HCC) cell line SNU398 was cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS). HCC cell line SK-HEP-1 and colorectal adenocarcinoma cell line DLD-ZEB2 were cultured in DMEM medium supplemented with 10% FBS at 37 °C in a humidified 5% CO2 atmosphere. Monoclonal anti-ZEB2 secreting hybridoma clone 6E5 was cultured in DMEM supplemented with 10% Hyclone Super Low IgG FBS (Thermo Scientific, #SH3089803) and penicillin/streptomycin. Supernatants were collected in sterile containers and stored at −20 °C until use. The SNU398 and SK-HEP-1 cell lines were obtained from Dr. Mehmet Ozturk [28]. The DLD-ZEB2 cell line [13]. was a kind gift of Dr. Geert Berx to A.E. Sayan. The isolation of the anti-ZEB2 antibody producing hybridoma clone 6E5 has been reported before [7, 24]. All cell lines were routinely checked for mycoplasma contamination. The identities of the SNU398 and SK-HEP-1 cell lines were validated by STR analysis.

2.3. Western blotting

Cell lysates were suspended in SDS lysis buffer (10% Glycerol, 2% SDS in 62,5 mM Tris-HCl) including a Protease Inhibitor Cocktail (Roche), and protein concentrations were determined using a BCA assay. Equal amounts of protein (50 μg) were separated by 8% or 12% SDS-PAGE and transferred to PVDF membranes (Millipore, #IPVH00010). After blocking for 1 h in Tris Buffered Saline (TBS: 50 mM Tris-Cl, pH 7.5. 150 mM NaCl) containing 5% non-fat dry milk, the membranes were incubated with primary antibodies at 4 °C overnight and with a secondary antibody for one hour at room temperature. After rigorous washing in TBS-Tween 20, the membranes were immersed in a Signal Fire chemiluminescent reagent (Cell Signaling Technology, #6883) after which the protein bands were visualized using a ChemiDocTMXRS system (Bio-Rad, CA).

2.4. ChIP assay and validation of anti-ZEB2 antibody

The ChIP experiments were carried out according to protocols of Active Motif Inc. (Carlsbad, CA, USA) (http://www.activemotif.com/documents/1848.pdf) as reported by Gévry et al. and Carey et al. [29, 30] with minor modifications. For crosslinking proteins to DNA, 1/10 volume formaldehyde solution (11% formaldehyde, 100 mM NaCl, 1 mM EDTA pH:8,0, 50 mM HEPES pH:7,9) was added to the medium in culture flasks containing 5×10^6 cells. After treatment for 15 min at room temperature, the reaction was stopped by adding glycine (final concentration 0.125 M). Next, the cells were washed with ice-cold PBS and cell pellets were resuspended in PIPES buffer (5 mM PIPES pH:8,0, 85 mM KCl, 0,5% NP40 and protease inhibitor cocktail) after which nuclear pellets were generated by centrifugation. The nuclei were subsequently lysed in Tris-EDTA-SDS (10 mM Tris-HCl, pH:8,0. 1 mM EDTA, 1% SDS) buffer. The resulting lysates were sonicated to generate 200 to 1000 bp long DNA fragments after which 10% of the lysate volume was preserved as control (input). The remaining volume was split into two equal fractions, which were each incubated at 4 °C overnight with Dynabeads® Protein G magnetic beads (Thermo Fisher Scientific, #10004D) that were previously coated with anti-ZEB2 (6E5) antibody or mouse IgG2a isotype control (Abcam, #ab18413) antibody. Magnetically collected beads were washed 5 times in LiCl buffer (10 mM Tris pH:7,5, 500 mM LiCl, 1%NP40, 1% sodium deoxycholate, 1 mM EDTA) and 2 times in Tris-EDTA buffer (10 mM Tris-HCl, pH:8,0, 1 mM EDTA), after which DNA-protein complexes were eluted with Tris-EDTA-SDS buffer. Next, the eluates were treated with RNase A (Thermo Fisher Scientific, #EN0531) for 30 min at 37 °C and then with proteinase K (Thermo Fisher Scientific, #AM2548) overnight. To reverse formaldehyde crosslinking, samples were incubated at 65 °C for 6 h. ChIP-DNA was extracted using a QIAquick PCR purification kit (Qiagen,

#28104). For ChIP validation of the 6E5 antibody, beads obtained through the above-mentioned ChIP protocol were treated with RNase and DNase and boiled in Laemmli buffer for 10 min. After removal of the beads by centrifugation, the samples were loaded onto 8% SDS-PAGE gels after which standard Western blotting procedures were carried out as described above.

2.5. ChIP-sequencing

SNU398 cells were fixed according to the cell fixation protocol provided by Active Motif Inc. (http://www.activemotif.com/documents/1848.pdf). Briefly, cells were incubated for 15 min at RT in a cell fixation solution consisting of 1% formaldehyde, 10 mM NaCl, 0.1 mM EDTA, pH 8,0 and 5 mM HEPES, pH 7.9. The fixation was stopped through the addition of glycine (final concentration 125 mM), after which the cells were washed 3 times with PBS-0.5% Igepal 40. PMSF (final concentration 1 mM) was added to the last wash. Next, cell pellets were snap-frozen and sent to Active Motif Inc. (Carlsbad, CA, USA) on dry ice for further analysis. DNA sequencing and computational analyses of the ChIP-Seq data were performed after immunoprecipitation of ZEB2-bound chromatin fragments using anti-ZEB2 monoclonal antibody 6E5. Briefly, 75-nucleotide sequence reads obtained through NextSeq 500 Illumina sequencing yielded 21 million tags. These tags (reads) were aligned to the human reference genome (NCBI Genome database, hg19) using the BWA algorithm [31]. Since the 5′-ends of the aligned reads represent the ends of ChIP-IP-fragments, the tags were extended *in silico* (using Active Motif software) at their 3′-ends to a length of 150–250 bp, depending on the average fragment length present in the size selected library (normally 200 bp). To identify the density of fragments (extended tags), the genome was divided into 32-nt bins after which the number of fragments in each bin was determined. The MACS peak caller was used to identify peaks with a cutoff value of 1e-7 [32]. Genomic regions with local enrichments in tag numbers were denoted as intervals. These intervals were defined by chromosome number and start and end coordinates. Gene margins were set to 10,000 bp, and intervals within 10,000 bp upstream and downstream of a gene region were considered as being associated with that gene. The TOMTOM algorithm present at the MEME suite database (meme-suite.org version 4.12.0) was used for the identification of ZEB2 binding motifs within ±200 bp of the peak maxima [33]. Functional annotations of ZEB2 targets were performed using the Protein Analysis Through Evolutionary Relationships (PANTHER) classification system (www.pantherdb.org).

2.6. Generation of stable ZEB2 shRNA cell clones

SNU398 cells were transduced with ZEB2-shRNA and control-shRNA lentiviral particles in the presence of 8 μg/ml polybrene (Sigma-Aldrich, #TR-1003-G). After 24 h 5 μg/ml puromycin (Thermo Fisher Scientific, #A1113802) was added to the cultures for selecting stable ZEB2-shRNA and control-shRNA cell clones.

2.7. RT-qPCR and ChIP-qPCR

Total RNA was isolated from SNU398 ZEB2-shRNA and control-shRNA cell clones, as well as doxycycline-induced and uninduced DLD-ZEB2 cells using a NucleoSpin RNA kit (Macharey-Nagel, #740955.250). cDNA was synthesized using a ProtoScript M-MuLV Taq RT-PCR kit (New England Biolabs, #E6400S) according to the manufacturer's instructions. Real-time quantitative PCR (RT-qPCR) reactions were performed using a Maxima SYBR Green qPCR master mix (Thermo Fisher Scientific, #K0223), 0.2 μM primers and 50 ng cDNA in a final volume of 20 μl. The PCR reactions were started by denaturation for 10 min. at 95 °C followed 45 cycles of denaturation at 95 °C for 15 s, annealing at 60 °C for 30 s and extension at 72 °C for 30 s. Ct values were normalized to those of *GAPDH* and calibrated using Ct values of controls. Relative gene expression levels were calculated using the ΔΔCt method and presented as log2 fold change values [34]. A cutoff of ±0.8 was used to define differential expression. The primers used for cDNA amplification of *ZEB2* were: 5′-CAAGGAGCAGGTAATCGCAAGT-3′ and 5′-GGAACCAGAATGGGAGAAACG-3′, whereas those used for mouse *Zeb2* were: 5′-ACCGACTCAAGGAGACAGAT-3′ and 5′-GAGTGGATATGCTGTGGTTCTC-3′, for *MAP7* were: 5′-GTGGACTTATCTTTGCCTTACCT-3′ and 5′-GCCAAACACAGTCAACATTCC-3′, for *BOK* were: 5′-GATGGACTGATGTCCTCAAGTG-3′ and 5′-GGCAGCAGCACGAAGAA-3′ and for *GALNT3* were: 5′-GGCACAACATCCAGAAGGAA-3′ and 5′-CATATCTGCTCTCCAGTGACAAC-3′. The primers used for cDNA amplification of *GAPDH* (normalization) were: 5′- GGCTGAGAACGGGAAGCTTGTCAT-3′ and 5′-CAGCCTTCTCCATGGTGGTGAAGA-3′.

For the ChIP-qPCR assays, ChIP-DNAs immunoprecipitated by 6E5 and isotype control antibodies, as well as 1% input DNA, were amplified using the primers mentioned below. Input Ct values were adjusted to 100% as follows: Ct(1% input)-6,64. The percent input values for the ChIP DNAs immunoprecipitated by 6E5 and isotype control antibodies were determined using the following formula: $100 \times 2^{(\text{Adjusted input-Ct}_{IP})}$. The fold enrichment of the immunoprecipitated

chromatin was calculated by dividing the 6E5 percent input value to that of the isotype control antibody. The primers used for ChIP-qPCR analysis of the proximal promoter region (−323 to −188) of *GALNT3* were: 5′-TCCCGCCTGCTCTAACG-3′ and 5′-GACAGCAACCGGAGTCG-3′, of the proximal promoter region (−323 to −175) of *MAP7* were: 5′-GACAGGTGAGCGCAGTC-3′ and 5′-CCGCTCTTCCAGCCGAG-3′, of the −22 k region (−21,967 to −22,080) of *LPCAT1* were: 5′-CAAACCGGGAAGTCATCTTG-3′ and 5′-TGCACAAGGGTTCTGTTCTG-3′ and of the −35 k region (−34,719 to −34,627) of *CDH4* were: 5′-AGATGCAGATGGGATGTAGC-3′ and 5′-TGCCTGTATTGCACCCACAC-3′. The primers used for the +13 k region (+13,394 to +13,491) of *BOK* were: 5′-TCTTGTGTTGGAGCCTTCAG-3′ and 5′-CACCGTGCACCAGAATATCA-3′ and the +165 k region (+164.973 to +165.085) of *ADARB2* were: 5′-CGCCACCAGCATATAGAAGAG-3′ and 5′- CAAAACAAACACCCTCACAGTC-3′. The primers used for the miR200c-141 promoter region were as reported before [35]: 5′-CAGGAGGACACACCTGTGC-3′ and 5′-TCCCCTGGTGGCCTTTAC-3′, whereas those used for the miR200b promoter region were: 5′-AAGGTGGGGGCGGGACGGA and 5′-TGGGAGGCGGTGGCGAGGTG-3′. The location of the respective amplicons was determined using the NCBI Genome database reference genome (hg19).

2.8. ZEB2 and GALNT3 expression analyses in primary human cancers

The *ZEB2* and *GALNT3* mRNA expression levels were assessed by RT-qPCR using 96-well cancer cDNA arrays (TissueScan) covering 8 different tumor types, including breast adenocarcinoma, colon adenocarcinoma, kidney carcinoma, liver carcinoma, ovarian adenocarcinoma, thyroid carcinoma, lung carcinoma and prostate adenocarcinoma. For each cancer type, cDNAs from 9 tumors and 3 corresponding normal samples were included. The RT-qPCR assays were performed in duplicate after which the *GALNT3* and *ZEB2* expression levels in the tumor tissues were calculated relative to those in the corresponding normal tissues using the ΔΔCt method. The same tumor types were analyzed using Expression Project for Oncology (exp*O*) gene expression datasets (GEO accession #GSE2109), whereas for comparative GALNT3 and ZEB2 protein expression analyses immunohistochemistry data available through the Human Protein Atlas database (http://www.proteinatlas.org) were used. The protein expression levels were scored according to staining intensities and fractions of stained cells, and presented as not detectable (ND), low, medium or high.

2.9. Statistical analyses

Gene expression levels in cell lines and primary tumors on the TissueScan cDNA array are presented as mean ± SD (standard deviation). Paired Student's *t* test was used for statistical analyses of experimental data. Significant differences were denoted as follows: $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$. Correlations of *ZEB2* and *GALNT3* mRNA expression levels derived from the TissueScan cDNA array and exp*O* dataset analyses, and correlations of ZEB2 and GALNT3 protein expression levels derived from the Human Protein Atlas dataset analysis, were calculated by Pearson's correlation coefficient using Hmisc 4.0.3 R package version 3.4.0. Negative and positive *r*-values denote downhill and uphill linear relationships, respectively. Statistical analyses of differential ZEB2 and GALNT3 protein expression levels were carried out by Wilcoxon Rank Sum test using MASS R package version 3.4.0.

Data Availability The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE103048 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103048).

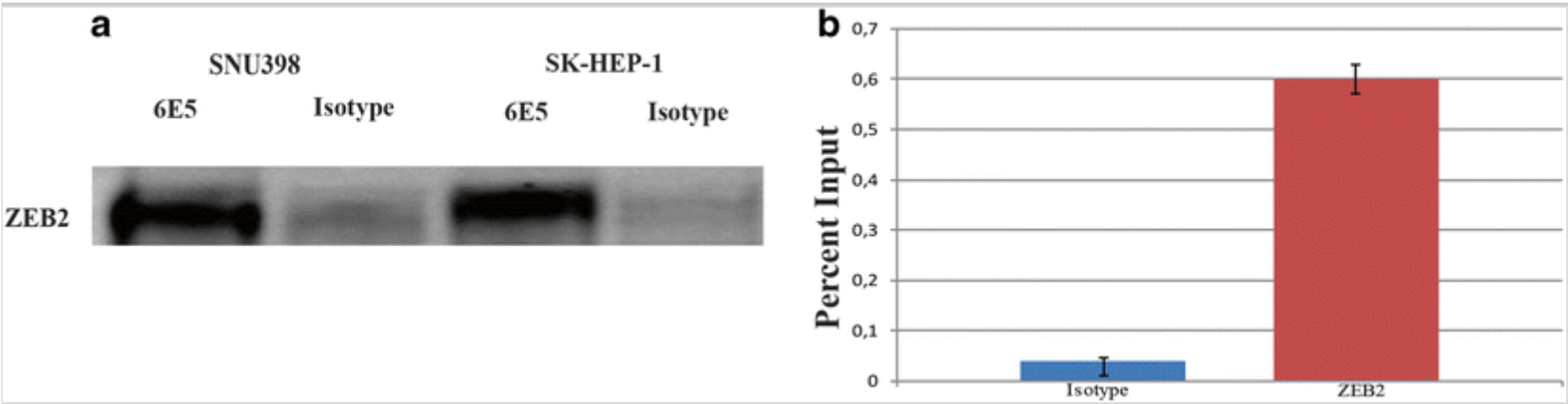
3. Results

3.1. Selection of cells and validation of anti-ZEB2 antibody for ChIP sequencing

Previously, we found that the hepatocellular carcinoma (HCC) cell line SNU398 exhibits a high endogenous ZEB2 protein expression level [23]. Therefore, we selected this cell line for our current ChIP sequencing experiments. For these experiments, we first assessed our homemade anti-ZEB2 monoclonal antibody 6E5 for its suitability. To this end, we performed a ChIP-western experiment using SNU398 and another (mesenchymal-like) HCC cell line, SK-HEP-1, and found that the 6E5 antibody reacts with the ZEB2 protein under ChIP conditions, including formaldehyde fixation and fragmentation by sonication (Fig. 1a). Next, we set out to validate our 6E5 antibody in a ChIP-qPCR assay in which the enrichment of a well-known ZEB2 target region within the miR200c-141 promoter was assessed. We found that, compared to the isotype control antibody used, 6E5 enriched ChIP DNA fragments encompassing the miR200c-141 promoter region by 15 fold (Fig. 1b).

Fig. 1

Validation of ZEB2 antibody 6E5 for ChIP analysis. **a** ChIP-Western blot using SNU398 and SK-HEP-1 cell lysates after fixation with formaldehyde and sonication. Immunoprecipitations were carried out with 6E5 and isotype control antibodies, after which the protein bands were detected using the 6E5 (ZEB2) antibody. **b** RT-PCR results of the miR200c-141 promoter region using DNA derived from the 6E5 and isotype antibody ChIP experiments. Enrichment of the target region relative to the ChIP input control is 15-fold



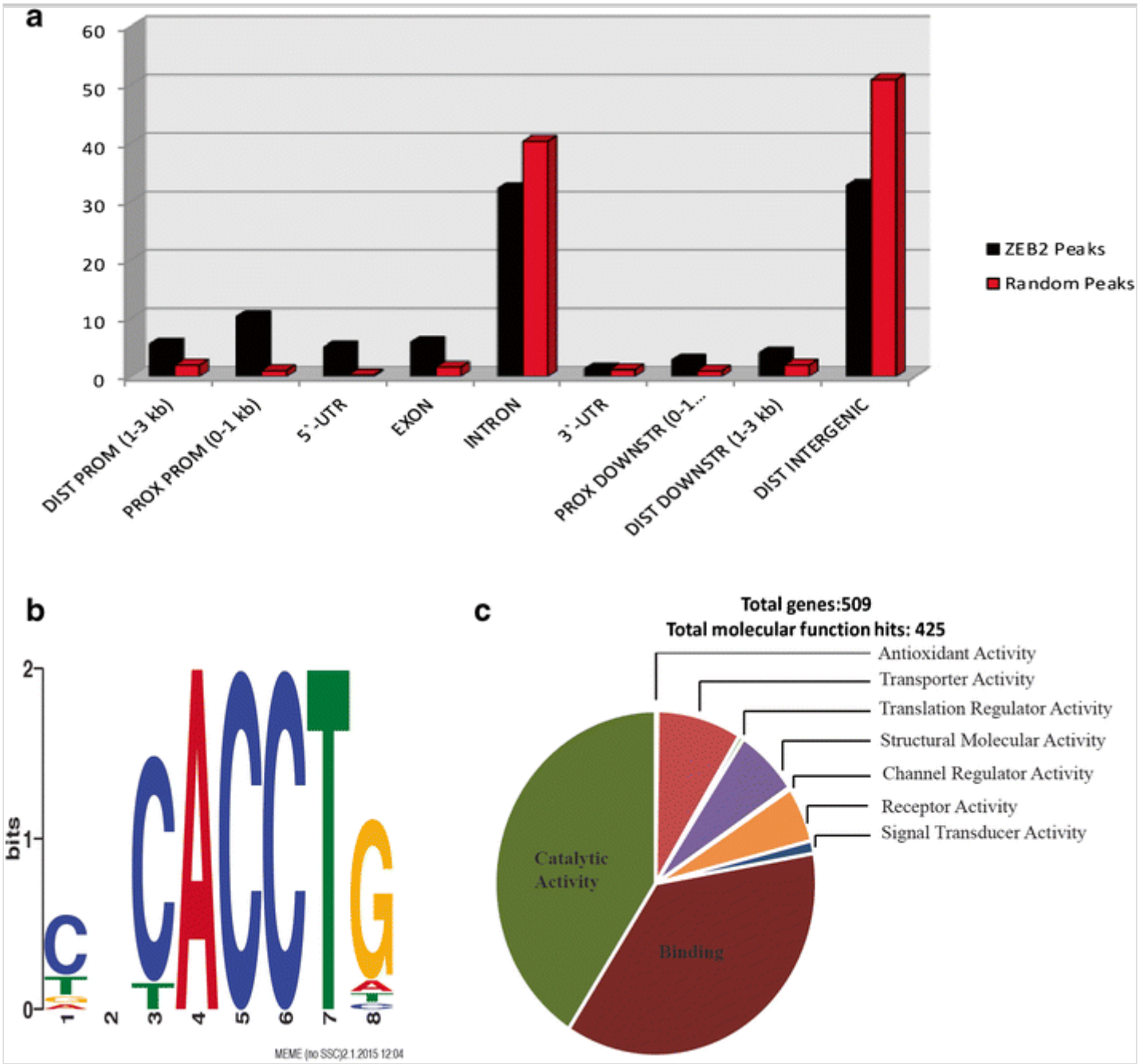
3.2. ChIP sequencing and genomic mapping of ZEB2 binding sites

Next, our ChIP-validated anti-ZEB2 monoclonal antibody 6E5 was used to assess genome-wide binding of ZEB2 by ChIP sequencing (ChIP-seq) in SNU398 cells. We found that 21 million reads in the ZEB2-immunoprecipitated and input samples analyzed by the MACS algorithm with reference to the human reference genome (hg19) generated 554 genomic intervals. Further analysis of these intervals indicated that ZEB2 was bound to the margins of 509 genes. Next, a more detailed genomic map of ZEB2 binding sites was established for the distal promoter, proximal promoter, 5′-UTR, exon, intron, 3′-UTR, proximal downstream, distal downstream and distal intergenic regions. Compared to the distribution of random controls, an enrichment was observed for both the promoter (16% versus 3%) and 5′-UTR (5% versus 0%) regions (Fig. 2a). A subsequent MEME-TOMTOM based analysis of interval sequences pulled down by the 6E5 antibody revealed a consensus CACCTG E-box sequence as common ZEB2 motif (Fig. 2b). Gene Ontology classification using the

PANTHER database led to a clustering of the 509 ZEB2 targets to various functions and processes. Of these, two subcategories of catalytic activity, i.e., kinase activity and enzyme regulator activity, showed a significant enrichment (Fig. 2c).

Fig. 2

Analysis of ZEB2 ChIP-seq data. **a** Genome-wide mapping of ZEB2-binding regions (ZEB2 peaks) within the SNU398 genome relative to controls (random peaks). A detailed genomic map of ZEB2 binding sites was established for distal promoter (−3 kb to −1 kb from TSS), proximal promoter (−1 kb to 0 kb from TSS), 5′-UTR, exon, intron, 3′-UTR, proximal downstream (0 to +1 kb from TSS), distal downstream (+1 kb to +3 kb from TSS) and distal intergenic (> 3 kb from gene) regions. **b** ZEB2 ChIP-seq target motif matching with the ZEB2 consensus motif CACCTG. (**c**) ZEB2 target classification through PANTHER gene ontology analysis and Bonferroni correction. The kinase activity and enzyme regulator activity subcategories (not shown) exhibited 2.02 and 1.98-fold enrichments, respectively ($p < 0.05$). Both were classified under the catalytic activity function with a 1.34-fold enrichment ($p < 0.01$)


a
60
50
40
30
20
10
0
DIST PROM (1-3 kb)
PROX PROM (0-1 kb)
5'-UTR
EXON
INTRON
3'-UTR
PROX DOWNSTR (0-1...
DIST DOWNSTR (1-3 kb)
DIST INTERGENIC
ZEB2 Peaks
Random Peaks
b
bits
2
1
0
1
2
3
4
5
6
7
8
MEME (no SSC)2.1.2015 12:04
c
Total genes:509
Total molecular function hits: 425
Antioxidant Activity
Transporter Activity
Translation Regulator Activity
Structural Molecular Activity
Channel Regulator Activity
Receptor Activity
Signal Transducer Activity
Catalytic Activity
Binding

3.3. ChIP validation of gene targets and their ZEB2-dependent expression

After screening of genomic intervals for the presence of E-box motifs, *CDH4*, *ADARB2* and *LPCAT1* gene regions that were found to contain multiple ZEB2 consensus elements were selected for ChIP validation assays using the 6E5 and isotype control antibodies. ChIP-qPCR was performed using primers flanking the ZEB2 binding motifs within the intervals. By doing so, enrichments comparable to that for miR200b (29.5) were found for the *CDH4*, *ADARB2* and *LPCAT1* loci (22.4, 19.7 and 21.7, respectively) (Fig. 3). Next, we generated stable silencing ZEB2-shRNA-SNU398 (sh-ZEB2) and non-silencing control-shRNA-SNU398 (NSC) cell clones to assess ZEB2-dependent expression of target genes by RT-qPCR (Supplementary Fig. S1). To this end, we narrowed down our list of 509 targets to 62 genes whose associated intervals were found to display high peak values, to comprise ZEB2 DNA-binding motifs and/or to be located in putative promoter regions (−7500 bp to +2500 bp from the transcription start site, TSS). By doing so, we found differential expression of the *MAP7*, *BOK* and *GALNT3* genes between the sh-ZEB2 and NSC cell clones. In the ZEB2 knockdown cells, the expression of *BOK* and *MAP7* was found to be decreased (−0.8 and −1.2 fold, respectively), whereas a 3.96-fold increase in *GALNT3* expression was detected. These results suggest a positive correlation of *BOK* and *MAP7*, and a negative correlation of *GALNT3*, with *ZEB2* expression (Fig. 4). In addition, we validated the binding of ZEB2 to its consensus motif within the intervals associated to these genes by ChIP-qPCR and found 25.5 and 27.8 fold enrichments for *MAP7* and *BOK*, respectively. The highest enrichment (85 fold) was observed for the *GALNT3* gene (Fig. 5).

Fig. 3

ChIP validation of ChIP-seq targets. RT-qPCR was performed with ChIP DNA obtained from SNU398 cells using the 6E5 and isotype control antibodies. The fold enrichments for ZEB2-binding regions of the *LPCAT1*, *ADARB2* and *CDH4* intervals were calculated by dividing the ZEB2% input by the control % input. The ZEB2 target miR200b was used as positive control

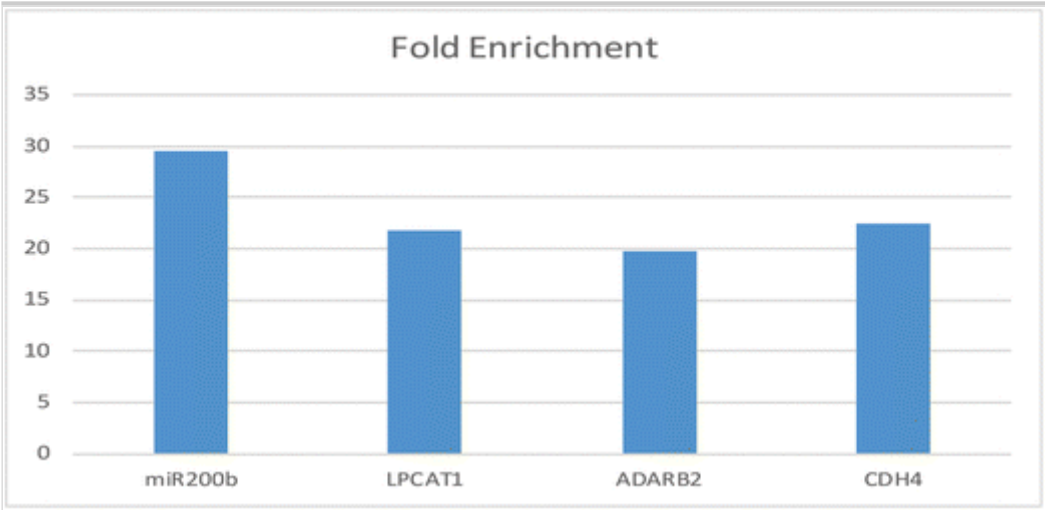


Fig. 4

Differential gene expression of the *BOK*, *MAP7* and *GALNT3* genes after ZEB2 knockdown in SNU398 cells. RT-qPCR results from sh-ZEB2 and control NSC cells are shown. Relative gene expression levels were calculated using the ΔΔCt method

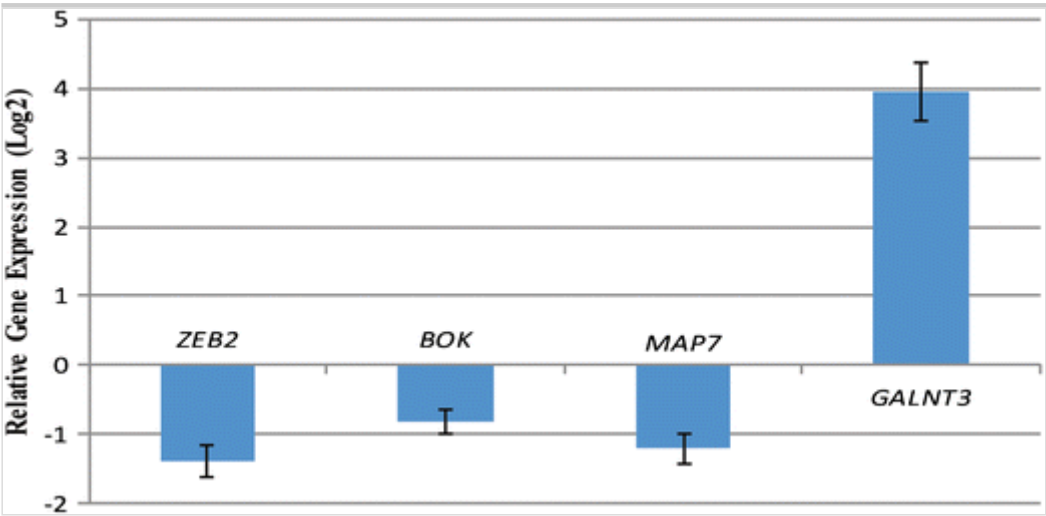
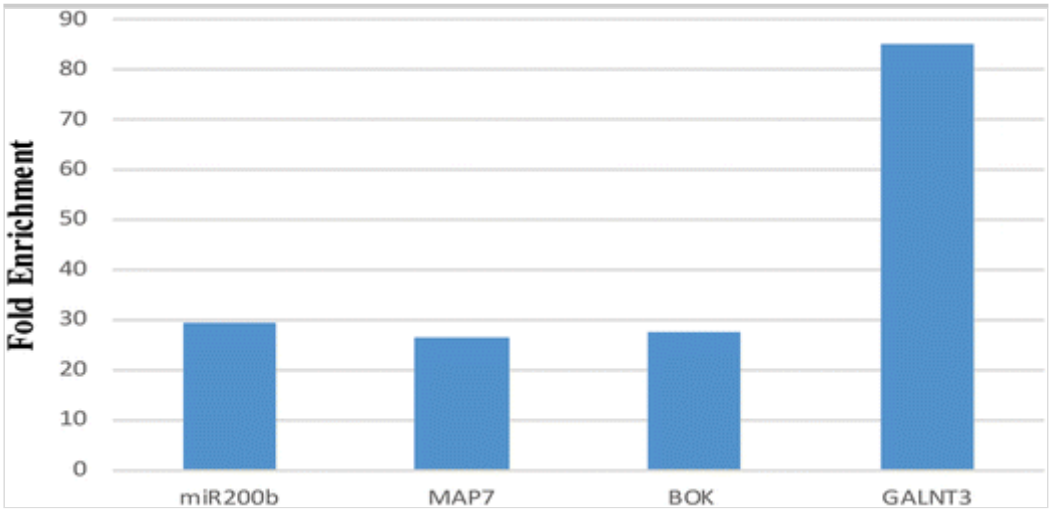


Fig. 5

ChIP validation of differentially expressed genes in ZEB2 knockdown and control cells. Primers flanking ZEB2 binding motifs within the *MAP7* (+80 bp from TSS), *BOK* (+13,936 bp from TSS) and *GALNT3* (+179 bp from TSS) genes were used for ChIP-qPCR to amplify DNA fragments immunoprecipitated with the 6E5 and isotype (control) antibodies. Fold enrichment is presented as the ratio of % 6E5 antibody input to that of the control antibody. miR200b (29.5-fold enrichment) was used as positive control



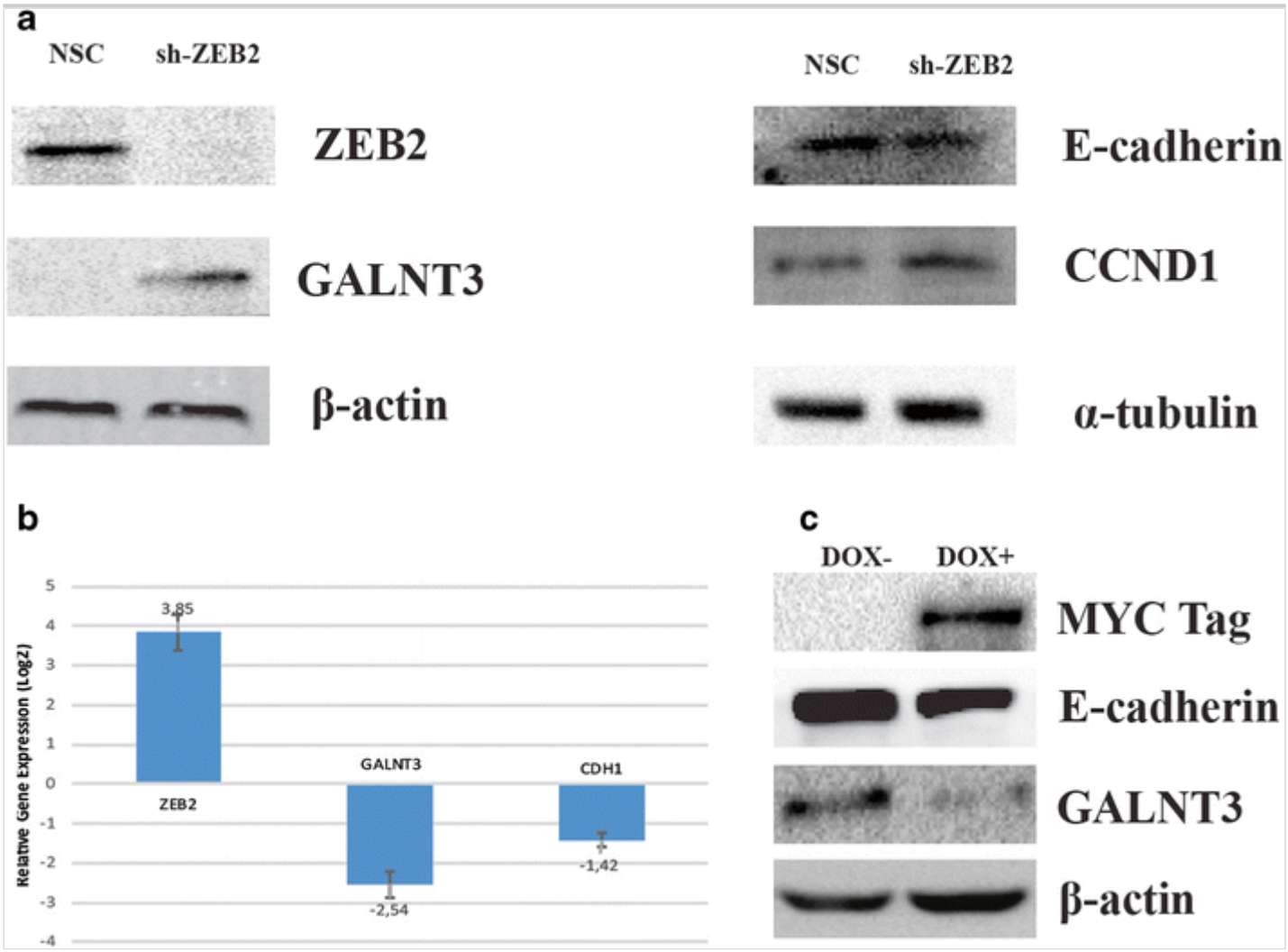
3.4. ZEB2-dependent expression of GALNT3 in cancer-derived cell lines

The observed upregulation of *GALNT3* in ZEB2 knockdown cells and its significant fold enrichment in the ChIP assay ($p < 0.001$) prompted us to study the putative ZEB2-dependent expression of GALNT3 in more detail. Parallel to its high transcript level, we also observed a marked increase in GALNT3 protein level in ZEB2 knockdown cells (sh-ZEB2 clone, Fig. 6a). The same sh-ZEB2 clone was used to assess the protein expression levels of the well-established ZEB2 targets CCND1 and CDH1. We found that the level of CCND1 was increased, but that of CDH1 was unchanged. Next, we tested the expression of GALNT3 in a doxycycline-inducible ZEB2 model (DLD-ZEB2) at both its transcript and protein levels.

Compared to the uninduced cells, we found that doxycycline induction resulted in a 3.85 fold increase in *ZEB2* expression as assessed by RT-qPCR. This increase was accompanied by 1.42 and 2.54 fold decreases in *CDH1* ($p < 0{,}001$) and *GALNT3* ($p < 0{,}001$) expression, respectively (Fig. 6b). Subsequent Western blot results were in complete accordance with the RT-qPCR results, suggesting that ZEB2 plays a direct role in GALNT3 downregulation (Fig. 6c). The decreases in both GALNT3 transcript and protein levels were more apparent than those of CDH1 (Fig. 6b, c).

Fig. 6

GALNT3 is negatively regulated by ZEB2. **a** Western blots showing differential GALNT3 and CCND1 protein expression, but unaltered E-cadherin expression in ZEB2 knockdown and NSC control cells. β-actin and α-tubulin were used as loading controls. **b** RT-qPCR results using cDNA derived from doxycycline-induced (2 μg/ml) and uninduced DLD1 cells. Bars represent relative *ZEB2*, *GALNT3* and *CDH1* expression levels determined using the ΔΔCt method. **c** Western blot analysis of E-cadherin and GALNT3 in doxycycline-induced and uninduced DLD-ZEB2 cells. ZEB2 was detected using a Myc-tag antibody. β-actin was used as loading control



3.5. Negative correlation between ZEB2 and GALNT3 expression in primary tumor tissues

The changes observed in GALNT3 expression in ZEB2 knockdown and ZEB2 induced cell lines led us to assess their relative expression *in vivo*. To this end, the individual expression of *ZEB2* or *GALNT3* and their co-expression was assessed using a cancer cDNA array encompassing eight different cancer types, each represented by nine tumor samples and three

corresponding normal tissue samples. Detailed clinicopathological information on the patients and tumors is listed in Supplementary Table 1 (Table S1). Compared to the normal tissues, an apparent decrease in *ZEB2* expression was found in most of the tumor tissues included. The only exception was breast cancer, showing an apparent increase in *ZEB2* expression. Conversely, we found that the *GALNT3* expression levels were substantially increased in ovarian cancers and sharply decreased in kidney cancers. Following previous cues, we next asked whether a reciprocal correlation exists between *GALNT3* and *ZEB2* expression. Significant reciprocal *ZEB2* and *GALNT3* expression levels were indeed observed in kidney, ovarian, prostate and liver tumors. In contrast, we found that the expression level of *ZEB2* exhibited a positive correlation with that of *GALNT3* in breast, lung and thyroid cancers, of which the breast and lung cancer data were non-significant. Overall, we conclude that several primary cancers show a reciprocal correlation, suggesting that ZEB2 may act as a *GALNT3* repressor in these cancers (Table 1 and Supplementary Fig. S2).

Table 1

Correlations between *ZEB2* and *GALNT3* mRNA expression in primary human tumors

Cancer type	*N*	Gene ID	ΔΔCt mean[a]	Differential expression	Correlation coefficient *r*
Breast	9	*ZEB2*	0.70	NS	0.34
		GALNT3	0.66		
Colorectal	9	*ZEB2*	−0.90	NS	−0.30
		GALNT3	0.33		
Kidney	9	*ZEB2*	0.10	$p < 0.001$	−0.02
		GALNT3	−2.74		
Ovarian	9	*ZEB2*	−2.60	$p < 0.001$	−0.14

[a] Mean relative mRNA expression values for *ZEB2* and *GALNT3* per cancer type and all cancers together. For the individual tumor samples, relative *ZEB2* and *GALNT3* mRNA expression levels were estimated by calibration of the mean Ct values of 3 normal tissues in each cancer type and normalization to *ACTNB1*. Differential expression was analyzed by Student's *t* test. The correlation coefficient *r* was calculated using Pearson's correlation test

NS not significant

Cancer type	*N*	Gene ID	ΔΔCt mean	Differential expression	Correlation coefficient *r*
Lung	9	*GALNT3*	2.98	NS	0.30
		ZEB2	−3.10		
		GALNT3	−0.53		
Prostate	9	*ZEB2*	−0.50	$p < 0.05$	−0.46
		GALNT3	0.44		
Liver	9	*ZEB2*	−1.10	$p < 0.05$	−0.36
		GALNT3	0.06		
Thyroid	9	*ZEB2*	−2.6	$p < 0.05$	0.11
		GALNT3	−1.48		
All	72	*ZEB2*	−1.25	$p < 0.05$	−0.10
		GALNT3	−0.04		

[a]Mean relative mRNA expression values for *ZEB2* and *GALNT3* per cancer type and all cancers together. For the individual tumor samples, relative *ZEB2* and *GALNT3* mRNA expression levels were estimated by calibration of the mean Ct values of 3 normal tissues in each cancer type and normalization to *ACTNB1*. Differential expression was analyzed by Student's *t* test. The correlation coefficient *r* was calculated using Pearson's correlation test

NS not significant

Because of the limited number of samples available through the cDNA arrays, we extended our *ZEB2* and *GALNT3* expression analyses to publicly available Expression Project for Oncology (exp*O*) datasets (GEO accession GSE2109, encompassing data from 2158 solid tumor arrays) from tumor tissues of the same eight histotypes (www.intgen.org/expo /). From the 2158 arrays, data from 1415 arrays encompassing 352 breast, 293 colorectal, 280 kidney, 199 ovarian, 132 lung, 83 prostate, 44 liver and 32 thyroid tumors were subjected to Pearson's correlation analysis. By doing so, a significant negative correlation ($r = -0.36$; $p < 0.001$) between *ZEB2* and *GALNT3* expression was found for all 1415

samples. When cancer types were individually analyzed, we found that *ZEB2* expression showed a significant inverse correlation with *GALNT3* expression in breast, colorectal, ovarian and prostate cancers. In addition, thyroid, liver, kidney and lung cancers showed a negative correlation (Table 2 and Supplementary Fig. S3). A significant differential expression between *ZEB2* and *GALNT3* was found for all compiled cancer types and datasets (Table 2).

Table 2

Correlations of ZEB2 and GALNT3 mRNA expression levels in human tumors

Cancer type	*n*	Gene ID	Probe intensity mean	Differential expression	Correlation coefficient *r*
Breast	352	*ZEB2*	643,72	$p < 0.0001$	−0.19***
		GALNT3	1821,33		
Colorectal	293	*ZEB2*	508,71	$p < 0.0001$	−0.38***
		GALNT3	4244,68		
Kidney	280	*ZEB2*	995,30	$p < 0.0001$	−0.10
		GALNT3	383,40		
Ovarian	199	*ZEB2*	455,94	$p < 0.0001$	−0.25***
		GALNT3	3490,91		
Lung	132	*ZEB2*	664,44	$p < 0.0001$	−0.05
		GALNT3	2309,37		
Prostate	83	*ZEB2*	500,79	$p < 0.0001$	−0.44***
		GALNT3	3665,87		

n indicates the number of array datasets per cancer type. Out of the probes targeting *ZEB2* and *GALNT3* (5 and 2, respectively) probes with the highest mean intensities were selected and analyses were performed using one single probe value for each gene. Differential expression levels and correlations between *ZEB2* and *GALNT3* were analyzed using Student's t-test and Pearson's correlation coefficient test, respectively. *** indicates a significant negative correlation ($p < 0.001$)

Cancer type	*n*	Gene ID	Probe intensity mean	Differential expression	Correlation coefficient *r*
Liver	44	*ZEB2*	554,73	$p < 0.0001$	−0.16
		GALNT3	2083,65		
Thyroid	32	*ZEB2*	556,1375	$p < 0.0001$	−0.23
		GALNT3	1975,45		
All	1415	*ZEB2*	1269,46	$p < 0.0001$	−0.36***
		GALNT3	2438,76		

n indicates the number of array datasets per cancer type. Out of the probes targeting *ZEB2* and *GALNT3* (5 and 2, respectively) probes with the highest mean intensities were selected and analyses were performed using one single probe value for each gene. Differential expression levels and correlations between *ZEB2* and *GALNT3* were analyzed using Student's t-test and Pearson's correlation coefficient test, respectively. *** indicates a significant negative correlation ($p < 0.001$)

Next, we compared the protein expression levels of ZEB2 and GALNT3 in human tumor tissues as listed in the Human Protein Atlas database (http://www.proteinatlas.org/ENSG00000169554-ZEB2/cancer for ZEB2 and http://www.proteinatlas.org/ENSG00000115339-GALNT3/cancer for GALNT3, respectively) [36]. The ZEB2 and GALNT3 protein expression levels were assessed in paired normal-tumor tissues from 9 breast, 7 colorectal, 7 lung, 7 melanoma, 5 liver, 5 kidney, 4 glioma and 3 ovarian cancer patients. Consistent with our RT-qPCR data, we found that ZEB2 was downregulated in most of the cancers, except for gliomas and a few melanoma cases. A significant difference between ZEB2 and GALNT3 protein expression was found in all 48 paired tumor samples ($p < 0.001$), and individually in the breast ($p < 0.001$), colorectal, liver, lung ($p < 0.01$) and kidney ($p < 0.05$) tumor samples (Supplementary Fig. S4). Correlation analyses were performed in the same set of tumors and, by doing so, an inverse correlation between ZEB2 and GALNT3 expression was found for all tumors, except melanoma (Table 3).

Table 3

Correlations of ZEB2 and GALNT3 protein expression levels in human tumors

Cancer type	Gene ID	Protein expression in tumor tissues (*n*)				Correlation coefficient
		High	Medium	Low	ND	*r*
Breast	ZEB2	0	0	0	9	−0.67
	GALNT3	5	3	1	0	
Colorectal	ZEB2	0	0	0	7	−0.22
	GALNT3	5	1	0	1	
Lung	ZEB2	0	0	0	7	−0.52
	GALNT3	3	2	1	1	
Melanoma	ZEB2	0	1	0	7	0.56
	GALNT3	0	3	2	3	
Liver	ZEB2	0	0	0	5	−0.87
	GALNT3	2	2	1	0	
Renal	ZEB2	0	0	0	5	−0.13
	GALNT3	0	1	3	1	
Glioma	ZEB2	1	2	0	1	−0.71
	GALNT3	0	0	2	2	
Ovarian	ZEB2	0	0	0	3	−0.52
	GALNT3	2	1	0	0	

Tumor tissues were paired according to the patient information available through the Human Protein Atlas database. "*n*" represents the number of patients per staining pattern. "*r*" represents correlation coefficient values as calculated by Pearson's correlation coefficient test

ND not detectable

Cancer type	Gene ID	Protein expression in tumor tissues (*n*)				Correlation coefficient
		High	Medium	Low	ND	*r*
All	ZEB2	1	3	0	44	−0.67
	GALNT3	17	13	10	8	
Tumor tissues were paired according to the patient information available through the Human Protein Atlas database. "*n*" represents the number of patients per staining pattern. "*r*" represents correlation coefficient values as calculated by Pearson's correlation coefficient test						
ND not detectable						

4. Discussion

Epithelial-to-mesenchymal transition (EMT) is not only associated with normal embryonic development, but also with cancer invasion and metastasis [37, 38, 39, 40]. ZEB2, also known as SIP1, is one of the EMT inducers that were initially shown to downregulate E-cadherin expression by binding to E-box sequences present in the minimal promoter of its encoding gene, *CDH1* [4]. Repression of other epithelial junction proteins by ZEB2 has also been observed in different cellular and pathophysiological contexts. In addition, it has been found that doxycycline-induced expression of ZEB2 in DLD1 and A431 cells may result in epithelial dedifferentiation and decreases in E-cadherin, P-cadherin, claudin 4, tight junction protein 3, plakophilin 2, desmoplakin, connexin 26 and connexin 31 mRNA expression levels, and significant decreases in desmoplakin, plakophilin 2 and claudin 4 protein expression levels [13]. Besides its role in regulating the expression of junction proteins, other studies have indicated a role of ZEB2 in distinct cellular processes, including those regulating cell cycle progression, cellular senescence, neural development, osteogenesis and tumor stemness. ZEB2 was also found to act as a direct negative regulator of CCND1, causing G1 cell cycle arrest in ZEB2-induced cells, suggesting impaired cell cycle progression as one of the characteristics of cells undergoing EMT [16]. The finding that ZEB2 acts as a mediator of TGF-β induced repression of hTERT led us to define a role of ZEB2 in replicative senescence that spontaneously occurs in HCC cells with a p53/p16^{INK4a} deficient background [14, 15]. Later, GADD45G has been described as an upstream inducer of ZEB2 expression in liver cells, and it has been suggested that GADDG45/ZEB2

downregulation may act as a senescence evasion mechanism during hepatocarcinogenesis [41]. However, as the telomerase promoter is commonly mutated allowing activation in HCC, ZEB2 can be expressed at later stages of HCC development and induce EMT. Another target of ZEB2, i.e., liver/bone/kidney alkaline phosphatase (LBK-ALP), has been identified in a study on bone morphogenetic protein (BMP)-induced osteogenesis. It was found that ZEB2 can bind to the E2 box present in the *ALP* gene promoter and, by doing so, represses its expression [42]. It has also been found that ZEB2 inactivation in mouse cortical precursor cells results in an entire deficiency of the hippocampus, with a concomitant activation of the Wnt antagonist Secreted Frizzled-Related Protein 1 (Sfrp1) and downregulation of the non-canonical Wnt effector JNK [26]. The putative role of ZEB2 in stemness was substantiated by another study pointing out that normal and neoplastic human breast stem cells with a $CD44^{high}CD24^{low}$ phenotype express EMT markers, including ZEB2 [43].

In spite of the multiplicity of biological processes regulated by ZEB2, and the presence of a myriad of E-box sequences within the human genome, only a few direct ZEB2 target genes have so far been identified [4, 13, 14, 16, 26, 44]. A further understanding of the biology of ZEB2 is, however, of critical importance in order to elucidate its role in EMT and cancer progression/metastasis. Therefore, we aimed at establishing a genome-wide map of ZEB2 binding sites in a high endogenous ZEB2 expressing cell line, SNU398, by ChIP sequencing using our previously described monoclonal antibody 6E5 [24]. After ChIP-seq, 554 genomic intervals were identified corresponding to 509 annotated gene regions. A MEME-TOMTOM database search that revealed and confirmed the presence of a ZEB2 consensus sequence (CACCTG) within the intervals identified, indicated the effectiveness of our ChIP-seq approach. We validated our results through ChIP-qPCR analysis of the *CDH4*, *ADARB2* and *LPCAT1* gene regions containing multiple ZEB2-binding sequences. Although these genes have not been defined as ZEB2 targets before, they displayed enrichment comparable to a well-established ZEB2-regulated gene, miR200b. Interestingly, *CDH1* was not among the ZEB2 targets in our ChIP-Seq study, neither were we able to detect any enrichment by ChIP in SNU398 cells. It may, therefore, be plausible to state that E-cadherin expression regulation may not depend on ZEB2 in these cells, which have been shown to express CDH1 and other EMT inducing transcription factors [28]. In addition, E-cadherin silencing has previously been shown to be mediated by CpG methylation of the *CDH1* gene promoter in cell lines derived from liver, stomach, lung and bladder cancers [45]. Another well-established ZEB2 target, CCND1, was however found to exhibit increased protein expression in shZEB2-SNU398 cells.

Using a limited ZEB2 target list, resulting from selecting gene regions with high peak values, multiple E-boxes and/or ZEB2-binding sites within their promoters, 62 genes that met these criteria were subjected to expression studies in ZEB2-

silenced and control SNU398 cells. By doing so, we found that only a few genes, including *BOK*, *MAP7* and *GALNT3*, exhibited a differential expression, i.e., *BOK* and *MAP7* expression paralleled ZEB2 downregulation indicating a ZEB2-dependent induction, whereas *GALNT3* was found to be upregulated upon ZEB2 knockdown, hinting at ZEB2-dependent repression. To assess whether ZEB2 may directly modulate the expression of these genes, we performed ChIP assays and, by doing so, readily found an enrichment of *BOK* and *MAP7*, comparable to that of control miR200b. Remarkably, we found that the enrichment of *GALNT3* was about three-fold higher than that of miR200b. As of yet, the expression regulation of *BOK*, *MAP7* and *GALNT3* has not been linked to ZEB2. Here, we propose them as novel ZEB2 targets. For our further analyses, we focused on *GALNT3* since it displayed the highest expression and ChIP enrichment values.

GALNT3 is a member of N-Acetyl Galactosamine Transferase enzyme family, and mediates the transfer of N-acetyl galactosamine to serine or threonine residues. Hence, it controls the initial steps of mucin type *O*-glycosylation of polypeptides within the Golgi apparatus [46]. Mucin type *O*-glycosylation is one of the most abundant forms of glycosylation of both membrane bound and secreted proteins, and represents an important process underlying the biology of cancer development and metastasis, given the fact that posttranslational glycosylation that accompanies neoplastic transformation has a significant impact on the behavior of cancer cells and the progression of malignant diseases [47]. Comparison of GALNT3 protein levels between ZEB2 knockdown and control SNU398 cells revealed a significant increase in GALNT3 protein levels upon ZEB2 knockdown, validating our previous observation that *GALNT3* expression may be negatively regulated by *ZEB2*. To exclude the possibility that *GALNT3* expression modulation by ZEB2 is confined to our cell systems, we assessed changes in both GALNT3 transcript and protein levels in DLD-ZEB2 cells that ectopically expresses ZEB2 upon doxycycline induction. We found that exogenous ZEB2 expression in these colorectal carcinoma-derived cells repressed GALNT3 expression at both levels, along with that of the canonical ZEB2 target E-cadherin. Our observation that GALNT3 is negatively regulated by ZEB2 in these different cell systems prompted us to extend our studies to investigate ZEB2 and GALNT3 expression correlations in primary human tumor samples. Through the analysis of cDNA samples derived from eight different human cancer types, we observed an apparent downregulation of ZEB2 in most of the tumor tissues compared to that in normal tissues. We also found that GALNT3 exhibited a clear decrease in kidney and thyroid cancers and a marked increase in ovarian cancer, with no distinctive changes in other cancer types. These results are in line with other studies indicating differential GALNT3 expression in cancer. Comparison of GALNT3 expression in orthotopic xenograft tumors in mice showed increased GALNT3 expression in HEP3B-derived tumors

compared to tumors derived from HCC cell lines with metastatic potential. No significant difference of GALNT3 expression was, however, observed in primary metastatic and non-metastatic human HCC tissues [48]. Through our analysis of GALNT3 expression in nine HCC cell lines, we found that HEP3B cells exhibited the highest transcript levels (data not shown). In colorectal cancer, decreased GALNT3 levels have been associated with a poor survival [49]. These investigators also observed an association of GALNT3 expression with histologic differentiation, with well-differentiated tumors having the highest and poorly differentiated tumors having the lowest GALNT3 expression. In contrast, GALNT3 expression has been found to corresponded to high-grade tumors and a poor survival in renal cell carcinoma patients [50]. In line with this, a strong GALNT3 expression has been observed in more than 1/3 of oral squamous cell carcinoma patients with increased tumor recurrences and decreased disease-free survival rates [51]. Here, we found a ~3-fold increased GALNT3 expression in ovarian cancers compared to its corresponding normal tissues. This result confirms previous findings indicating that ovarian cancer patients whose tumors exhibit high GALNT3 expression levels have a shorter survival, and that the expression of GALNT3 gradually increases from normal ovarian tissues to tumors with a low-malignant potential to high-grade tumors [52]. Our correlation analysis of *GALNT3* and *ZEB2* expression data revealed an inverse correlation between these genes in most of the tumors, as expected. But because of the limited number of tumor samples included, we decided to further explore the publicly available exp*O* dataset. This exercise substantiated our observation that the expression of *ZEB2* is significantly lower than that of *GALNT3*. In addition, we found that this *in silico* analysis strongly supported our *in vitro* and *in vivo* results indicating that ZEB2 negatively regulates *GALNT3* expression in human tumors. Subsequent ZEB2 and GALNT3 protein expression analyses revealed similar findings, including a negative correlation between ZEB2 and GALNT3 expression in all analyzed tumor types, except melanomas. Interestingly, we noticed a decreased GALNT3 protein expression in gliomas which, unlike other tumors, displayed increased ZEB2 protein levels, again substantiating an inverse correlation between ZEB2 and GALNT3 expression. Taken together, we conclude that our results indicate that GALNT3 is negatively regulated by ZEB2.

So far, only a limited number of studies has dealt with the regulation of GALNT3 expression, and most of these were related to bone remodeling in a parathyroid hormone (PTH)-dependent manner [53]. Besides a negative regulation of GALNT3 by PTH, two other transcription factors have been found to positively regulate GALNT3. RUNX2, an essential regulator of chondrocyte maturation, has been found to upregulate GALNT3 and mice overexpressing GALNT3 have been found to exhibit dwarfism, dysregulated chondrocyte proliferation and apoptosis [54]. In addition, it has been reported that

treatment of MCF7 breast cancer cells with the histone deacetylase inhibitor sodium butyrate resulted in p300/CBP-associated factor acetylation and enhanced binding of the nuclear respiratory factor-1 (NRF1) to DNA, resulting in GALNT3 upregulation [55]. Posttranslational glycosylation that accompanies neoplastic transformation may have a significant impact on the behavior of cancer cells and malignant disease progression [47]. A gene expression analysis of pancreatic cancer-derived cell lines clustered these cell lines in two groups, i.e., cell lines with a mesenchymal phenotype showing lower GALNT3 expression levels than those with an epithelial-like phenotype [56]. This study underscores our observation that an inverse correlation exists between EMT induction and GALNT3 expression.

Taken together, we conclude that through our ChIP-Seq analysis new targets have been identified for the transcriptional repressor ZEB2, of which GalNAc-T3 (GALNT3) showed differential expression in a ZEB2-dependent manner in both tumor-derived cell lines and primary tumor samples. The negative regulation of O-glycosylation by ZEB2 adds a new level of complexity to our understanding of the role of ZEB2 and GALNT3 in cancer progression and metastasis. The characterization of proteins glycosylated by GALNT3 may facilitate the identification of new diagnostic and/or therapeutic targets for human cancers.

Acknowledgements

The study was funded by the Scientific and Technical Research Council of Turkey (TUBITAK) grant 111S484 (to T.Y.).

Author’s contributions

T.Y., A.E.S. and P.B-E. designed the study. P.B-E., M.C., I.Y-C. and G.O. performed the experiments. P.B-E., N.T. and T.Y. analyzed and interpreted the data. T.Y., A.E.S. and P.B-E. wrote the manuscript. All authors read and approved the final manuscript.

Compliance with ethical standards

Competing interests The authors declare that they have no conflict of interest.

5. Electronic supplementary material

Supplementary Fig. S1 (PDF 460 kb)

Supplementary Fig. S2 (PDF 523 kb)

Supplementary Fig. S3 (PDF 798 kb)

Supplementary Fig. S4 (PDF 360 kb)

Supplementary Table 1

(XLS 48 kb)

References

1. T. Nagase, K.-i. Ishikawa, N. Miyajima, A. Tanaka, H. Kotani, N. Nomura, O. Ohara, Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. DNA Res **5**, 31–39 (1998)

2. K. Verschueren, J.E. Remacle, C. Collart, H. Kraft, B.S. Baker, P. Tylzanowski, L. Nelles, G. Wuytens, M.-T. Su, R. Bodmer, SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5′-CACCT sequences in candidate target genes. J Biol Chem **274**, 20489–20498 (1999)

3. J.E. Remacle, H. Kraft, W. Lerchner, G. Wuytens, C. Collart, K. Verschueren, J.C. Smith, D. Huylebroeck, New mode of DNA binding of multi-zinc finger transcription factors: δEF1 family members bind with two hands to two target sites. EMBO J **18**, 5073–5084 (1999)

4. J. Comijn, G. Berx, P. Vermassen, K. Verschueren, L. van Grunsven, E. Bruyneel, M. Mareel, D. Huylebroeck, F. Van Roy, The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. Mol Cell **7**, 1267–1278 (2001)

5. A.A. Postigo, J.L. Depp, J.J. Taylor, K.L. Kroll, Regulation of Smad signaling through a differential recruitment of coactivators and corepressors by ZEB proteins. EMBO J **22**, 2453–2462 (2003)

6. L.A. van Grunsven, C. Michiels, T. Van de Putte, L. Nelles, G. Wuytens, K. Verschueren, D. Huylebroeck, Interaction between Smad-interacting protein-1 and the corepressor C-terminal binding protein is dispensable for transcriptional repression of E-cadherin. J Biol Chem **278**, 26135–26145 (2003)

7. A.E. Sayan, T.R. Griffiths, R. Pal, G.J. Browne, A. Ruddick, T. Yagci, R. Edwards, N.J. Mayer, H. Qazi, S. Goyal, SIP1 protein protects cells from DNA damage-induced apoptosis and has independent prognostic value in bladder cancer. Proc Natl Acad Sci U S A **106**, 14884–14889 (2009)

8. G. Maeda, T. Chiba, M. Okazaki, T. Satoh, Y. Taya, T. Aoba, K. Kato, S. Kawashiri, K. Imai, Expression of SIP1 in oral squamous cell carcinomas: implications for E-cadherin expression and tumor progression. Int J Oncol **27**, 1535–1541 (2005)

9. E. Rosivatz, I. Becker, K. Specht, E. Fricke, B. Luber, R. Busch, H. Höfler, K.-F. Becker, Differential expression of the epithelial-mesenchymal transition regulators snail, SIP1, and twist in gastric cancer. Am J Pathol **161**, 1881–1891 (2002)

10. J. Caramel, E. Papadogeorgakis, L. Hill, G.J. Browne, G. Richard, A. Wierinckx, G. Saldanha, J. Osborne, P. Hutchinson, G. Tse, A switch in the expression of embryonic EMT-inducers drives the development of malignant melanoma. Cancer Cell **24**, 466–480 (2013)

11. N. Wakamatsu, Y. Yamada, K. Yamada, T. Ono, N. Nomura, H. Taniguchi, H. Kitoh, N. Mutoh, T. Yamanaka, K. Mushiake, Mutations in SIP1, encoding Smad interacting protein-1, cause a form of Hirschsprung disease. Nat Genet **27**, 369–370 (2001)

12. G. Verstappen, L.A. van Grunsven, C. Michiels, T. Van de Putte, J. Souopgui, J. Van Damme, E. Bellefroid, J. Vandekerckhove, D. Huylebroeck, Atypical Mowat–Wilson patient confirms the importance of the novel association between ZFHX1B/SIP1 and NuRD corepressor complex. Hum Mol Genet **17**, 1175–1183 (2008)

13. C. Vandewalle, J. Comijn, B. De Craene, P. Vermassen, E. Bruyneel, H. Andersen, E. Tulchinsky, F. Van Roy, G. Berx, SIP1/ZEB2 induces EMT by repressing genes of different epithelial cell–cell junctions. Nucleic Acids Res **33**, 6566–6578 (2005)

14. S.-Y. Lin, S.J. Elledge, Multiple tumor suppressor pathways negatively regulate telomerase. Cell **113**, 881–889 (2003)

15. N. Ozturk, E. Erdal, M. Mumcuoglu, K.C. Akcali, O. Yalcin, S. Senturk, A. Arslan-Ergul, B. Gur, I. Yulug, R. Cetin-Atalay, Reprogramming of replicative senescence in hepatocellular carcinoma-derived cells. Proc Natl Acad Sci U S A **103**, 2178–2183 (2006)

16. J. Mejlvang, M. Kriajevska, C. Vandewalle, T. Chernova, A.E. Sayan, G. Berx, J.K. Mellon, E. Tulchinsky, Direct repression of cyclin D1 by SIP1 attenuates cell cycle progression in cells undergoing an epithelial mesenchymal transition. Mol Biol Cell **18**, 4615–4624 (2007)

17. M. Kato, J. Zhang, M. Wang, L. Lanting, H. Yuan, J.J. Rossi, R. Natarajan, MicroRNA-192 in diabetic kidney glomeruli and its function in TGF-β-induced collagen expression via inhibition of E-box repressors. Proc Natl Acad Sci U S A **104**, 3432–3437 (2007)

18. T. Shirakihara, M. Saitoh, K. Miyazono, Differential regulation of epithelial and mesenchymal markers by δEF1 proteins in epithelial–mesenchymal transition induced by TGF-β. Mol Biol Cell **18**, 3533–3544 (2007)

19. P.A. Gregory, A.G. Bert, E.L. Paterson, S.C. Barry, A. Tsykin, G. Farshid, M.A. Vadas, Y. Khew-Goodall, G.J. Goodall, The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1.

Nat Cell Biol **10**, 593–601 (2008)

20. S.-M. Park, A.B. Gaur, E. Lengyel, M.E. Peter, The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. Genes Dev **22**, 894–907 (2008)

21. S. Brabletz, T. Brabletz, The ZEB/miR-200 feedback loop—a motor of cellular plasticity in development and cancer? EMBO Rep **11**, 670–677 (2010)

22. P.A. Gregory, C.P. Bracken, E. Smith, A.G. Bert, J.A. Wright, S. Roslan, M. Morris, L. Wyatt, G. Farshid, Y.-Y. Lim, An autocrine TGF-β/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. Mol Biol Cell **22**, 1686–1698 (2011)

23. T. Acun, E. Oztas, T. Yagci, M.C. Yakicier, SIP1 is downregulated in hepatocellular carcinoma by promoter hypermethylation. BMC Cancer **11**, 223 (2011)

24. E. Oztas, M.E. Avci, A. Ozcan, A.E. Sayan, E. Tulchinsky, T. Yagci, Novel monoclonal antibodies detect Smad-interacting protein 1 (SIP1) in the cytoplasm of human cells from multiple tumor tissue arrays. Exp Mol Pathol **89**, 182–189 (2010)

25. C. Grandori, J. Mac, F. Siëbelt, D.E. Ayer, R.N. Eisenman, Myc-Max heterodimers activate a DEAD box gene and interact with multiple E box-related sites in vivo. EMBO J **15**, 4344–4357 (1996)

26. A. Miquelajauregui, T. Van de Putte, A. Polyakov, A. Nityanandam, S. Boppana, E. Seuntjens, A. Karabinos, Y. Higashi, D. Huylebroeck, V. Tarabykin, Smad-interacting protein-1 (Zfhx1b) acts upstream of Wnt signaling in the mouse hippocampus and controls its formation. Proc Natl Acad Sci U S A **104**, 12919–12924 (2007)

27. A. Stryjewska, R. Dries, T. Pieters, G. Verstappen, A. Conidi, K. Coddens, A. Francis, L. Umans, W.F. van IJcken, G. Berx, Zeb2 regulates cell fate at the exit from epiblast state in mouse embryonic stem cells. Stem Cells **35**, 611–625

(2017)

28. H. Yuzugullu, K. Benhaj, N. Ozturk, S. Senturk, E. Celik, A. Toylu, N. Tasdemir, M. Yilmaz, E. Erdal, K.C. Akcali, Canonical Wnt signaling is antagonized by noncanonical Wnt5a in hepatocellular carcinoma cells. Mol Cancer **8**, 90 (2009)

29. M.F. Carey, C.L. Peterson, S.T. Smale, Chromatin immunoprecipitation (chip). Cold Spring Harb Protoc **2009**, pdb. prot5279 (2009)

30. N. Gévry, S. Hardy, P.-É. Jacques, L. Laflamme, A. Svotelis, F. Robert, L. Gaudreau, Histone H2A. Z is essential for estrogen receptor signaling. Genes Dev **23**, 1522–1533 (2009)

31. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**, 1754–1760 (2009)

32. Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, Model-based analysis of ChIP-Seq (MACS). Genome Biol **9**, R137 (2008)

33. S. Gupta, J.A. Stamatoyannopoulos, T.L. Bailey, W.S. Noble, Quantifying similarity between motifs. Genome Biol **8**, R24 (2007)

34. M.E. Avci, O. Konu, T. Yagci, Quantification of SLIT-ROBO transcripts in hepatocellular carcinoma reveals two groups of genes with coordinate expression. BMC Cancer **8**, 392 (2008)

35. U. Burk, J. Schubert, U. Wellner, O. Schmalhofer, E. Vincan, S. Spaderna, T. Brabletz, A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO Rep **9**, 582–589 (2008)

36. F. Pontén, K. Jirström, M. Uhlen, The human protein atlas—a tool for pathology. J Pathol **216**, 387–393 (2008)

37. J.P. Thiery, Epithelial–mesenchymal transitions in tumour progression. Nat Rev Cancer **2**, 442–454 (2002)

38. A. Sathyanarayanan, K.S. Chandrasekaran, D. Karunagaran, microRNA-145 modulates epithelial-mesenchymal transition and suppresses proliferation, migration and invasion by targeting SIP1 in human cervical cancer cells. Cell Oncol **40**, 119–131 (2017)

39. S. Bugide, V.K. Gonugunta, V. Penugurti, V.L. Malisetty, R.K. Vadlamudi, B. Manavathi, HPIP promotes epithelial-mesenchymal transition and cisplatin resistance in ovarian cancer cells through PI3K/AKT pathway activation. Cell Oncol **40**, 133–144 (2017)

40. A. Fortunato, The role of hERG1 ion channels in epithelial-mesenchymal transition and the capacity of riluzole to reduce cisplatin resistance in colorectal cancer cells. Cell Oncol **40**, 367–378 (2017)

41. G. Xu, L. Zhang, A. Ma, Y. Qian, Q. Ding, Y. Liu, B. Wang, Z. Yang, Y. Liu, SIP1 is a downstream effector of GADD45G in senescence induction and growth inhibition of liver tumor cells. Oncotarget **6**, 33636 (2015)

42. P. Tylzanowski, K. Verschueren, D. Huylebroeck, F.P. Luyten, Smad-interacting protein 1 is a repressor of liver/bone/kidney alkaline phosphatase transcription in bone morphogenetic protein-induced osteogenic differentiation of C2C12 cells. J Biol Chem **276**, 40001–40007 (2001)

43. S.A. Mani, W. Guo, M.-J. Liao, E.N. Eaton, A. Ayyanan, A.Y. Zhou, M. Brooks, F. Reinhard, C.C. Zhang, M. Shipitsin, The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell **133**, 704–715 (2008)

44. N.E. Renthal, C.-C. Chen, C.W. Koriand'r, R.D. Gerard, J. Prange-Kiel, C.R. Mendelson, miR-200 family and targets, ZEB1 and ZEB2, modulate uterine quiescence and contractility during pregnancy and labor. Proc Natl Acad Sci U S A **107**, 20828–20833 (2010)

45. K. Yoshiura, Y. Kanai, A. Ochiai, Y. Shimoyama, T. Sugimura, S. Hirohashi, Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. Proc Natl Acad Sci U S A **92**, 7416–7419 (1995)

46. E.P. Bennett, U. Mandel, H. Clausen, T.A. Gerken, T.A. Fritz, L.A. Tabak, Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology **22**, 736–756 (2011)

47. S.R. Stowell, T. Ju, R.D. Cummings, Protein glycosylation in cancer. Annu Rev Pathol Mech **10**, 473–510 (2015)

48. T. Liu, S. Zhang, J. Chen, K. Jiang, Q. Zhang, K. Guo, Y. Liu, The transcriptional profiling of glycogenes associated with hepatocellular carcinoma metastasis. PLoS One **9**, e107941 (2014)

49. K. Shibao, H. Izumi, Y. Nakayama, R. Ohta, N. Nagata, M. Nomoto, K. Matsuo, Y. Yamada, K. Kitazato, H. Itoh, Expression of UDP-N-acetyl-α-D-galactosamine–polypeptide galNAc N-acetylgalactosaminyl transferase-3 in relation to differentiation and prognosis in patients with colorectal carcinoma. Cancer **94**, 1939–1946 (2002)

50. S. Kitada, S. Yamada, A. Kuma, S. Ouchi, T. Tasaki, A. Nabeshima, H. Noguchi, K. Wang, S. Shimajiri, R. Nakano, Polypeptide N-acetylgalactosaminyl transferase 3 independently predicts high-grade tumours and poor prognosis in patients with renal cell carcinomas. Br J Cancer **109**, 472–481 (2013)

51. Y. Harada, H. Izumi, H. Noguchi, A. Kuma, Y. Kawatsu, T. Kimura, S. Kitada, H. Uramoto, K.-Y. Wang, Y. Sasaguri, Strong expression of polypeptide N. Tumour Biol **37**, 1357–1368 (2016)

52. Z.-Q. Wang, M. Bachvarova, C. Morin, M. Plante, J. Gregoire, M.-C. Renaud, A. Sebastianelli, D. Bachvarov, Role of the polypeptide N-acetylgalactosaminyltransferase 3 in ovarian cancer progression: possible implications in abnormal mucin O-glycosylation. Oncotarget **5**, 544–560 (2014)

53. Y. Fan, R. Bi, M.J. Densmore, T. Sato, T. Kobayashi, Q. Yuan, X. Zhou, R.G. Erben, B. Lanske, Parathyroid hormone 1 receptor is essential to induce FGF23 production and maintain systemic mineral ion homeostasis. FASEB J

30, 428–440 (2016)

54. C.A. Yoshida, T. Kawane, T. Moriishi, A. Purushothaman, T. Miyazaki, H. Komori, M. Mori, X. Qin, A. Hashimoto, K. Sugahara, Overexpression of Galnt3 in chondrocytes resulted in dwarfism due to the increase of mucin-type O-glycans and reduction of glycosaminoglycans. J Biol Chem **289**, 26584–26596 (2014)

55. H. Izumi, O. Ryo, G. Nagatani, I. Tomoko, Y. Nakayama, M. Nomoto, K. Kohno, p300/CBP-associated factor (P/CAF) interacts with nuclear respiratory factor-1 to regulate the UDP-N-acetyl-alpha-d-galactosamine: polypeptide N-acetylgalactosaminyltransferase-3 gene. Biochem J **373**, 713–722 (2003)

56. K.A. Maupin, A. Sinha, E. Eugster, J. Miller, J. Ross, V. Paulino, V.G. Keshamouni, N. Tran, M. Berens, C. Webb, Glycogene expression alterations associated with pancreatic cancer epithelial-mesenchymal transition in complementary model systems. PLoS One **5**, e13002 (2010)