IJC
International Journal of Cancer

Identification of several potential chromatin binding sites of HOXB7 and its downstream target genes in breast cancer

Henna Heinonen[1], Tatiana Lepikhova[1], Biswajyoti Sahu[2], Henna Pehkonen[1], Päivi Pihlajamaa[2], Riku Louhimo[1], Ping Gao[3], Gong-Hong Wei[3], Sampsa Hautaniemi[1], Olli A. Jänne[2] and Outi Monni[1]

[1] Research Programs' Unit, Genome-Scale Biology and Institute of Biomedicine, Medical Biochemistry and Developmental Biology, University of Helsinki, Helsinki, Finland

[2] Institute of Biomedicine/Physiology, University of Helsinki, Helsinki, Finland

[3] Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland

HOXB7 **encodes a transcription factor that is overexpressed in a number of cancers and encompasses many oncogenic functions. Previous results have shown it to promote cell proliferation, angiogenesis, epithelial–mesenchymal transition, DNA repair and cell survival. Because of its role in many cancers and tumorigenic processes, HOXB7 has been suggested to be a potential drug target. However, HOXB7 binding sites on chromatin and its targets are poorly known. The aim of our study was to identify HOXB7 binding sites on breast cancer cell chromatin and to delineate direct target genes located nearby these binding sites. We found 1,504 HOXB7 chromatin binding sites in BT-474 breast cancer cell line that overexpresses *HOXB7*. Seventeen selected binding sites were validated by ChIP-qPCR in several breast cancer cell lines. Furthermore, we analyzed expression of a large number of genes located nearby HOXB7 binding sites and found several new direct targets, such as *CTNND2* and *SCGB1D2*. Identification of HOXB7 chromatin binding sites and target genes is essential to understand better the role of HOXB7 in breast cancer and mechanisms by which it regulates tumorigenic processes.**

Cancer Genetics

HOX genes belong to a family of transcription factors called homeobox genes that contain a 183-base-pair DNA region named homeobox. This region encodes a 61-amino acid protein domain called homeodomain, which is able to bind to DNA. Humans have 39 *HOX* genes organized in four chromosomal clusters, A–D. These clusters of about 100 kb in length are situated on different chromosomes: A at 7p14-15, B at 17q21-22, C at 12q12-13 and D at 2q31-37. Each cluster contains 9–11 *HOX* genes assigned to 13 paralogous groups on the basis of their homeobox sequence similarities and position within each cluster.[1] HOX proteins play key roles in the regulation of embryonal development and maintenance of adult organ homeostasis. *HOX* gene expression is stage-related and tissue- or region-specific, and it follows a unique rule of temporal and spatial colinearity. The genes at the 3′ end of the cluster are expressed first and control the development of anterior body parts, whereas genes at the 5′ end are expressed later during development and control more posterior regions of the developing embryo.[1–7] Because of their important role in normal tissues, altered expression of *HOX* genes has been observed in numerous cancers.[1,8] *HOXB7* is upregulated in melanoma,[9] leukemia,[10] breast,[11] lung,[12] colorectal[13] and pancreatic cancers,[14] and it has a role in promoting cell proliferation, angiogenesis,[9,15–17] epithelial–mesenchymal transition (EMT),[18] DNA repair and cell survival.[19] Additionally, *HOXB7* has been observed to play a dual role in HER2-induced breast cancer tumorigenesis by delaying tumor onset but promoting metastatic tumor progression.[20]

Although all HOX proteins have DNA-binding homeodomains, these domains do not determine binding specificity. It has been observed that HOX proteins share a large homology in amino acid sequence of their homeodomains, and these domains bind to a very similar core sequence (TAAT) *in vitro*. HOX transcription factors require cofactors for the affinity and specificity of binding to the target sites.[21–23] HOX cofactors are proteins that bind to DNA cooperatively with HOX proteins and help in selection of the proper binding site.[24] The best known group of HOX cofactors is TALE (three amino acid loop extension) proteins.[25] These TALE proteins in mammals are encoded by four *Pbx* genes (*Pbx1*,

Key words: HOXB7, ChIP-seq, binding site, target gene, breast cancer
Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Academy of Finland, Finnish Cancer Organizations and Sigrid Jusélius Foundation
DOI: 10.1002/ijc.29616
History: Received 4 Nov 2014; Accepted 11 May 2015; Online 25 May 2015
Correspondence to: Outi Monni, Research Programs' Unit, Genome-Scale Biology, University of Helsinki, PO Box 63 (Haartmaninkatu 8), FIN-00014 Helsinki, Finland, Tel.: +358-40-7639302, Fax: +358-9-19125444, E-mail: outi.monni@helsinki.fi

What's new?
The transcription factor HOXB7 is overexpressed in various cancers, but it's not yet known just which genes HOXB7 activates. How does it influence cancer on a molecular level? This study found 1500 sequences where HOXB7 binds the chromatin in breast cancer cells. They went on to identify several potential target genes near the HOXB7 binding sites. Not only will finding these genes help explain how HOXB7 overexpression promotes tumor growth, it will help understand what side effects might result from hindering HOXB7 expression.

Pbx2, *Pbx3* and *Pbx4*), three Meis genes (*Meis1*, *Meis2* and *Meis3*) and two *Prep* genes (*Prep1* and *Prep2*). There are several Pbx interaction domains in Hox proteins suggesting that Hox–Pbx complexes may assemble onto *cis*-regulatory elements in several different ways depending on the target gene.[24] It has also been observed that the binding sites for Hox–Pbx complex are often used for gene activation, whereas sites with no clear Pbx input are used for both activation and repression. This suggests that Hox–Pbx complexes would recruit transcriptional co-activators rather than co-repressors. Furthermore, it seems that Hox paralogs 1–5 use Pbx cofactors more than paralogs 6–13 and that *cis*-regulatory elements without Pbx input have several Hox binding sites. The same cofactor directs different Hox proteins to different binding sites suggesting that Hox protein contains the specificity information, but it becomes revealed only in the presence of the cofactor. HOXB7 is reported to require PBX1, PBX2 and PREP1 cofactors for its oncogenic activity, and it also regulates their expression.[26] HOX collaborators are proteins that bind in parallel to *cis*-elements of HOX proteins and determine whether the target gene expression is activated or repressed and the strength of this regulation.[24] Multiprotein complexes on *cis*-regulatory elements containing HOX proteins and their cofactors and collaborators are called "Hoxasomes."

We have previously found *HOXB7* to be highly amplified and overexpressed in breast cancer.[11] In our study, our aim was to delineate genome-wide chromatin binding sites of HOXB7 in breast cancer using ChIP-seq and to identify potential target genes that are directly regulated by this transcription factor. This information is important in clarifying the molecular basis and mechanisms of action behind its oncogenic potential as well as in predicting the potential side effects if using HOXB7 as a drug target.

Material and Methods

Human breast cancer cell lines and culture conditions

BT-474 ductal carcinoma, MCF7 adenocarcinoma, MDA-MB-361 adenocarcinoma and T-47D ductal carcinoma cell lines were purchased from ATCC (Manassas, VA). Dulbecco's Modified Eagle's Medium for BT-474 was supplemented with 15% Fetal Bovine Serum (FBS), 1% 100 mM sodium pyruvate, 1% 200 mM L-glutamine and 1% Pen-Strep (penicillin 10,000 U/ml, streptomycin 10,000 μg/ml). Leibovitz's L15 medium for MDA-MB-361 was supplemented with 20% FBS, 1% 200 mM L-glutamine and 1% Pen-Strep. Eagle's Minimum Essential Medium for MCF7 and RPMI-1640 for T-47D were supplemented with 10% FBS, 1% 200 mM L-glutamine and 1% Pen-Strep.

ChIP-seq

Fifty million BT-474 cells were used per sample, and chromatin immunoprecipitation (ChIP) was performed according to Affymetrix® Chromatin Immunoprecipitation Assay Protocol. Fifteen micrograms of HOXB7 antibody (Sigma-Aldrich, St. Louis, MO) and normal rabbit IgG (Alpha Diagnostic International, San Antonio, TX) were used for immunoprecipitation. The ChIP-seq samples were prepared using NEBNext DNA reagents according to the manufacturer's protocol (New England Biolabs, Ipswich, MA), and the DNA library was sequenced using Illumina Genome Analyzer II. Four replicates of IgG control were merged together and used as a reference. From the four sequenced HOXB7 ChIP samples, two and two samples were merged together to yield two biological replicates of HOXB7, which both contained approximately the same amount of reads as the IgG control data set. The ChIP-seq data have been deposited to Gene Expression Omnibus database with accession number: GSE47164.

Data analyses

ChIP-seq peak calling was performed using the MACS algorithm[27] with a p values cut-off of 1e−05. The overlap analysis, motif enrichment analysis and mapping of the binding sites to nearest gene using a ±30 kb window were performed using Cistrome.[28] The pathway analyses were performed with WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt) using KEGG, Wikipathways and Pathway Commons databases. The gene ontology (GO) enrichment analyses were performed with Chipster (http://chipster.csc.fi/).

Correlation analysis

All microarray data were analyzed with the Anduril framework,[29] and all annotations were based on the Ensembl v70 with GRCh37. We downloaded gene expression microarrays from The Cancer Genome Atlas for 522 primary breast carcinoma tumors and 59 controls.[30] First, probes matching either multiple or no genes were removed. Then, the data were normalized to a mean of 0. Each gene was assigned one of three categories: upregulated, downregulated or normal. A gene was considered upregulated if its expression was more than three standard deviations over the median of normal samples. Twelve samples displayed upregulation of *HOXB7*. To determine which of the genes that were located nearby the newly found HOXB7 binding sites had an association with *HOXB7* expression status, samples were divided into two groups (*HOXB7* upregulated and *HOXB7* normal), and a signal-to-

noise statistic was calculated on the expression of each gene.[31] A *p*-value for each signal-to-noise score was calculated by a permutation test (1,000 permutations) and *p*-values were corrected for multiple hypotheses using the Benjamini–Hochberg procedure.

ChIP-qPCR

Five million cells were plated onto 150-mm dishes and cultured for 4 days. Antibody conjugation and ChIP were performed as described previously.[32] Ten micrograms of HOXB7 antibody and normal rabbit IgG were used. The sites for validation were selected based on the following criteria: (*i*) strong enrichment by ChIP-seq, (*ii*) location in proximal promoter or (*iii*) location close to genes associated with epithelial cancers. The primers for the binding sites were designed using Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and the sequences are listed in Supporting Information Table 1. qPCR was carried out using LightCycler480 and SYBR Green I detection system (Roche, Mannheim, Germany) according to the manufacturer's instructions. The enrichment of ChIP relative to input chromatin was calculated using % input method and the formula: $100*2^{(\text{Adjusted input} - \text{Cp IP})}$. Statistical significance of enrichment for HOXB7 *vs*. IgG was calculated using Student's *t*-test in SPSS or T.TEST in Excel. If the *p*-value was <0.05, the result was regarded as statistically significant.

Transductions

Five shRNA constructs targeting *HOXB7* mRNA from the MISSION® RNAi library (Sigma-Aldrich) as well as controls, including non-target shRNA and an empty pLKO-1 vector were used. Lentiviral particles were produced at Biomedicum Functional Genomics Unit (University of Helsinki, Finland). One day before transductions, 100,000 BT-474 cells per well were plated onto a 12-well plate. Next day, the medium was changed, and 1 μl of polybrene and 500 μl of virus suspension were added to 500 μl of medium. The plate was incubated at 37°C for 10 min and then centrifuged at 2,500 rpm for 30 min. After centrifugation, the plate was incubated at 37°C for 6 hr after which the medium was changed. The transduced cells were selected by puromycin treatment for a week.

Immunoblotting

The cells were lysed in cold RIPA buffer (Sigma-Aldrich) supplemented with phosphatase and protease inhibitors (Thermo Fisher Scientific, Waltham, MA and Roche, Basel, Switzerland, respectively) and 1 mM Ethylenediaminetetraacetic acid. HOXB7 antigen was detected with HOXB7 primary antibody (diluted in 1:250) and anti-rabbit secondary antibody (1:20,000) both from Invitrogen (Camarillo, CA). The primary antibodies for CCND1 (Abcam, Cambridge, UK), EGFR, phospho-EGFR, AKT, phospho-AKT, ERK, phospho-ERK, p38 and phospho p38 (Cell Signaling, Boston, MA) were used at 1:1,000 dilution and incubated overnight in 1% milk/TBS-Tween. Five to ten micrograms of protein lysates were loaded on SDS-gels (BioRad, Hercules, CA) and after electrophoresis, blotted to Polyvinylidene difluoride membranes with TurboTransferBlot equipment (BioRad, Hercules). Membranes were blocked with 5% milk Tris Buffered Saline with Tween for 1 hr and incubated with primary antibody dilution (1% milk/Tris Buffered Saline with Tween (TBST), 1:1,000) overnight. Membranes were washed with TBST and incubated with secondary antibody Horse radish Peroxidase-goat-anti-rabbit or mouse IgG (H + L; Invitrogen) in 1% milk/TBST (1:20,000). Membranes were washed with TBST. Equal protein loading was verified with α-tubulin primary antibody (1:2,000, Sigma-Aldrich) and anti-mouse secondary antibody (1:20,000, Invitrogen). Immobilon Western reagents (Millipore, Billerica, MA) were used for chemiluminescence. Knockdown of HOXB7 was quantified with ImageJ software, and alpha-tubulin was used as a loading control.

RNA isolation and qRT-PCR

Total RNA was isolated using NucleoSpin RNA II kit (Macherey-Nagel, Düren, Germany) and transcribed to cDNA using random hexamers supplied with SuperScript™ III First-Strand Synthesis System for RT-PCR (Invitrogen). qPCR was carried out using LightCycler480 and SYBR Green I detection system (Roche). Primer sequences of the genes located nearby HOXB7 binding sites are shown in Supporting Information Table 2. Each sample was measured in duplicate, and the gene expression results were normalized using *ACTB* mRNA levels as reference. The results were confirmed using also *RPLPO* mRNA levels as a reference (data not shown).

Results

Validation of HOXB7 antibody function

Two known genes, *FGF2*[9,15] and *EGFR*,[33] have previously been reported to possess HOXB7 binding sites nearby and to be directly regulated by this transcription factor. We tested first whether HOXB7 antibody used for ChIP can recognize these sites using following primers (5′–3′) for *FGF2*: forward GTCATCTGTGGCACCTGCT and reverse AGATGAGGAT GGGGATAAGGA and for *EGFR*: forward CAAGGCCAG CCTCTGAT and reverse CCCCTTTCCCTTCTTTTGTT. Strong enrichment was observed at both binding sites in samples immunoprecipitated with HOXB7 antibody as opposed to those immunoprecipitated with non-specific IgG, thus validating the functionality of the selected HOXB7 antibody (Fig. 1).

ChIP-seq analysis of HOXB7 binding sites

We identified 1,504 new HOXB7 binding sites (false discovery rate, FDR $<5\%$) on BT-474 breast cancer cell chromatin. Figure 2*a* contains examples of the binding sites representing sites nearby *C10orf10*, *ADORA2B* and *ABHD11* genes. Thirty most enriched binding sites with an annotated nearest gene are listed in Table 1, and the list encompassing all HOXB7 binding sites is shown in Supporting Information Table 3. The majority of the binding sites are located at distal intergenic regions (52%) and introns (39%), and only 4% are situated at promoter regions (Fig. 2*b*).

Motif and *cis*-element analyses of the HOXB7 binding sites

We performed *de novo* motif search to define a DNA sequence that is enriched among the HOXB7 binding sites on BT-474 cell chromatin. Figure 2*c* shows the most common motif that was present in 99.4% (1,496/1,504, *z* scores: −19.6898, $p = 1e-30$) of the binding sites. We also defined the top-scoring *cis*-elements for the HOXB7 binding sites. The top 15 *cis*-elements are shown in Supporting Information Table 4. In addition, HOXB7 binding sites had hits to known motifs of several other HOX transcription factors (Supporting Information Table 5). For example, Hoxd1 and HOXA3 motifs hit 97% of the found HOXB7 binding sites.

ChIP-qPCR validation of selected HOXB7 binding sites

Seventeen HOXB7 binding sites identified by ChIP-seq were validated with ChIP-qPCR using a ChIP protocol different from that for ChIP-seq. We first validated the binding sites in two breast cancer cell lines, BT-474 and MDA-MB-361, showing high-level expression of *HOXB7* (Fig. 3 and Supporting Information Fig. 1). BT-474 shows high-level amplification and expression of *HOXB7*.[11] In MDA-MB-361 cell line, *HOXB7* is highly expressed but not amplified. To further validate the binding sites, we included T-47D with intermediate *HOXB7* expression as well as MCF7 with moderate expression of *HOXB7* (Supporting Information Figs. 2 and 3). Enrichment was seen in HOXB7 antibody *versus* IgG immunoprecipitated samples at all the selected sites and in the vast majority of the cases the result was statistically significant ($p < 0.05$). Therefore, we conclude that the 1,504 sites identified by ChIP-seq represent true chromatin binding sites of HOXB7 in breast cancer cells.

Correlation analysis in clinical breast cancer data

To see whether the genes located nearby HOXB7 binding sites correlate with *HOXB7* expression in clinical breast cancer data, we compared the expression of these genes with *HOXB7* in 12 *HOXB7*-overexpressing breast tumors to 510 normal tumors in the breast cancer dataset of The Cancer Genome Atlas. We found 37 positively and 27 negatively correlating genes ($q < 0.05$) with *HOXB7* expression (Supporting Information Fig. 4 and Supporting Information Table 6), suggesting genes potentially activated and repressed by HOXB7, respectively.

Expression of genes located nearby HOXB7 binding sites

BT-474 breast cancer cells were transduced with five different lentiviral shRNA constructs targeting *HOXB7* mRNA, but only one construct resulted in a sufficient knockdown of *HOXB7*. The cells transduced with empty vector and non-targeting shRNA construct were used as controls. Both HOXB7 protein (Fig. 4*a*) and mRNA levels (Fig. 4*b*) were decreased in cells transduced with *HOXB7* shRNA construct as compared to the controls. HOXB6 shares ~40% homology with HOXB7 and is expressed in BT-474 cell line. To test the specificity of the *HOXB7* shRNA construct, we determined its effect also on *HOXB6* expression, and no change in *HOXB6* mRNA level was observed (Fig. 4*c*).

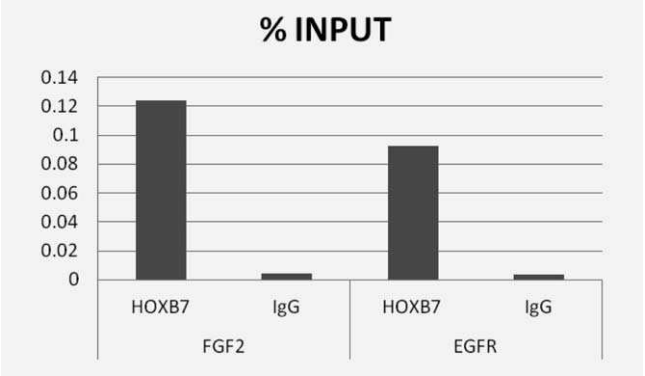

Figure 1. ChIP-qPCR validation of HOXB7 antibody in BT-474 breast cancer cell line with two known binding sites located near *FGF2* and *EGFR* genes. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

In order to examine whether the HOXB7 binding sites were functional, the expression of nearby genes was analyzed in HOXB7 depleted *vs.* control cells by measuring the levels of encoded mRNAs with qRT-PCR. Altogether, 45 genes nearby 31 binding sites were selected for further analysis. These genes were chosen either because they were the nearest genes of the binding sites or they were located nearby the binding site and had been reported to be associated with breast or some other epithelial tumor type. Eight genes, *SCGB1D1*, *SCGB1D2*, *SCGB1D4*, *SCGB2A1*, *SCGB2A2*, *DAP*, *DHRS2* and *ADORA2B*, were found to be highly upregulated (Fig. 5), whereas *CCNB1*, *DFFA*, *NRL*, *TARDBP*, *PBX1* and *MAL2* showed moderate upregulation in *HOXB7* knockdown cells (Supporting Information Fig. 5). In addition, *TRPC6* and *ING3* were observed to be moderately upregulated and correlated negatively with *HOXB7* expression also in the clinical breast cancer dataset of 12 *HOXB7*-overexpressing samples (The Cancer Genome Atlas). By contrast, *APMAP*, *CTNND2*, *DCAF11*, *EEF2K*, *PI3KR1* and *WNT16* were moderately downregulated in *HOXB7*-depleted cells (Fig. 5 and Supporting Information Fig. 5). *WNT16* correlated positively with *HOXB7* also in the clinical breast cancer data set ($p = 0.025$, $q = 0.154$, Supporting Information Table 6).

Enriched pathways and GO categories among genes located nearby the new HOXB7 binding sites

Pathways and GO categories enriched among genes located nearby the HOXB7 binding sites are shown in Supporting Information Tables 7 and 8, respectively. Enriched pathways included, for example, pathways associated with FGF, EGFR as well as MAPK signaling. Furthermore, enriched GO classes were involved in the regulation of cell proliferation, transcription, apoptosis and angiogenesis, supporting the previous findings on HOXB7 function.

Effect of HOXB7 on regulation of signaling pathways

HOXB7 has been linked to enhanced cellular growth and proliferation, and GO analysis showed enrichment of cellular

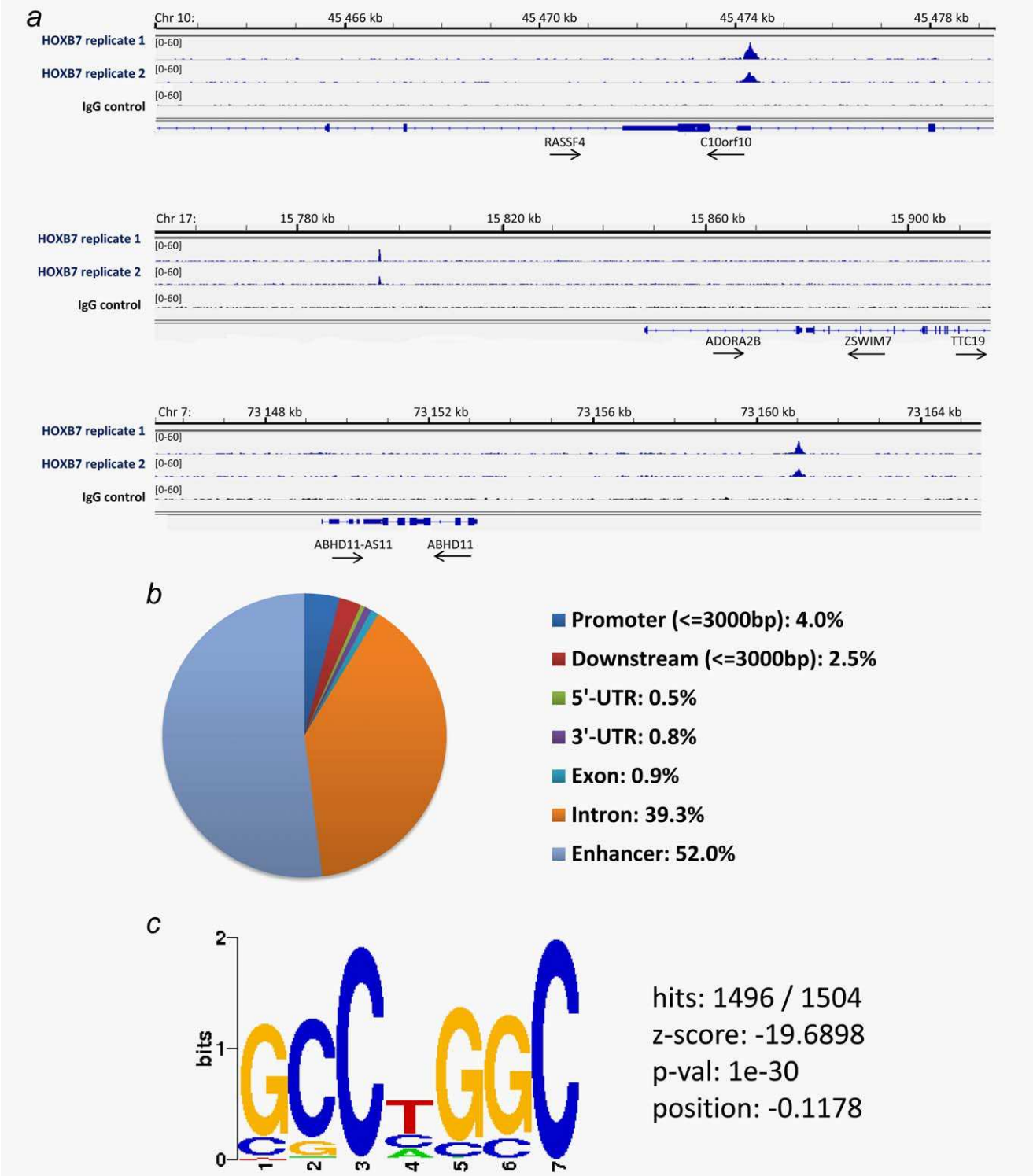

Figure 2. (*a*) Examples of the HOXB7 binding sites which are located close to annotated genes *C10orf10*, *ADORA2B* and *ABHD11*. (*b*) Distribution of the HOXB7 binding sites between different genomic regions in BT-474 breast cancer cell line. (*c*) *De novo* motif search in BT-474 breast cancer cell line. The most enriched motif among the HOXB7 binding sites found with ChIP-seq.

processes and pathways associated with these activities. Therefore, it was imperative to study whether HOXB7 is involved in regulating the activity of EGFR as well as downstream PI3K/AKT and MAPK signaling pathways, which are the two critical pathways promoting cell growth and survival in cancer. Interestingly, phospho-ERK protein levels

Table 1. Thirty most enriched new HOXB7 binding sites for which an annotated nearest gene was found

Chromosome	Start	End	Fold enrichment	Location	Nearest gene	Description
chr10	45474137	45474515	66,71	overlapStart	*C10orf10*	Chromosome 10 open reading frame 10
chr17	15795989	15796350	50,83	Upstream	*ADORA2B*	Adenosine A2b receptor
chr5	151743562	151743627	49,79	Downstream	*NMUR2*	Neuromedin U receptor 2
chr7	73160787	73161225	49,66	Upstream	*ABHD11*	Abhydrolase domain containing 11
chr5	67845023	67845460	48,51	Downstream	*PIK3R1*	Phosphoinositide-3-kinase, regulatory subunit 1 (alpha)
chr16	30382250	30382612	46,4	Upstream	*TBC1D10B*	TBC1 domain family, member 10B
chr11	93398671	93398756	45,81	Inside	*KIAA1731*	KIAA1731
chr6	4739139	4739232	45,11	Inside	*CDYL*	Chromodomain protein, Y-like
chr17	41499069	41499197	43,82	Downstream	*ARL4D*	ADP-ribosylation factor-like 4D
chr19	16895537	16895599	43,82	Upstream	*SIN3B*	SIN3 transcription regulator homolog B (yeast)
chr4	122694216	122694334	43,82	Upstream	*TMEM155*	Transmembrane protein 155
chr12	8179103	8179513	43,2	Upstream	*FOXJ2*	Forkhead box J2
chrX	48065011	48065133	42,51	Upstream	*SSX5*	Synovial sarcoma, X breakpoint 5
chr15	76317972	76318114	41,83	Upstream	*NRG4*	Neuregulin 4
chr3	52754097	52754183	41,83	Upstream	*GLT8D1*	Glycosyltransferase 8 domain containing 1
chr4	118875392	118875509	41,83	Upstream	*NDST3*	*N*-deacetylase/*N*-sulfotransferase (heparan glucosaminyl) 3
chr5	67797322	67797381	41,83	Downstream	*PIK3R1*	Phosphoinositide-3-kinase, regulatory subunit 1 (alpha)
chr5	85485023	85485099	41,83	Upstream	*NBPF22P*	Neuroblastoma breakpoint family, member 22, pseudogene
chrX	54447865	54448010	41,83	Upstream	*TSR2*	TSR2, 20S rRNA accumulation, homolog (*S. cerevisiae*)
chrX	119035398	119035530	41,83	Inside	*AKAP14*	A kinase (PRKA) anchor protein 14
chr19	3723507	3723664	40,39	Inside	*TJP3*	Tight junction protein 3
chr10	104133091	104133237	39,83	Upstream	*NFKB2*	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)
chr12	133489094	133489155	39,83	Inside	*PGAM5*	Phosphoglycerate mutase family member 5
chr17	16806704	16806826	39,83	Downstream	*TBC1D27*	TBC1 domain family, member 27
chr6	41330382	41330498	39,83	Downstream	*NCR2*	Natural cytotoxicity triggering receptor 2
chr2	202144389	202144585	39,72	Inside	*CASP8*	Caspase 8, apoptosis-related cysteine peptidase
chr19	4778968	4779146	39,6	Downstream	*MIR7-3*	MicroRNA 7-3
chr19	5430778	5430841	38,4	Upstream	*ZNRF4*	Zinc and ring finger 4
chr11	43335391	43335498	37,84	Inside	*API5*	Apoptosis inhibitor 5
chr2	107095275	107095369	37,84	Upstream	*RGPD3*	RANBP2-like and GRIP domain containing 3

The list of all 1,504 new binding sites of HOXB7 in BT-474 breast cancer cell line identified by ChIP-seq is shown in Supporting Information Table 3.

decreased upon HOXB7 knockdown, whereas AKT, p38 MAPK and EGFR signaling were activated when HOXB7 was knocked down (Supporting Information Fig. 6).

Discussion

As *HOX* genes play key roles in regulating the homeostasis of normal tissues and organs, deregulation of their expression often results in serious consequences such as promotion of oncogenic processes.[1] Although the identification of target genes of HOX transcription factors is very essential, they still remain largely uncovered. This is due to the fact that DNA binding of different HOX proteins is highly similar *in vitro* and they require a large amount of cofactors and collaborators for binding specificity and direction of regulation *in vivo*.[24] This makes identification of HOX targets very challenging. HOX proteins often share *cis*-elements with other transcription factors and, therefore, it is unlikely that there are targets that are specific for HOXB7 only. Based on

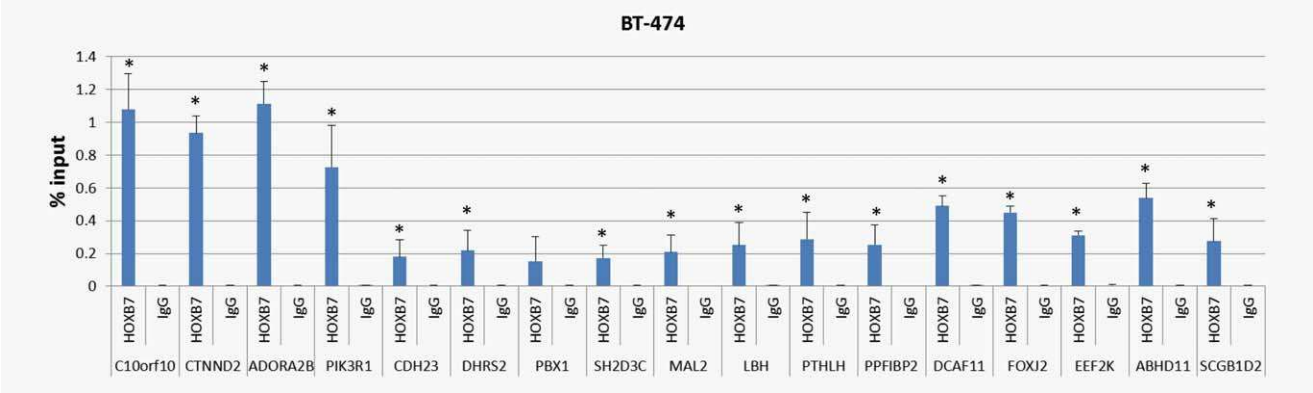

Figure 3. ChIP-qPCR validation of 17 HOXB7 binding sites on BT-474 breast cancer cell chromatin. The binding sites are named by the gene nearest to the binding site. Results are averages of two biological replicates, which were done in duplicates. Statistical significance is marked with an asterisk, $p < 0.05$. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

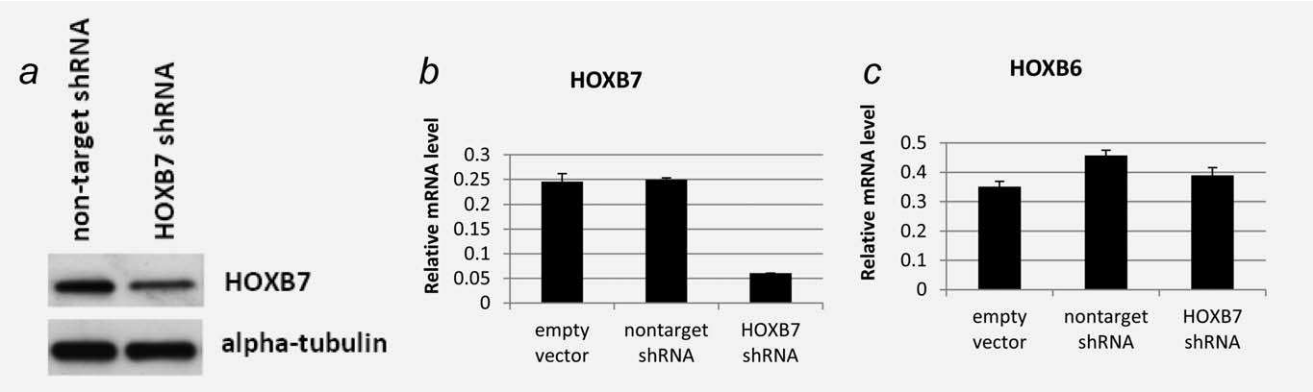

Figure 4. Expression of *HOXB7* in knockdown *vs*. control cells at (*a*) protein and at (*b*) mRNA levels in BT-474 breast cancer cell line. (*c*) *HOXB6* mRNA expression was not affected by *HOXB7* shRNA. HOXB7 knockdown showed 45% of HOXB7 protein expression as compared to 100% in non-target cells in (*a*).

studies with *Drosophila*, Mann *et al.*[24] suggest that there are three types of Hox target genes: (*i*) paralog-specific or genes that are highly specific for one Hox paralog, (*ii*) semi-paralog-specific or genes that are shared by a subset of Hox proteins and (*iii*) general target genes of Hox proteins, which are regulated by most or all Hox genes. Moreover, the binding sites of HOX transcription factors do not have to be paralog-specific as long as the other genes using the same sites are not expressed at the same time or in the same tissues.

HOXB7 is overexpressed in many different cancers and is associated with several oncogenic processes.[12–18] We decided to carry out pathway analysis to see whether our results are in line with this data. For this analysis, we included all the genes that were located close to HOXB7 binding sites. Our assumption is that not all of these genes are directly regulated by HOXB7, but the analysis is likely to highlight biological processes that are important for HOXB7 regulation. We identified several oncogenic pathways, such as EGFR, MAPK and PI3K signaling, that were enriched among HOXB7 targets. Therefore, we wanted to test whether HOXB7 manipulation has any impact on the activation of these pathways. HOXB7 knockdown had moderate effect on AKT and p38 MAPK kinase signaling. On the other hand, HOXB7 knockdown caused clear decrease in levels of ERK phosphorylation as shown previously in colorectal and breast cancer.[13,18] Surprisingly, knockdown of HOXB7 showed increase in phosphorylated EGFR (Y1068) which is interesting, because Y1068 is one of the phosphorylation sites for ERK activation. Our data proposes that HOXB7 promotes activation of ERK, but it could occur either through other tyrosine kinases than EGFR or activation of threonine phosphorylation of EGFR.[34,35] Alternatively, ERK activation could result in feedback loop leading to inhibition of Y1068 of EGFR as well as AKT signaling in BT-474.

HOXB7 has been suggested to be a potential drug target due to its role in promoting cell proliferation and survival.[12,14,33] However, very little is known about the HOXB7 binding sites on chromatin. In our study, we have identified 1,504 HOXB7 binding sites on BT-474 breast cancer cell chromatin using ChIP-seq technique and validated these results with ChIP-qPCR in four different breast cancer cell lines. The vast majority of the binding sites were located at distal intergenic regions and introns, suggesting that HOXB7

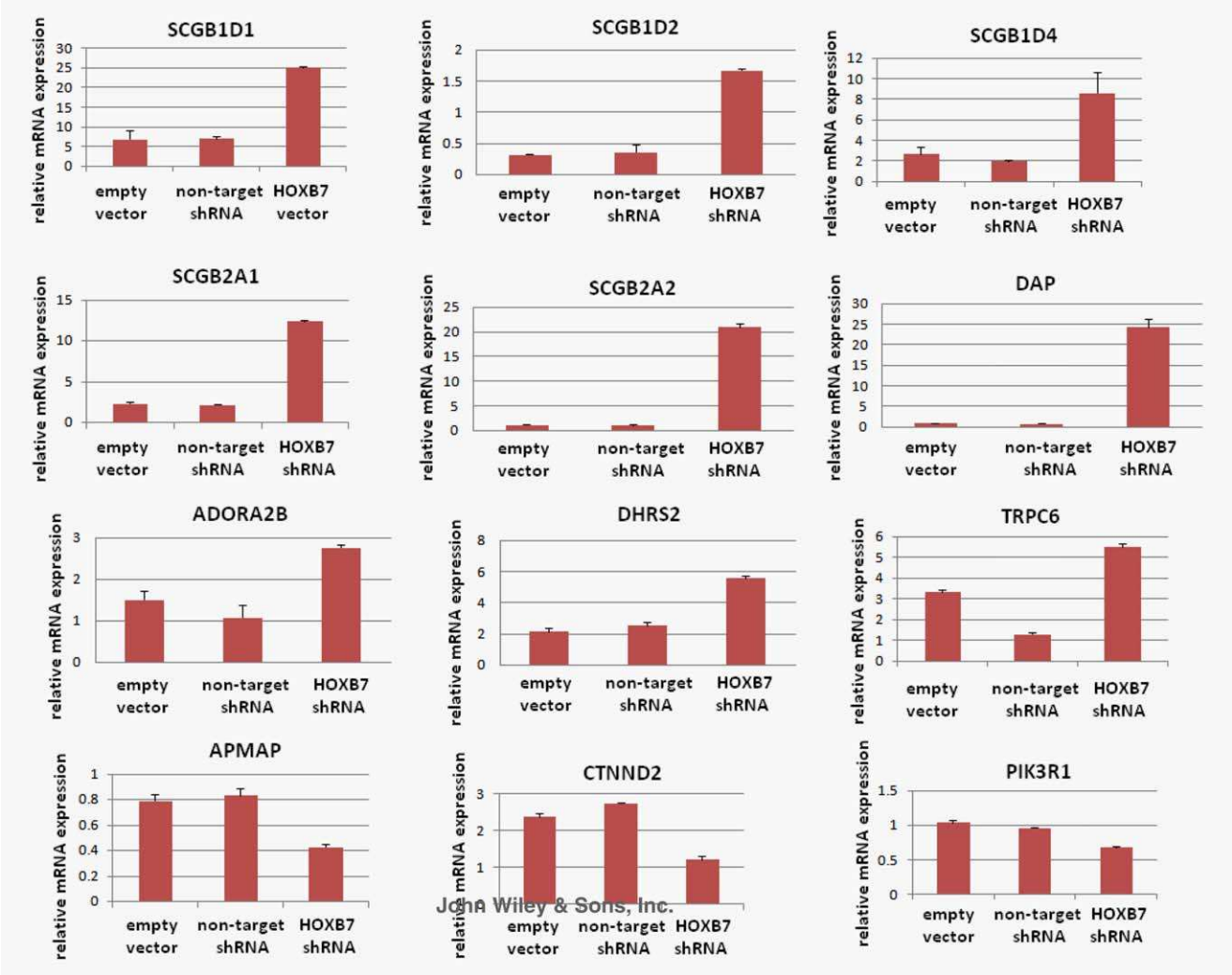

Figure 5. Deregulated expression of 12 genes located nearby HOXB7 binding sites in *HOXB7* knockdown cells *vs*. empty vector and non-target controls analyzed with qRT-PCR in BT-474 breast cancer cell line. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

mainly binds to distal enhancer regions. HOXB7 might regulate the expression of its target genes by interacting with the basic transcriptional machinery through DNA looping rather than binding to the promoter of its target gene itself. Similar distribution of binding sites was seen with Hoxa2 in mouse embryos.[36] We studied expression of the genes located nearby HOXB7 binding sites in a *HOXB7* knockdown cell line and clinical breast cancer samples. Correlation analysis for *HOXB7* and its potential target genes in publicly available clinical breast cancer data revealed 64 genes of which 37 correlated positively and 27 correlated negatively with *HOXB7* expression, suggesting direct target genes potentially activated and repressed by HOXB7, respectively. Many of the targets are suggested to have a role in breast cancer pathogenesis. To name few, driver mutations were recently found on *ARID1B* in breast cancer.[37] *LBH* is a novel transcription co-factor and a direct target of the WNT/β-catenin signaling pathway. It was recently shown to be involved in the regulation of mammary gland development and basal mammary stem cell lineage.[38] *FOXJ2*, which negatively correlated with *HOXB7* expression, was shown to inhibit migratory properties of human breast cancer.[39]

Interestingly, five genes from the secretoglobin gene family, *SCGB1D1*, *SCGB1D2*, *SCGB1D4*, *SCGB2A1* and *SCGB2A2*, became highly upregulated upon HOXB7 knockdown. These genes are located around a HOXB7 binding site at 11q13 and seem to be repressed by HOXB7. SCGB1D1 and SCGB1D2 are also known as lipophilin A and lipophilin B, respectively. These lipophilins form heterodimers with SCGB2A proteins, for example, SCGB1D2 heterodimerizes with SCGB2A2,[40] and the heterodimers further associate to form tetramers. SCGB1D4 has been reported to regulate chemotactic migration and invasion,[41] but the functions of other members of the SCGB1D subfamily are still largely unknown. SCGB2A1 and SCGB2A2 are also known as mammaglobins. *SCGB2A1* expression has been observed to be associated with

favorable outcome in ovarian cancer.[42] Its role in breast cancer prognosis is still unknown, but in light of the published data, it is interesting that *SCGB2A1* potentially promoting favorable outcome is suppressed by HOXB7. Two genes around a binding site at 5p15, *CTNND2* and *DAP*, showed also a change in their expression upon HOXB7 depletion. *CTNND2* has been reported to be overexpressed in prostate cancer and involved in tumor formation,[43] making it an intriguing target activated by HOXB7. Furthermore, suppression of *DAP* by HOXB7 is interesting due to its function as a potential tumor-suppressor gene.[44] HOXB7 has been reported to downregulate its cofactor PBX1,[26] which was confirmed also in our study, because *PBX1* became moderately upregulated after *HOXB7* suppression. Altogether, we were able to find an expression change in 49% of the genes tested (22/45). This may be due to the fact that HOXB7 requires cofactors and collaborators for target gene regulation. Therefore, suppressing only HOXB7 expression may not be sufficient to result in detectable expression changes in the majority of its target genes. Furthermore, *HOXB7* is part of the *HOXB* cluster located at 17q21-22, the site that is amplified in 10% of primary breast cancers and in several breast cancer cell lines,[11] so when HOXB7 is overexpressed due to this amplification, as is the case in BT-474 cell line, several other *HOXB* genes are also amplified and at least *HOXB2* is also overexpressed. It is not clear whether HOXB7 shares targets with other *HOXB* genes, whereas functional redundancy has been suggested for *HOX* genes belonging to the same paralogous group.[45] In this light, HOXB7 and HOXA7 could share target genes if they are expressed in the same cell.

Although the significance of *HOXB7* overexpression has been acknowledged in a number of cancers, the molecular mechanisms behind its oncogenic potential are still largely unknown. Our study reports >1,500 chromatin binding sites in the genome of a breast cancer cell line overexpressing *HOXB7*. The genes located nearby these binding sites are potential direct targets of HOXB7. Only a handful of direct targets of HOXB7 have been published so far and therefore, our study provides valuable data for further investigation. In the future, functional studies of these genes are required to reveal specific mechanisms through which HOXB7 executes its oncogenic potential.

Acknowledgements

Janica Djupsjöbacka and Matti Kankainen at the Biomedicum Functional Genomics Unit are acknowledged for excellent technical assistance. Olli Kallioniemi and Sami Kilpinen are acknowledged for the collaboration and fruitful discussions in the beginning of the project. Mikko Kivento is thanked for the help in statistical analyses. The results published here are in part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at http://cancergenome.nih.gov. The TSP study accession number in the database of Genotype and Phenotype (dbGaP) for the TCGA study used here is phs000569.v1.p7.

References

1. Grier DG, Thompson A, Kwasniewska A, et al. The pathophysiology of *HOX* genes and their role in cancer. *J Pathol* 2005;205:154–71.
2. Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature* 1978;276:565–70.
3. Graham A, Papalopulu N, Krumlauf R. The murine and Drosophila homeobox gene complexes have common features of organization and expression. *Cell* 1989;57:367–78.
4. Krumlauf R. *Hox* genes in vertebrate development. *Cell* 1994;78:191–201.
5. Abate-Shen C. Deregulated homeobox gene expression in cancer. *Nat Rev Cancer* 2002;2:777–85.
6. Takahashi Y, Hamada JI, Murakawa K, et al. Expression profiles of 39 *HOX* genes in normal human adult organs and anaplastic thyroid cancer cell lines by quantitative real-time RT-PCR system. *Exp Cell Res* 2004;293:144–53.
7. Challen G, Gardiner B, Caruana G, et al. Temporal and spatial transcriptional programs in murine kidney development. *Physiol Genomics* 2005;23:159–71.
8. Shah N, Sukumar S. The Hox genes and their roles in oncogenesis. *Nat Rev Cancer* 2010;10:361–71.
9. Carè A, Silvani A, Meccia E, et al. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. *Mol Cell Biol* 1996;16:4842–51.
10. Storti P, Donofrio G, Colla S, et al. HOXB7 expression by myeloma cells regulates their pro-angiogenic properties in multiple myeloma patients. *Leukemia* 2010;3:527–37.
11. Hyman E, Kauraniemi P, Hautaniemi S, et al. Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res* 2002;62:6240–5.
12. Yuan W, Zhang X, Xu Y, et al. Role of HOXB7 in regulation of progression and metastasis of human lung adenocarcinoma. *Mol Carcinog* 2014;53:49–57.
13. Liao WT, Jiang D, Yuan J, et al. HOXB7 as a prognostic factor and mediator of colorectal cancer progression. *Clin Cancer Res* 2011;17:3569–78.
14. Nguyen Kovochich A, Arensman M, Lay AR, et al. HOXB7 promotes invasion and predicts survival in pancreatic adenocarcinoma. *Cancer* 2013;119:529–39.
15. Carè A, Silvani A, Meccia E, et al. Transduction of the SkBr3 breast carcinoma cell line with the HOXB7 gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. *Oncogene* 1998;16:3285–9.
16. Carè A, Valtieri M, Mattia G, et al. Enforced expression of HOXB7 promotes hematopoietic stem cell proliferation and myeloid-restricted progenitor differentiation. *Oncogene* 1999;18:1993–2001.
17. Carè A, Felicetti F, Meccia E, et al. HOXB7: a key factor for tumor-associated angiogenic switch. *Cancer Res* 2001;61:6532–9.
18. Wu X, Chen H, Parker B, et al. HOXB7, a homeodomain protein, is overexpressed in breast cancer and confers epithelial-mesenchymal transition. *Cancer Res* 2006;66:9527–34.
19. Rubin E, Wu X, Zhu T, et al. A role for the HOXB7 homeodomain protein in DNA repair. *Cancer Res* 2007;67:1527–35.
20. Chen H, Lee JS, Liang X, et al. Hoxb7 inhibits transgenic HER2/neu-induced mouse mammary tumor onset but promotes progression and lung metastasis. *Cancer Res* 2008; 68:3637–44.
21. Gehring WJ, Qian YQ, Billeter M, et al. Homeodomain-DNA recognition. *Cell* 1994;78:211–23.
22. Chang CP, Brocchieri L, Shen WF, et al. Pbx modulation of Hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. *Mol Cell Biol* 1996;16:1734–45.
23. Mann RS, Affolter M. Hox proteins meet more partners. *Curr Opin Genet Dev* 1998;8:423–9.
24. Mann RS, Lelli KM, Joshi R. Hox specificity: unique roles for cofactors and collaborators. *Curr Top Dev Biol* 2009;88:63–101.
25. Moens CB, Selleri L. Hox cofactors in vertebrate development. *Dev Biol* 2006;291:193–206.
26. Fernandez LC, Errico MC, Bottero L, et al. Oncogenic HoxB7 requires TALE cofactors and is inactivated by a dominant-negative Pbx1 mutant in a cell-specific manner. *Cancer Lett* 2008;266:144–55.
27. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;9:R1137.
28. Liu T, Ortiz JA, Taing L, et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* 2011;12:R83.
29. Ovaska K, Laakso M, Haapa-Paananen S, et al. Large-scale data integration framework provides a comprehensive view on glioblastoma multiforme. *Genome Med* 2010;2:65.
30. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490:61–70.

31. Hautaniemi S, Ringnér M, Kauraniemi P, et al. A strategy for identifying putative causes of gene expression variation in human cancers. *J Franklin Inst* 2004;341:77–88.
32. Sahu B, Laakso M, Ovaska K, et al. Dual role of FoxA1 in androgen signaling and prostate cancer. *EMBO J* 2011;30:3962–76.
33. Jin K, Kong X, Shah T, et al. The HOXB7 protein renders breast cancer cells resistant to tamoxifen through activation of the EGFR pathway. *Proc Natl Acad Sci USA* 2012;109:2736–41.
34. Gan Y, Shi C, Inge L, et al. Differential roles of ERK and Akt pathways in regulation of EGFR-mediated signaling and motility in prostate cancer cells. *Oncogene* 2010;29:4947–58.
35. Li X, Huang Y, Jiang J, et al. ERK-dependent threonine phosphorylation of EGF receptor modulates receptor downregulation and signaling. *Cell Signal* 2008;20:2145–55.
36. Donaldson IJ, Amin S, Hensman JJ, et al. Genome-wide occupancy links Hoxa2 to Wnt-β-catenin signalling in mouse embryonic development. *Nucleic Acids Res* 2012;40:3990–4001.
37. Stephen MJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* 2012;486:400–4.
38. Lindley LE, Curtis KM, Sanchez-Mejias A, et al. The WNT-controlled transcriptional regulator LBH is required for mammary stem cell expansion and maintenance of the basal lineage. *Development* 2015;142:893–904.
39. Wang Y, Yang S, Ni Q, et al. Overexpression of forkhead box J2 can decrease the migration of breast cancer cells. *J Cell Biochem* 2012;113:2729–37.
40. Jackson BC, Thompson DC, Wright MW, et al. Update of the human secretoglobin (*SCGB*) gene superfamily and an example of 'evolutionary bloom' of androgen-binding protein genes within the mouse *Scgb* gene superfamily. *Hum Genomics* 2011;5:691–702.
41. Choi MS, Ray R, Zhang Z, et al. IFN-gamma stimulates the expression of a novel secretoglobin that regulates chemotactic cell migration and invasion. *J Immunol* 2004;172:4245–52.
42. Tassi RA, Calza S, Ravaggi A, et al. Mammaglobin B is an independent prognostic marker in epithelial ovarian cancer and its expression is associated with reduced risk of disease recurrence. *BMC Cancer* 2009;9:253.
43. Lu Q, Dobbs LJ, Gregory CW, et al. Increased expression of delta-catenin/neural plakophilin-related armadillo protein is associated with the down-regulation and redistribution of E-cadherin and p120ctn in human prostate cancer. *Hum Pathol* 2005;36:1037–48.
44. Levy-Strumpf N, Kimchi A. Death associated proteins (DAPs): from gene identification to the analysis of their apoptotic and tumor suppressive functions. *Oncogene* 1998;17:3331–40.
45. Maconochie M, Nonchev S, Morrison A, et al. Paralogous HOX genes: function and regulation. *Annu Rev Genet* 1996;30:529–56.