486–501 *Nucleic Acids Research, 2011, Vol. 39, No. 2* *Published online 17 September 2010*
doi:10.1093/nar/gkq819

A highly conserved SOX6 double binding site mediates SOX6 gene downregulation in erythroid cells

Claudio Cantu'[1,*], Vito Grande[1], Ilaria Alborelli[1], Letizia Cassinelli[1], Ileana Cantu'[1], Maria Teresa Colzani[1], Rossella Ierardi[2], Luisa Ronzoni[3], Maria Domenica Cappellini[3], Giuliana Ferrari[2,4], Sergio Ottolenghi[1] and Antonella Ronchi[1,*]

[1]Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, [2]San Raffaele Telethon Institute for Gene Therapy (HSR-TIGET), San Raffaele Scientific Institute, 20132 Milan, [3]Dipartimento di Medicina Interna, Università di Milano, Fondazione Policlinico Mangiagalli, Regina Elena, IRCCS, Milano and [4]Vita-Salute San Raffaele University, Milan, Italy

Received December 23, 2009; Revised August 27, 2010; Accepted August 31, 2010

ABSTRACT

The Sox6 transcription factor plays critical roles in various cell types, including erythroid cells. Sox6-deficient mice are anemic due to impaired red cell maturation and show inappropriate globin gene expression in definitive erythrocytes. To identify new Sox6 target genes in erythroid cells, we used the known repressive double Sox6 consensus within the εy-globin promoter to perform a bioinformatic genome-wide search for similar, evolutionarily conserved motifs located within genes whose expression changes during erythropoiesis. We found a highly conserved Sox6 consensus within the Sox6 human gene promoter itself. This sequence is bound by Sox6 *in vitro* and *in vivo,* and mediates transcriptional repression in transient transfections in human erythroleukemic K562 cells and in primary erythroblasts. The binding of a lentiviral transduced Sox6FLAG protein to the endogenous Sox6 promoter is accompanied, in erythroid cells, by strong downregulation of the endogenous Sox6 transcript and by decreased *in vivo* chromatin accessibility of this region to the PstI restriction enzyme. These observations suggest that the negative Sox6 autoregulation, mediated by the double Sox6 binding site within its own promoter, may be relevant to control the Sox6 transcriptional downregulation that we observe in human erythroid cultures and in mouse bone marrow cells in late erythroid maturation.

INTRODUCTION

Sox6 is a member of the Sox (Sry-type HMG box) family of transcription factors, characterized by the presence of an HMG domain that recognizes the minor groove on DNA. The binding of Sox proteins to DNA forces it to bend at about 75°, introducing local conformational changes. The ability of Sox proteins to bind in close proximity to other transcription factors and to distort DNA suggests that they can act as 'architectural proteins', possibly by promoting the assembly of biologically active multiprotein complexes. These complexes, in turn, mediate the interactions between distant chromatin domains, bringing together promoter/enhancer regions, finally assembling the 'chromatin hubs' that control gene expression regulation. Within this general frame, Sox6 has been suggested to act both as activator or repressor, depending on its interactions and on its target sequences (1–3). The Sox domain recognizes a very degenerate (A/T)(A/T)CAA(A/T)G consensus, making it very difficult to identify its *in vivo* targets: the best-characterized and validated Sox6 target sequence on the regulatory elements of the chondrocyte gene Col2a1 is, for example, composed of four sites each having different mismatches relative to the HMG box consensus (4). Moreover, the presence of adjacent pairs of Sox sites on different known targets suggests that double Sox sites might likely be the preferred Sox6 targets, although the relative arrangement and orientation of the two sites is not yet clearly defined (1–3 and references therein).

Murine Sox6 null mutants (p100H) show delayed growth, myopathy, arterioventricular heart block and die within 2 weeks following birth (5). Sox6 is indeed

*To whom correspondence should be addressed. Tel: +39 02 6448 3337; Fax: +39 02 6448 3314; Email: antonella.ronchi@unimib.it
Correspondence may also be addressed to Claudio Cantu'. Email: mcclaudio061281@gmail.com

required for proper formation of heart, nervous system (5,6,7), cartilage (4,8) and cardiac and skeletal muscle (9,10). Recent reports indicate that complete Sox6 ablation causes a perturbation of erythropoiesis resulting in the presence of increased numbers of nucleated and misshaped red cells in the fetal circulation and in a strong relative increase of embryonic (εy) globin gene expression (11–13). In particular, Sox6 directly silences εy-globin expression in murine definitive erythropoiesis by binding to a double Sox6 site lying within a 36-bp region on the εy proximal promoter (13). Moreover, embryonic liver stem cells from Sox6 null mice engrafted into lethally irradiated wild-type (WT) adult mice, show levels of εy expression in the spleen and bone marrow that are higher than those observed in control mice transplanted with wild type cells (12). Finally, Sox6 cooperates with BCL11a to downregulate the γ-globin gene in adult erythroid cells (14).

The emerging critical role of Sox6 in erythropoiesis prompted us to search for new direct targets by combining a bioinformatic approach with DNA microarray analysis. Among the resulting candidate Sox6 binding sites, we found an evolutionarily conserved double Sox6 site lying within the Sox6 proximal promoter itself, and we demonstrated that Sox6 binds to this region repressing its transcriptional activity in K562 cells. Accordingly, overexpression of an exogenous Sox6 by lentiviral transduction in both K562 cells and primary human erythroblast represses the endogenous Sox6 transcript expression. To link this observation to the Sox6 expression profile during human erythropoiesis, we used as model system primary $CD34^+$ cells (from both Cord Blood and Peripheral Blood) undergoing *in vitro* erythroid differentiation. In both cultures, Sox6 expression peaks in early erythroblasts and decreases along with erythroid differentiation. Together, these data suggest that the Sox6 autoregulation mediated by the double Sox6 binding site within the promoter may be relevant for Sox6 downregulation in the final stages of erythroid differentiation.

MATERIALS AND METHODS

Cell cultures

$CD34^+$ cells were immunopurified from Human Cord Blood (HCB), cultured for 2 weeks in StemSpan (Stem Cell Technologies) containing 20% of fetal bovine serum (FBS, Hyclone) and supplemented with 10 ng/ml of human stem cell factor (hSCF), 1 U/ml of human erythropoietin (EPO), 1 ng/ml of human interleukin-3 (hIL-3), 10^{-6} M dexamethasone (Sigma) and 10^{-6} M β-estradiol (Sigma) (15). $CD34^+$ cells were seeded at a concentration of 10^5 cells/ml and diluted over time to maintain the cell concentration in the range of $1–2 \times 10^6$ cells/ml. Cells were collected and analysed on days 6, 8, 10, 12 and 14.

$CD34^+$ cells from human peripheral adult blood (HPB) were immunopurified after buffy coat isolation and erythroid cultures were done as detailed in ref. (16): cells were cultured at a density of 10^5 cells/ml in alpha-minimal essential medium in the presence of SCF, IL-3 and Epo. Cyclosporine A at 1 μg/ml was added to inhibit lymphocyte and monocyte growth. After 7 days of culture, cells were replated in erythroid differentiation medium. Cell samples were collected and analysed on days 7 and 14. For both cultures, erythroid differentiation was monitored by FACS analysis for the expression of GlycophorinA (CD235), and by cell-morphology analysis on cytocentrifuged samples stained with May–Grunwald–Giemsa. Primary human cells were obtained according to Institutional guidelines. K562 cells were cultured in RPMI medium supplemented by 10% FBS, PenStrep and L-glutamine.

Electrophoretic mobility shift assay (EMSA)

DNA oligonucleotide probes were labeled at the same specific activity: 3 pmol of each double-stranded oligonucleotide were labeled with γ-^{32}P ATP by using T4 Polynucleotide Kinase (NEB) and purified on 8% acrylamide gel. Labeled probe at 30 fmol was incubated with 10 μg of nuclear or total extracts, for 20 min at 15°C in a buffer containing 5% glycerol, 50 mM NaCl, 20 mM Tris, pH 7.9, 0.5 mM EDTA, 5 mM MgCl, 1 mM dithiothreitol (DTT), 500 ng/μl poly(dG-dC) and 50 ng/μl bovine serum albumin (BSA) in a 15-μl final reaction mixture. The reaction mixture was then loaded onto a 8% polyacrylamide gel (29:1 acrylamide-bisacrylamide ratio) and run at 4°C and 150V for 3 h. Nuclear extracts were prepared according to standard protocols (17,18). The antibodies used for the supershift analysis were: anti-FLAG, (Sigma F7425); anti-GATA1 (SantaCruz N6, sc-265).

Competitions were done using a 25-, 50- and 100-fold molar excess (with respect to the radiolabeled probe) of 'cold' oligonucleotides.

Oligonucleotide probes:

WT probe: Fw: 5′-CCTCTGTAACAAAGTTTCTTTGTTTTAATGG-3′,
Rev: 5′-CCATTAAAACAAAGAAACTTTGTTACAGAGG-3′;
Mut probe: Fw: 5′-CCTCTGTGGCAGAGTGTCTGTGTGTGAATGG-3,
Rev: 5′-CCATTCACACACAGACACTCTGCCACAGAGG-3′;
Mut1 probe: Fw: 5′-CCTCTGTGGCAGAGTGTCTTTGTTTTAATGG-3′,
Rev: 5′-CCATTAAAACAAAGACACTCTGCCACAGAGG-3′;
Mut2 probe: Fw: 5′-CCTCTGTAACAAAGTGTCTGTGTGTGAATGG-3′,
Rev 5′-CCATTCACACACAGACACTTTGTTACAGAGG-3′.

Sox6 overexpression vectors

The Sox6 murine cDNA (19) was kindly provided by Prof. Michiko Hamada-Kanazawa, Kobe-Gakuin University, Japan. The Sox6 cDNA was transferred into the pCMV-Tag 4B plasmid (Stratagene), in frame with a 3′ FLAG epitope, (EcoRI-EcoRV restriction sites), to

produce the Sox6FLAG expression vector used in transfection assays. The Sox6 recombinant protein lacks the 49 C-terminal aminoacids: this shorter molecule fully retains Sox6 biological proprieties (19) and allows discrimination between endogenous and exogenous Sox6. The Sox6FLAG cassette (EcoRI-KpnI) was blunted and cloned immediately upstream to the IRES-Emerald GFP cassette into the blunted BamHI site of the pHR SIN BX IR/EMW [derived from pHR SIN CSGW, (20)] lentiviral vector. The two packaging plasmids psPAX2 and pMD-VSVG were used to produce Lentiviral pseudo-particles in 293T cells (www.lentiweb.com).

The expression levels of exogenous Sox6FLAG versus the endogenous transcript was estimated in K562 cells and in cultures of primary Cord Blood-derived cells (at Day 10, peak of Sox6 expression) by Real Time PCR, using GAPDH mRNA as a standard for comparison (Supplementary Figure S1).

Lentiviral harvesting protocol

Exponentially growing HEK 293T cells were transfected using Lipofectamine 2000 (Invitrogen) with the three vectors lentiviral system. Nearly 48 h after transfection, the supernatant containing recombinant viruses was harvested, filtered (0.45 μm), centrifuged at 20.000*g* for 8 h and then frozen at −80°C. Lentiviruses were titrated on HEK 293T cells by measuring the percentage of GFP positive cells by FACS analysis.

Luciferase reporter plasmids

The human Sox6 promoter region from nt −1116 to −1 was obtained by direct amplification from genomic DNA with Phusion High-fidelity DNA Polymerase (Finnzymes) using the following primers:

Fw: 5′-ATCGGTACCGGGCTGAGTTAGATATTTATTTC-3′;
Rev: 5′-ATCTCGAGAGATCTGAATTCATGAAAGTGACCTG-3′

containing a KpnI and XhoI restriction site (underlined), respectively, for further cloning into the corresponding sites of the pGL2 luciferase reporter vector (Promega). To generate the corresponding plasmid mutated in the Sox6 consensus (−1116Mut) a two-step PCR approach was used: (i) in the first series of PCRs, mutations were introduced by amplifying the wild-type template sequence combining the above Fw and Rev primers with the following oligonucleotides, carrying the desired mutations:

RevMut 5′-GGCAGAGTGTCTGTGTGTGAATGGAACTAAAAATATGCTG-3′
FwMut 5′-CACACACAGACACTCTGCCACAGAGGCTGTATTCTTTC-3′.

(ii) The two amplified fragments were gel purified, annealed and used as a template for a second round of PCR with the external Fw and Rev primers. The resulting mutated fragment was cloned (KpnI-XhoI sites) into pGL3.

The DNA region of 234 nt (from nt −991 to −759) containing the double Sox6 site, was amplified with the following primers:

Fw: 5′-ACGTGGTACCGATCCATTGTTTTTCAGAAGG-3′
Rev: 5′-ACGTCTCGAGAACAAAGAAACTTTGTTACAGAGGC-3′

containing the KpnI and XhoI sites, as above, for cloning into the pGL2 luciferase reporter vector (Promega), upstream to the minimal Gata-1 promoter and to the Gata-1 erythroid-specific enhancer region HS2 (16). To mutate the double Sox consensus, the same region was amplified with a Reverse primer containing the same mutations proved to abolish Sox6 binding in EMSA assay:

Rev: 5′-ACGTCTCGAGCACACAGACACTCTGCCACAGAGGC-3′.

All the amplified DNA regions were sequenced to avoid undesired mutations.

Transfection experiments

1×10^5 Exponentially growing K562 cells, seeded in 24-well dishes, were transfected in 0.5 ml of Opti-MEM medium (Invitrogen), using 2 μl of Lipofectamine 2000 (Invitrogen), 800 ng of the reporter plasmid and increasing amounts (indicated in the figures) of the Sox6 expression plasmid (pCMV-Sox6Tag4B) per well. The pCMV-Tag4B empty vector was added to each transfection at the concentration required to equalize the total amount of DNA transfected in each reaction. After 24 h, total cellular extracts were prepared and Luciferase activity was measured according to the Promega Luciferase reporter system protocol. P19 cells were transfected in the same condition in a 12-well dish. Retinoic acid (RA) at 500 nM was used to induce Sox6 expression (19). All experiments were repeated in quadruplicate with at least two independent plasmid preparations.

Transfection of primary erythroblasts was obtained by nucleofection (Human $CD34^+$ cell Nucleofector Kit, Amaxa VPA-1003) of 5 μg of the Luciferase reporter constructs together with 50 ng of Renilla reporter plasmid to normalize for transfection efficiency. Each transfection was carried out in triplicate and two independent cultures were used from both peripheral and cord blood.

Chromatin immunoprecipitation assay

Briefly, K562, FACS-sorted mouse bone marrow cells or human primary erythroblasts ($1–2 \times 10^6$ cells for each immunoprecipitation reaction) were fixed with 0.4% formaldehyde for 10 min at room temperature, and chromatin was sonicated to a size of ∼500 nt. Immunoprecipitation was performed after overnight incubation and subsequent incubation with protein A agarose (Upstate biotechnology). The following antibodies were used: anti-FLAG antibody (Sigma F-7425), anti-Sox6 antibody (Millipore Ab5805), anti-H3K4me1 (Abcam 8895-100) and anti-H3K27me3 (Upstate 07-449),

anti-HDAC1 (Millipore 17-608). Immunoprecipitated DNA was then analysed by amplifying an equivalent of 10^4 cells DNA with the following oligonucleotides:

Sox6 promoter: Fw 5′-TTTGAAAGAATACAGCCT CTG-3′,
Rev 5′-ATGCATTAAGGTGGTTTGGTA-3′;
GAPDH locus: Fw 5′-CGGAGTCAACGGATTTGGTC GTAT-3′,
Rev 5′-AGCCTTCTCCATGGTGGTGAAGAC-3′;
Sox4 locus: Fw 5′-CATGGTGCAGCAAACCAACA-3′, Rev 5′-TTCATGGGTCGCTTGATGTG-3′;
Cyclin D1 promoter: Fw 5′-CTCCCGCTCCCATTCTC T-3′, Rev 5′-GAGGCTCCAGGACTTTGC-3′.
Mouse Sox6 promoter: Fw5′-GCATTAGAAAGTAGTT AGGTCG, Rev5′-CAGCCTCTGTAACAAAGTTCC;
Mouse Sox6 exon16: Fw5′-TGCGACAGTTCTTCACT GTGG, Rev 5′-CGTCCATCTTCATACCATACG.

Immunoprecipitation was repeated three times on independent chromatin preparations.

RNA isolation and RT–PCR

Total RNA from 10^5 cells (both K562 or differentiating human $CD34^+$ cells at the different time points) was purified with TRI Reagent (Applied Biosystem AM9738). Before cDNA synthesis, RNA was treated with RQ1 DNase (Promega) for 30 min at 37°C. cDNA was prepared using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystem cat n°4368814) and then diluted 1:50 for amplification. Negative control reactions (without reverse transcriptase) gave no signal. Real-time analysis was performed using ABI Prism 7500, (PE Applied Biosystems).

Primers were designed to amplify 100 to 150-bp amplicons and were based on sequences from the Ensembl database (http://www.ensembl.org/Homo_sapiens/Info/Index). Samples from three or more independent experiments were analyzed in triplicate. Specific PCR product accumulation was monitored by SYBR Green dye fluorescence in 25-µl reaction volume. Dissociation curves confirmed the homogeneity of PCR products.

The same forward primer was used to amplify both endogenous and exogenous Sox6 cDNA:

Fw: 5′-GAGGCAGTTCTTTACTGTGG-3′. To discriminate between endogenous and transduced RNA variants, two different reverse primers were used: Rev1: 5′-CCGCCATCTGTCTTCATAC-3′ complementary to the extreme 3′ of Sox6 transcript and Rev2. 5′-CTTATC GTCGTCATCCTTGTA-3′, which matches with the FLAG coding region (the combination of primers (F, R1 and R2) is schematized in Figure 6A). Primers for GAPDH were: Fw: 5′-ACGGATTTGGTCGTATTG GG-3′, Rev: 5′-TGATTTTGGAGGGATCTCGC-3′; Mouse Sox6 primers: Fw: 5′-TGCGACAGTTCTTCAC TGTGG-3′, Rev: 5′-CGTCCATCTTCATACCAT ACG-3′; Mouse HPRT primers: Fw: 5′-CCTGCTGGAT TACATTAAAGCACTG-3′, Rev: 5′-GTCAAGGGCAT ATCCAACAACAAAC-3′.

Western blot

K562 total and nuclear extracts were prepared according to standard protocol (17,18) and proteins were subjected to SDS–PAGE separation and blotting. The endogenous Sox6 protein was detected by the anti Sox6 (c-20) Santa Cruz sc-17332 antibody raised against the 20aa C-terminal epitope (absent in the exogenous Sox6FLAG protein). The Sox6FLAG and Sox4FLAG proteins were detected by the use of the anti-FLAG antibody Sigma F7425. Protein loading was checked by reprobing filters with a monoclonal anti hnRNP-C1/C2 antibody (Sigma R5028). Antibodies binding was detected by using appropriate horseradish peroxidise-conjugated IgG and revealed by ECL (LiteAblot, Euroclone).

Cell sorting

Freshly extracted mouse bone marrow cells were disaggregated in phosphate-buffered saline, and incubated with the following labeled antibodies: allophycocyanin (APC) anti-mouse CD117 (c-Kit); PE anti-mouse CD71, and FITC anti-mouse TER119, all from Becton Dickinson, RD. The sorting was performed by a MoFlo (DAKO-Cytomation) cell sorter and purity of the obtained cells populations was >95%.

Restriction enzyme accessibility assay

PstI chromatin accessibility assay was performed as previously described (21). Briefly, nuclei were collected from 2×10^6 Sox6- or empty vector-transduced K562 cells 72 h post infection, using RSB buffer (10 mM Tris–HCl, pH 7.5, 10 mM NaCl and 3 mM MgCl2) containing 0.2%(v/v) NP40 and 10 mM DTT. Nuclei were washed in RSB buffer without NP40 and resuspended in NEB Buffer 3 (100 mM NaCl, 50 mM Tris–HCl, 10 mM MgCl2, 1 mM dithiothreitol, pH 7.9) and aliquots were incubated at 37°C for 15 min with or without 30 units of PstI (NEB). Digestions were stopped adding the same volume of STOP Buffer (0.6 M NaCl, 20 mM Tris–HCl pH 8, 10 mM EDTA, 1% SDS). Samples were then treated overnight with proteinase K and DNA was recovered after phenol/chloroform extraction and subsequent ethanol precipitation.

Digested DNA was quantified by Real Time PCR (ABI Prism 7500, PE Applied Biosystem Real Time PCR system) as described in (22). A PstI non-containing amplicon (within 18S ribosomal subunit gene) was used as internal control to normalize for DNA loading.

Primers:

GATA Promoter Fw: 5′-CTCCCGCCAGCACTGGCC TG-3′
GATA Promoter Rw: 5′-GCCGGTGTGGGCTAGACT CC-3′
Sox6 Promoter Fw: 5′-CAAGATCCATTGTTTTTCAG AAGG-3′
Sox6 Promoter Rw: 5′-AACAAAGAAACTTTGTTACA GAGGC-3′
18S Fw: 5′-TTTCGGAACTGAGGCCATGATTAAG-3′
18S Rw: 5′-AGTTTCAGCTTTGCAACCATACTCC-3′

RESULTS

Sox6 expression during erythroid maturation

To get insight into the regulation of the Sox6 gene during erythropoiesis, we analyzed the Sox6 expression profile in different cellular models of erythroid differentiation.

To this end, we used primary human erythroid liquid cultures from $CD34^+$ cells purified either from adult peripheral [HPB; (16)], or umbilical cord blood [HCB; (15)], widely considered a physiological model of human erythropoietic differentiation. In these cultures, purified $CD34^+$ cells are amplified to erythroblasts and then induced to terminal differentiation in a semi-synchronous way into mature erythrocytes. In HCB-derived cultures, Sox6 expression was analyzed by RT–PCR at Day 0 (purified $CD34^+$ cells), during erythroblast amplification (Day6), at the end of the erythroid precursors amplification stage (Day 8) and then at days 10, 12 and 14 during erythroid induced differentiation. As shown in Figure 1A, Sox6 expression is absent in $CD34^+$ cells, starts to be detectable at Day 8 (72% of glycophorinA positive (GpA^+) cells, as estimated by FACS analysis, not shown), reaches a peak around day 12 (85% of GpA^+ cells), to finally decrease at Day 14, corresponding to the end of the culture and to the maximum extent of erythroid differentiation (~90% of GpA^+ cells).

Sox6 expression shows a similar pattern in the two-phase erythroid culture from $CD34^+$ cells purified from adult HPB (Figure 1B): Sox6 mRNA is absent in purified $CD34^+$ cells (Day 0), accumulates at the end of the erythroblasts amplification stage (Day 7, corresponding to 67% of GpA^+ cells) and then decreases at Day 14 of the culture, when erythroid differentiation is completed (97% of GpA^+ cells).

We then analyzed Sox6 expression *in vivo*, in mouse bone marrow FACS-sorted cell populations representing progressive stages of erythroid maturation. As shown in Figure 1D, Sox6 is modestly expressed in the more immature (Kit^+CD71^-) cell population, increases together with erythroid differentiation ($CD71^{high}Ter119^{low}$ and $CD71^{high}Ter119^{high}$ cells) and then falls in the most mature cells ($CD71^{low}TER119^{high}$).

These latter data confirm, in an *in vivo* system, the observation, made in the *ex vivo* primary cultures that, during mouse and man erythoid differentiation, Sox6 expression peaks during early erythroid differentiation, but then decreases at later stages.

Finally, we examined Sox6 expression in a widely used cell line, the erythroleukemic K562 cells. These cells express low levels of Sox6 that remain unchanged upon addition of hemin, a drug that induces a substantial degree of differentiation in these cells (Figure 1C).

Identification of Sox6 targets on DNA

Sox proteins share a conserved DNA-binding HMG box domain, which dictates their preference for the recognition of the hexameric core sequence 5′-WWCAAW-3′ (W = A/T) (3).

The only Sox6 target defined so far in erythropoietic cells is the double Sox6 binding site within the mouse εy-globin proximal promoter. This sequence, consisting of two AACAA(A/T)G sites in opposite orientation, spaced by 8 nt, is thought to concur to the εy-globin gene repression in adult erythroid cells (13) (Figure 2A).

To identify new Sox6 target genes in erythroid cells, we performed an *in-silico* genome-wide search using the web free tool TFBScluster (http://hscl.cimr.cam.ac.uk/TFBScluster), which allows identification of patterns of binding sites present in evolutionarily conserved regulatory regions (23).

By taking as a model the double Sox6 binding site from the *εy*-globin promoter, we set TFBScluster in order to search for two 5′-WWCAAW-3′ sites within a cluster of 30 nt. This search found 875 consensus sequences in mouse/man conserved regions, interspersed in the human genome (Figure 2B). Only a minority of them (6%) were found near the transcription start site of known genes, while the others were located in introns (56%), or downstream to the 3′-end of known genes (38%). In principle, every single identified target could be relevant, but to narrow the search to regions of high regulatory potential and to the erythroid system, we matched the 56 consensi found within 10-kb upstream to known genes with a list of differentially expressed genes (DEG) at different stages of erythroid maturation in mouse definitive erythropoiesis. This DEGs list (C.C., A.R. unpublished data) results from a DNA microarray analysis on three populations of hematopoietic cells purified by FACS sorting from E13.5 fetal livers, according to their relative expression of c-Kit (that marks early progenitor cells), and Ter119 (a protein associated to GlycophorinA, expressed only by erythroid committed cells). These populations were: $cKit^+/TER119^-$ cells, representing pluripotent hematopoietic progenitors; $cKit^+/TER119^+$ cells, corresponding to erythroid committed early progenitors and $cKit^-/TER119^{++}$ cells consisting of more differentiated erythroblasts and mature erythrocytes.

Among the seven genes identified by the merging of these two lists (see Figure 2D), there is Sox6 itself, suggesting a potential direct transcriptional auto-regulation. The position and the evolutionary conservation of this double Sox6 consensus was also confirmed by the use of two additional bioinformatics tools: the USCS database (http://genome.ucsc.edu/) and the TESS Transcription Element Search System software (http://www.cbil.upenn.edu/cgi-bin/tess/tess).

The Sox6 promoter contains a highly conserved double Sox6 consensus

The evolutionarily conserved putative double Sox6 binding site maps within a mouse/man homology region, extending from nt +1 [transcriptional start site, mapped in (24)] to −1116, in position −775/−759; Figure 3A and B. This site lies in a block of 31 conserved nucleotides, and is composed of two single sites in opposite orientation, spaced by three bases. The two Sox6 consensus binding sites are fully conserved from chicken to man whereas few substitutions are present within the spacing nucleotides (Figure 3C).

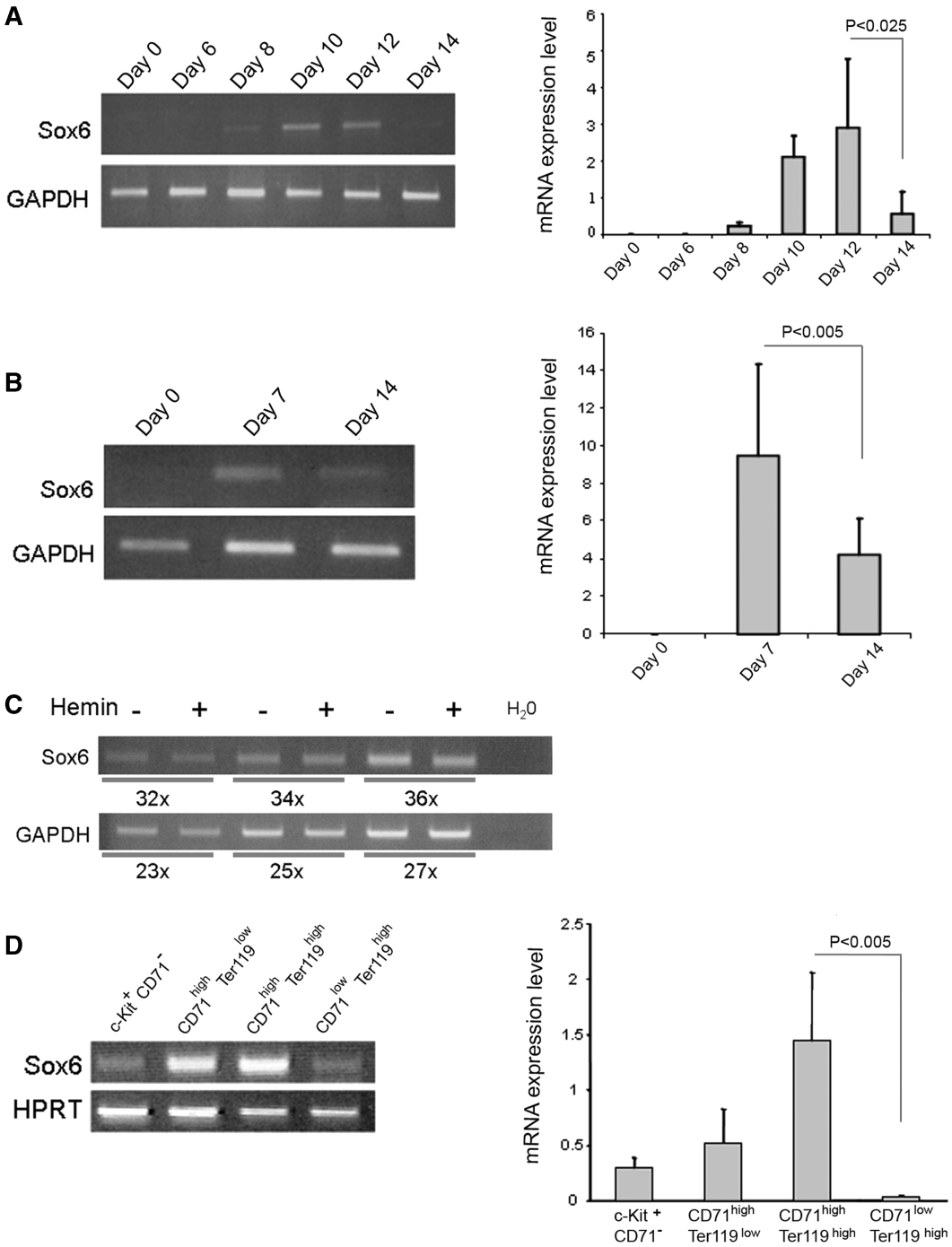

Figure 1. Sox 6 expression during erythroid maturation. (**A**) Erythroid liquid cultures from $CD34^+$ cells purified from HCB. Day 0: purified $CD34^+$ cells; Day 8: beginning of the erythroid differentiation phase. Day 14: end of the culture, when 90% of cells are GpA^+. (**B**) Erythroid liquid cultures from $CD34^+$ cells purified from HPB. Day 0: purified $CD34^+$ cells; Day 7: beginning of erythroid differentiation; Day 14: end of erythroid differentiation (97% GpA^+cells) (**C**) Semi-quantitative RT–PCR on cDNA from K562 cells, treated (+) or not (−) with hemin (50 µM for 4 days). PCR amplification cycles are indicated below the lanes. (**D**) Mouse Bone Marrow cells sorted according to their erythroid maturation, from more immature kit^+CD71^- to progressively more mature stages $CD71^{high}Ter119^{low}$, $CD71^{high}Ter119^{high}$, $CD71^{low}$ $Ter119^{high}$. Semi-quantitative (left) and RealTime (right) PCRs were performed on cDNA from cells indicated in the figures. The Sox6 expression level is normalized on GAPDH (human cells) and HPRT (mouse cells), respectively.

To assess the ability of the mouse/man conserved region (+1/−1116) to promote transcription in an erythroid context, we cloned it immediately upstream to the Luciferase reporter gene. This region (1116-S6), when compared with the promoterless pGL3 background plasmid and with the Thymidine kinase promoter (pTK), shows an intermediate transcriptional activity. This suggests that, although additional enhancer sequences

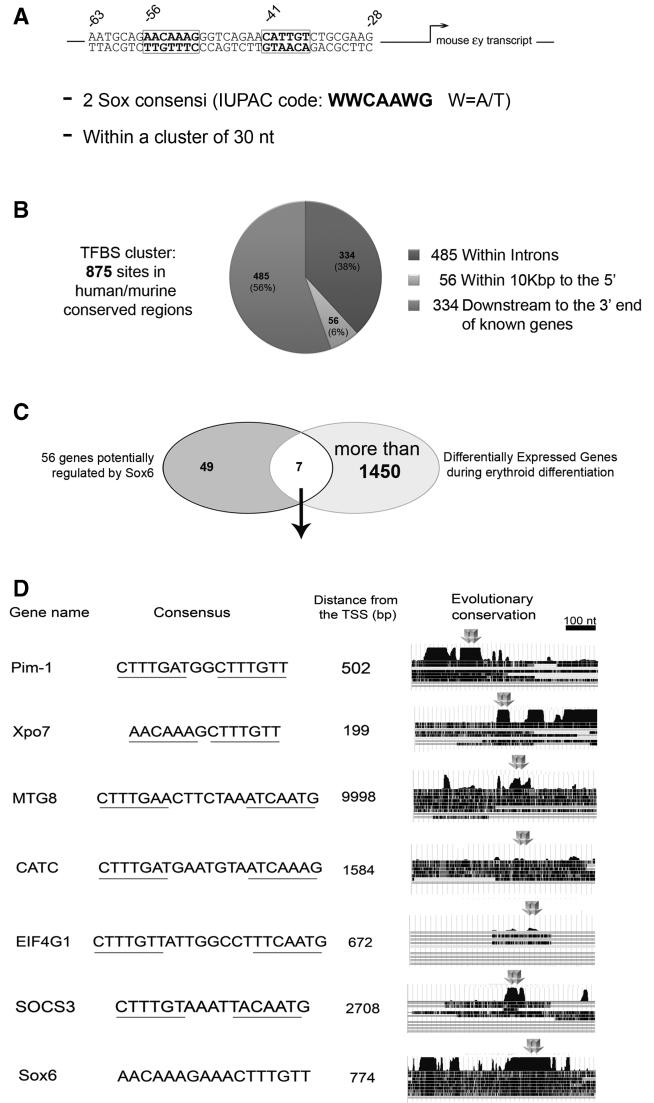

Figure 2. Schematic representation of the bioinformatic strategy used to identify new direct Sox6 targets. (**A**) Sequence of the double Sox6 binding site located within the mouse εy-globin promoter. Nucleotides positions are indicated on top of the sequence. Search criteria set in TFBScluster search are indicated. (**B**) TFBScluster search output. The identified potential Sox6 binding sites are clustered on the basis of their relative position with respect to known genes. The absolute number—and the corresponding percentages—of sites for each group is reported in the pie-chart. (**C**) The merge of the list of the 56 potential Sox6 binding sites located within 10Kb upstream to known genes with the list of genes differentially expressed during mouse erythropoiesis identified seven erythroid genes potentially regulated by Sox6 (listed in **D**). For each potential target site (including the one within the Sox6 locus itself) the distance from the transcriptional start site, the exact Sox consensus sequence, and the evolutionary conservation are indicated.

may be required for Sox6 optimal activity in K562, this element behaves as a promoter in an erythroid context (Figure 3D).

Sox6 binds to the putative double consensus identified within the Sox6 promoter

To test the ability of Sox6 to bind to the above identified consensus, we set up EMSA experiments using as a probe either an oligonucleotide encompassing the double Sox6 consensus site (WT) or the corresponding oligonucleotides mutated in only one (Mut1 and Mut2) or in both Sox6 binding sites (MUT; Figure 4A). As shown in Figure 4B the WT probe, when incubated with K562 nuclear extracts, gives a weak but consistent retarded band (lane 1), which becomes barely detectable when a similar amount of total protein extracts is used (lane 7). To overcome the low level of SOX6 in K562, we moved to protein extracts from K562 cells transfected with a pCMV-Sox6FLAG expressing vector. When these extracts are used, a stronger band, in the same position of that generated by the endogenous Sox6, is observed in both nuclear (lane 2) and total (lane 8) extracts. This band is specifically supershifted by increasing amounts of the anti-FLAG antibody (lanes 3–4 and 9), but it is not altered by the addition of the unrelated anti-GATA-1 antibody (lanes 5–6).

When tested in the same conditions, the probe mutated in both Sox6 binding sites (MUT) fails to give any retarded band, confirming that the above described band is specifically due to Sox6 binding to the double consensus binding site (Figure 4B; lanes 10–18).

To further confirm the specificity of Sox6 binding to the double consensus lying within its own promoter, we set up competition experiments with an oligonucleotide corresponding to a published Sox6 consensus located in the regulatory regions of the cartilage-specific gene Col2a1 (4). As shown in Figure 4C, the Col2a1 unlabeled probe efficiently competes for the Sox6 band (lanes 3–5). When the two oligonucleotides mutated in the single Sox6 binding sites (Mut 1 and Mut 2 in Figure 4A) are used as cold competitors, both oligonucleotides compete for the Sox6 band with a decreased affinity (lanes 6–11) if compared with the Col2a1 probe (lanes 3–5), and with the unlabelled WT probe (lane 2). Direct binding on the Mut1- and Mut2-labeled probes confirms the reduced ability of these oligonucleotides to bind Sox6 (Figure 3C; compare lanes 15–17). Taken together these data suggest that both single Sox6 binding sites contribute to the overall Sox6 binding observed with the WT probe.

The −775 Sox6 binding sites within the Sox6 promoter mediate transcriptional repression in K562 cells

To test the functional role of the −775/−759 Sox6 binding sites, we prepared a series of Luciferase reporter constructs that we assayed by transient transfection experiments in K562 cells. Each experiment was done with at least two independent DNA plasmid preparations, each of them transfected at least in triplicate.

We first introduced within the 1116-S6 construct described above the same point mutations previously shown to completely abolish Sox6 binding in EMSA (Figure 4A) and we compared the response of the resulting WT and Mut plasmids (1116-S6wt and 1116-S6mut; Figure 5A) to increasing amount of Sox6 in cotransfection experiments in K562 cells. When cotransfected, Sox6 has a general activation effect on several promoters (Supplementary Figure S2; including 1116-S6). Taking this into account, the ratio between the activation of the

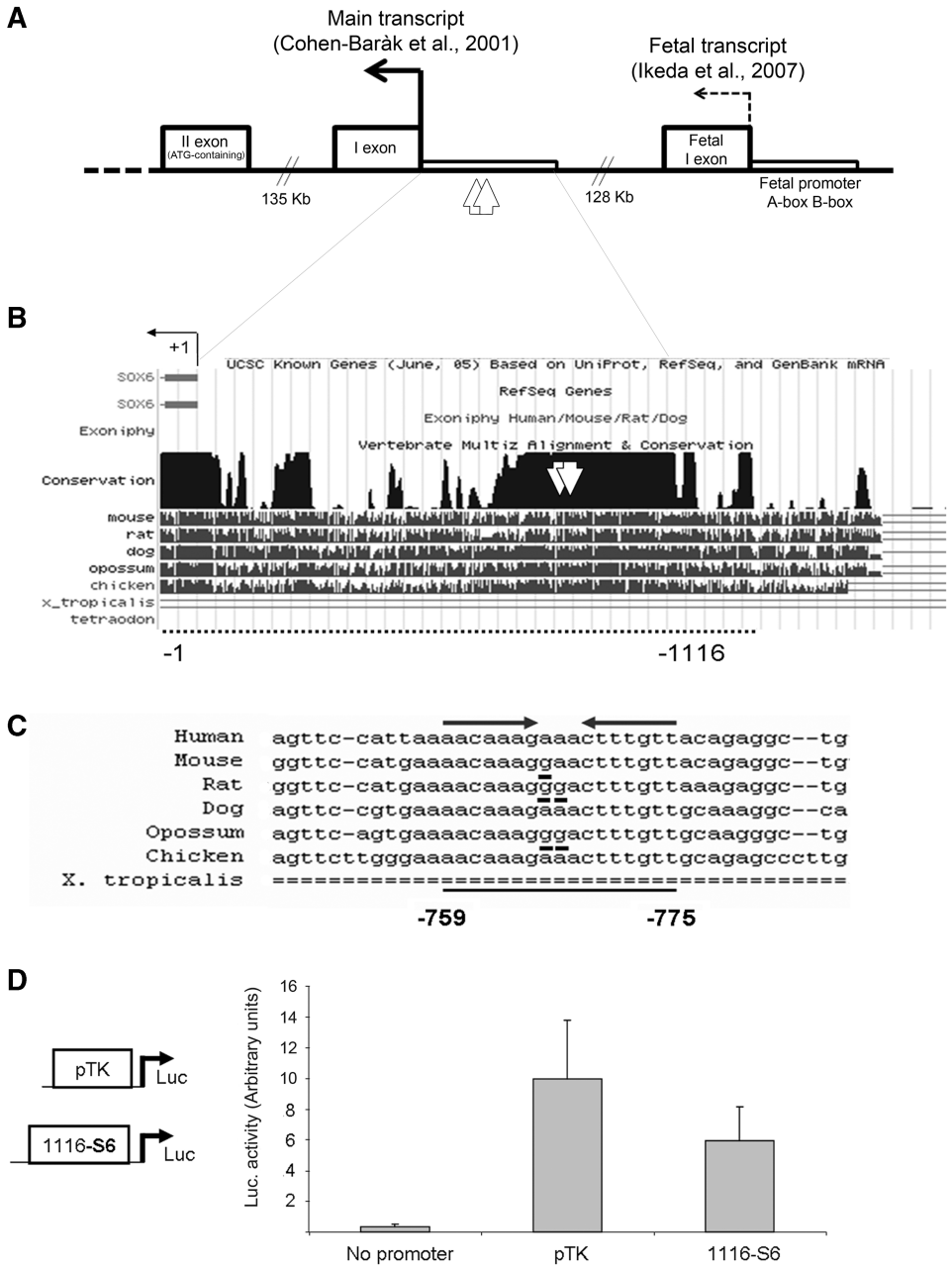

Figure 3. Fine mapping of the double Sox6 binding site on the human Sox6 promoter. (**A**) Schematic representation of the 5′-region of the human Sox6 locus on chromosome 11. The thick arrows indicate the position of the double Sox6 binding site, located within the region immediately upstream the main transcript start site. Note that the transcription orientation is from right to left. (**B**) UCSC genome browser (http://genome.ucsc.edu/) graphical map of the human Sox6 promoter region. The thick arrows indicate the same Sox6 binding sites as in **A**. The −1116 nt promoter region studied in transfection experiments is indicated with a dotted line under the panel. (**C**) Nucleotide conservation of the region (−775/−759) containing the two Sox6 binding sites (black arrows). The few substitutions in the spacer region between the two single sites are underlined. (**D**) The −1116 nt conserved region behaves as a promoter in K562 cells. The constructs used are schematically represented on the left of the figure. Luciferase activity is given in arbitrary units, standard deviations are represented on the top of each column.

WT versus the Mut construct is decreased in the presence of increasing amount of Sox6 (Figure 5B), suggesting a repressive role of Sox6 on the double Sox6 sequence.

To rule out the possibility that this effect might be due to non-physiological Sox6 concentrations in contransfected cells, we transfected the same constructs in P19 cells grown in the presence or in the absence of RA. Under RA induction, P19 cells accumulate Sox6 (as shown by RT–PCR, right panel), an event which preludes to their differentiation toward the neuronal lineage. As shown in Figure 5B the two constructs (1116-S6wt and 1116-S6mut) behave as in K562, confirming that the Sox6 double consensus mediates transcriptional repression in two independent cell contexts (K562 and P19 cells) in which Sox6 is actively transcribed.

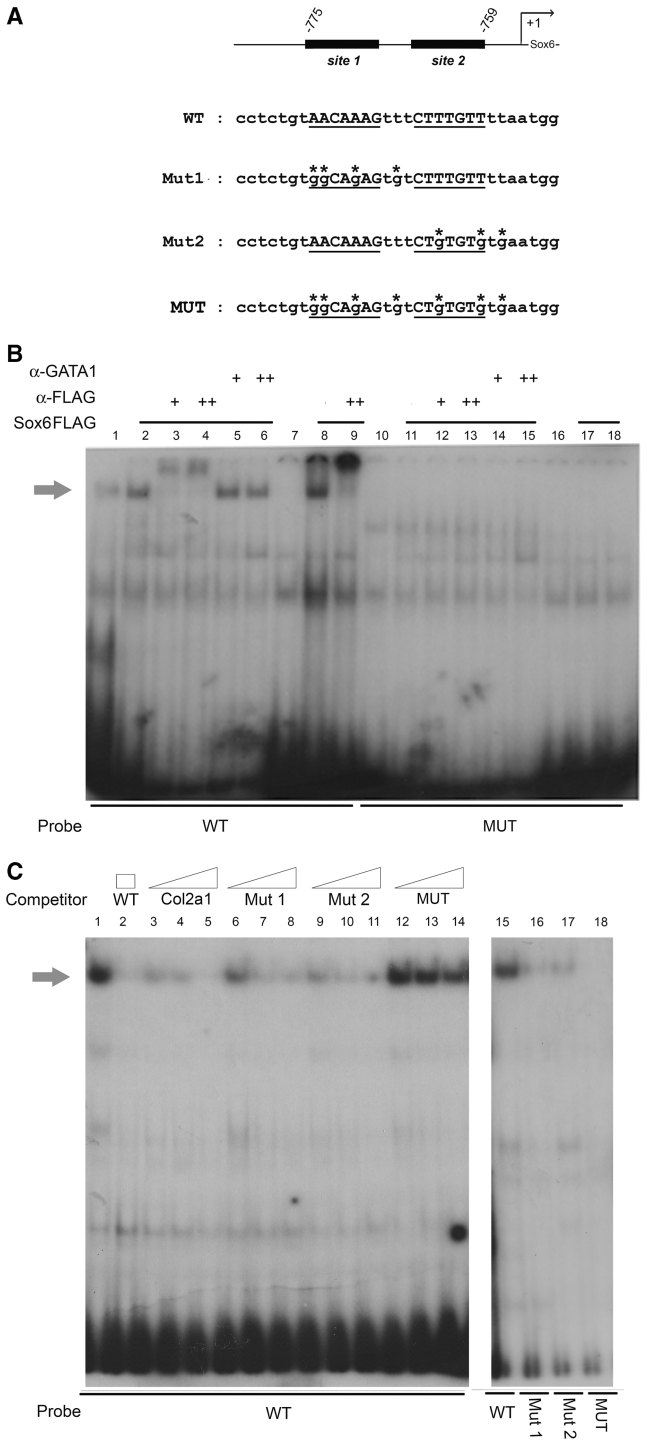

Figure 4. Sox6 binds with high affinity to the −775/−759 double consensus on human Sox6 promoter. (**A**) Nucleotide sequence of the four probes used in EMSA experiments. The two Sox6 sites are underlined and the mutations destroying the Sox6 consensus are marked by asterisks. (**B**) EMSA experiments: probes in panel A were ^{32}P labeled and incubated with either nuclear (lanes 1–6 and 10–16) or total K562 cell extracts (lanes 7–9 and 16–18). Extracts from K562 cells transfected with the pCMV-Sox6FLAG expressing vector: black bars over lane numbers; increasing amounts of antibodies are indicated as + and ++, respectively. ^{32}P labelled probes are indicated below the panel. The retarded band produced by Sox6 binding is indicated by the grey arrow: its specificity is demonstrated by the use of a specific anti-FLAG antibody (lanes 3 and 4). The MUT probe, mutated in both Sox6 consensus sites, fails to give any retarded band when tested in the same conditions as for the wild-type (WT) probe (lane 10–18). (**C**) Unlabeled competitor oligonucleotides (sequences in panel A) were added at three increasing concentrations and are indicated on top of the panel. WT: lane 2 (at the intermediate concentration). Col2a1-derived Sox6 consensus site (ref. 4): lanes 3–5. Mut1: lanes 6–8. Mut2: lanes 9–11 MUT: lanes 12–14.). Right panel: direct binding of the WT and Mutated probes.

The activity of the same two constructs as above (1116-S6wt and 1116-S6mut) was also tested in primary erythroblasts differentiated from CD34$^+$ cells purified from both peripheral and cord blood (Figure 5C). In these experiments cells were nucleofected at two time points, corresponding to the onset (Day 8) and to the maximum (Day 12) of Sox6 expression, respectively (as assayed by Real Time PCR, upper panels; see also Figure 1), and analyzed 24 h after transfection. Under these conditions, the 1116-S6wt construct (middle panels) is downregulated at Day 12 in correspondence of the Sox6 increase; in contrast, the activity of the 1116-S6mut construct, although slightly lower than that of the WT construct at Day 8, is not significantly modified between Day 8 and Day 12 (lower panels). By combining the data obtained with both the peripheral and cord blood cultures (Figure 5D), the mutated construct results to be significantly more active ($P < 0.03$) at late phases of differentiation (Day 12). This repression is associated to direct *in vivo* binding of SOX6 to its own promoter, as demonstrated by immunoprecipitation of chromatin from cells at the same Day 12 with an anti-Sox6 antibody (Figure 5E), further suggesting that the binding of SOX6 to this element contributes to the repression of Sox6 transcription when Sox6 expression reaches high levels.

Finally, to test whether this element is able to mediate transcriptional repression when transferred to an heterologous, highly active transcriptional context, we restricted the Sox6 5′-flanking region described earlier (the 1116nt, Figure 3) to 234 nt containing this site and we cloned it upstream to a strong erythroid-specific regulatory cassette (consisting of the core of the HS2 Gata-1 enhancer linked to the 330 nt Gata-1 minimal promoter, (16, pe+WT). We also made a similar construct (pe+MUT) in which the double Sox6 site is mutated as above (Figure 5F).

These constructs were cotransfected in K562 cells together with increasing amounts of the pCMV Sox6FLAG overexpressing plasmid.

As shown in Figure 5G, the pe+WT plasmid is progressively repressed in a dose dependent manner by the simultaneous cotransfection of increasing amounts of the Sox6 expressing vector (7, 11, 15 dark columns). On the other hand, the corresponding pe+MUT construct, mutated in the Sox6 consensus (columns 8, 12, 16) is insensitive to Sox6 repression. Moreover, the effect of Sox6 on the double Sox6 binding site is highly specific, since the cotransfection of a similar Sox4 overexpressing plasmid (at the highest concentration used for Sox6) fails to modify the activity of the pe+WT reporter (column 19). The amount of Sox6 and Sox4 produced by the two plasmids is similar, as shown by the western blot on the right panel.

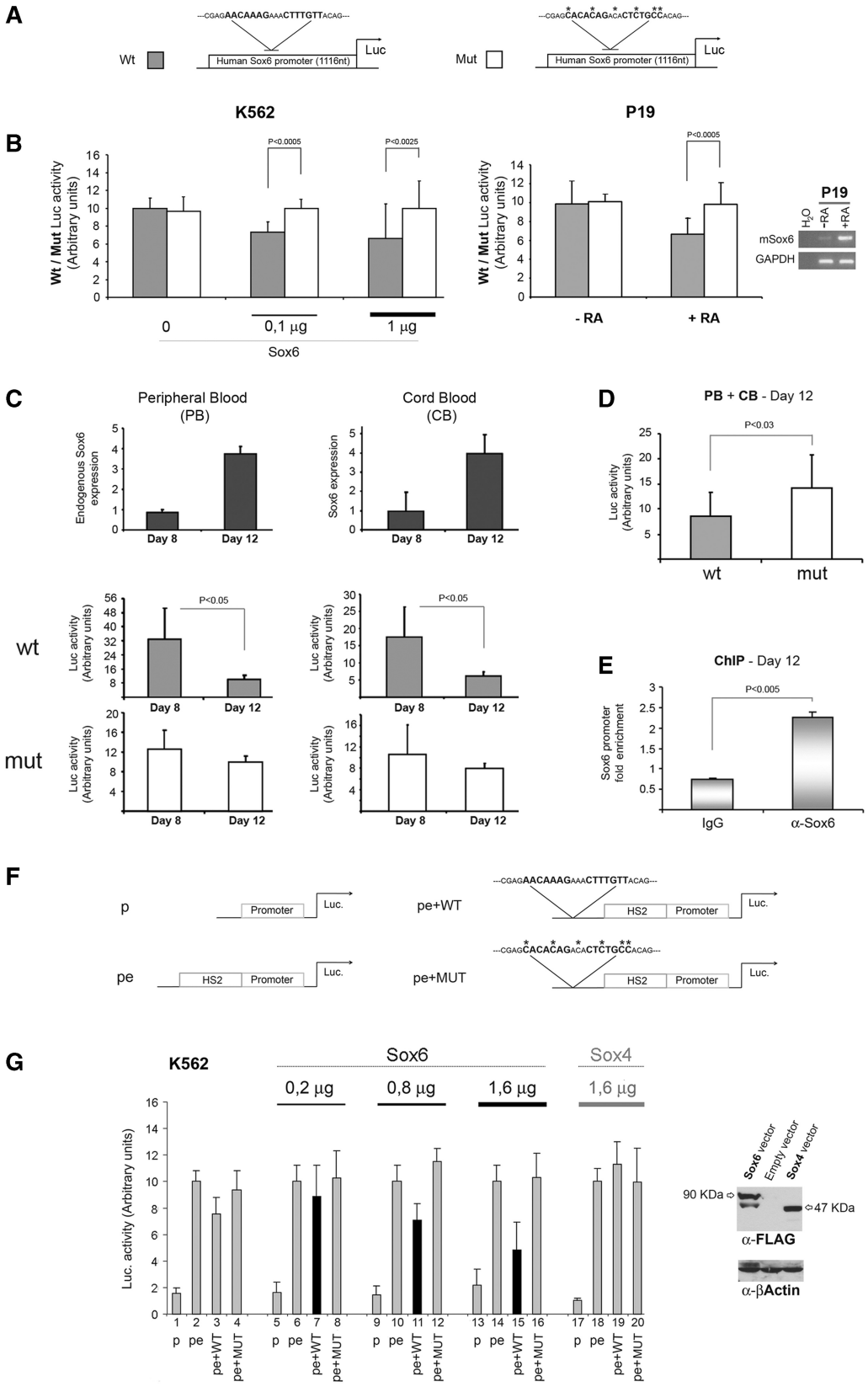

Figure 5. Sox6 represses its own promoter in cotransfection experiments in K562 cells. (**A**) The −1116/−1 region, either WT or mutated in the Sox6 consensus was cloned upstream to the Luciferase reporter gene. The mutations in the Sox6 consensus are the same as in Figure 4A, shown to abolish Sox6 binding in EMSA. (**B**) The two constructs were transfected either in K562 cells (left panel) with increasing doses of a Sox6 expressing plasmid, or in P19 cells (right panel) induced or not with RA. RA treatment induces Sox6 in P19 cells as shown by RT–PCR (right panel). In both cell lines, Sox6 induces a repression of the WT construct when compared to the mutated one. Statistical analyses are indicated above the chart. (**C**) Primary

Continued

The endogenous Sox6 gene is repressed by high levels of exogenously overexpressed Sox6

Since increasing amounts of Sox6 downregulate the Sox6 double consensus-Luciferase reporter construct (pe+WT; Figure 5), we wished to confirm the effect of Sox6 overexpression on the endogenous Sox6 gene transcription, by overexpressing in K562 cells a Sox6 protein that could be distinguished from the endogenous Sox6. To this end, we cloned a Sox6 cDNA lacking the 3′ 147 nt, in frame with a FLAG epitope, within a lentiviral expression vector. The resulting construct generates a protein lacking the C-terminal 49 aminoacids, still retaining its biological activity (19). Endogenous versus exogenous Sox6 can be distinguished both at RNA (by RT–PCR; Figure 6A, right panel) and at protein level (Figure 6B, right panel, and 'Materials and Methods' section).

In both K562 and Human Cord Blood (HCB) *in vitro*-derived erythoblasts, the overexpression of the exogenous Sox6 gene is mirrored by a dramatic reduction of both the endogenous Sox6 transcript and protein (Figure 6A and B, left panels), suggesting that high levels of Sox6 indeed strongly repress the transcription of the endogenous gene.

To test whether this repression is associated to direct *in-vivo* binding of the Sox6FLAG protein on the −775/−759 motif within the Sox6 promoter, we performed chromatin immunoprecipitation (ChIP) on K562 cells transduced with either the construct overexpressing the Sox6FLAG protein or the corresponding empty vector (K562-S6 and K562, respectively, in Figure 6C). The anti-FLAG antibody was used to detect the exogenous Sox6. As shown in Figure 6C, the anti-FLAG antibody efficiently immunoprecipitates the Sox6 double site region in cells expressing the exogenous Sox6FLAG protein (lane 6). The same sequence is not immunoprecipitated by the same anti-FLAG antibody when chromatin from K562 cells transduced with the empty vector (lane 5) is used. As a further negative controls, normal Rabbit-IgG fail to immunoprecipitate the Sox6 promoter region in both K562-S6 and empty vector-K562 chromatin samples (lanes 3–4) and primers designed on the GAPDH gene (lower gel) give no amplification on the same immunoprecipitated samples as above. Real Time PCR quantification of Immunoprecipitated bands demonstrates that the Sox6FLAG binding on the Sox6 promoter region is stronger than that observed for the cyclin D1 promoter, a known *in-vivo* target of Sox6 in insulinoma INS-1E and NIH-3T3 cells (25), and thus used as positive control (Figure 6C, lower panel).

Together, these data confirm that Sox6 is indeed able to repress its own transcription, possibly by direct binding *in vivo* to its own promoter.

This observation prompted us to address in more detail the chromatin dynamics of this region upon Sox6 overexpression (mimicking the accumulation of Sox6 during erythroid differentiation), by performing an *in vivo* Restriction Enzyme Accessibility Assay. To this end we exploited the presence of a PstI site 200 nt upstream to the double Sox6 consensus. As shown in Figure 6D, in K562 cells infected with the empty vector, this region is accessible to PstI cut, to an extent expected for an intermediate hypersensitive site (22). In the presence of high concentrations of Sox6 PstI is unable to access its target site, confirming that Sox6 binding to its target sequences correlates with a 'closed' status of this region. As a control, in both conditions (i.e. either basal level of or high level of Sox6) a PstI site within the Gata-1 gene results equally accessible in both conditions.

Sox6 binds to the conserved double consensus *in vivo* in murine bone marrow erythroid cells

To confirm the above observations *in vivo*, we moved to primary mouse Bone Marrow cells.

ChIP carried out with an anti-Sox6 antibody confirmed the ability of Sox6 to bind to the mouse conserved Sox6 double element (Figure 7A). To refine our analysis we then FACS sorted from mouse Bone Marrow two cell populations: the $CD71^{+}TER119^{+}$ cells (R2), representing erythroid progenitors, and cells negative for both markers (R1), representing more immature cells (Figure 7B). ChIP on chromatins from these populations reveals that Sox6 is indeed able to bind to its consensus in erythroid cells (R2) while it does not bind to the same sequence in $CD71^{-}TER119^{-}$ cells. Of note, Sox6 is already expressed at low levels in the more immature cells (Figure 1D) but it binds to its consensus only in more mature cells ($CD71^{+}TER119^{+}$ cells), where it is present at high concentration.

Since Sox6 can mediate transcriptional repression by recruiting HDAC1 [as on the CycD promoter; (25)], we tested whether HDAC1 was recruited on the Sox6 consensus. Figure 7C shows that the binding of Sox6 to its promoter is not associated with HDAC1 recruitment,

Figure 5. Continued
erythroblasts obtained from $CD34^{+}$ cells purified from either Peripheral Blood (left panels) or Cord Blood (right panels) were nucleofected with the 1116-S6wt and 1116-S6mut constructs at Days 8 and 12 of the cultures, in the presence of a Renilla-expressing plasmid to normalize for relative transfection efficiencies. Upper panels: RealTime PCR assessing the expression level of the endogenous Sox6 (relative to GAPDH). Middle panels: Luciferase activity driven by the 1116-S6wt element; Lower panels: Luciferase activity driven by the 1116-S6mut element at the same days (no statistical significant differences were found). Luciferase activity is given in arbitrary units. (**D**) The activity of the 1116-S6mut is higher than that of the 1116-S6wt construct at late phases of erythroid differentiation (Day 12) when data from Peripheral and Cord Blood cells are plotted together (normalized on Renilla expression). (**E**) ChIP demonstrates that at Day 12 SOX6 is indeed bound to its promoter. IgG antibodies were used as a control. (**F**) A 234-nt fragment (either the WT or the mutated Sox6 double site) was cloned upstream to a highly active Gata-1-derived erythroid cassette (pe construct). (**G**) All constructs were cotransfected in K562 cells together with increasing amounts (from 0.2 to 1.6 μg) of a Sox6 expressing plasmid. The pe+WT construct is repressed in a dose dependent manner by the addition of the cotransfected Sox6 expressing plasmid (black bars, columns 7, 11, 15), while the corresponding mutated element, pe+MUT, is not (columns 8, 12, 16). A Sox4 expressing plasmid (at the highest concentration of 1.6 μg used for Sox6) fails to repress pe+WT, suggesting a Sox6 highly specific effect (lane 19). Sox6-FLAG and Sox4-FLAG expressing plasmids produce comparable amounts of protein, as demonstrated by using the anti-FLAG antibody in the western blot on the right (1.6 μg of transfected plasmids, extracts from 5×10^5 transfected cells per lane).

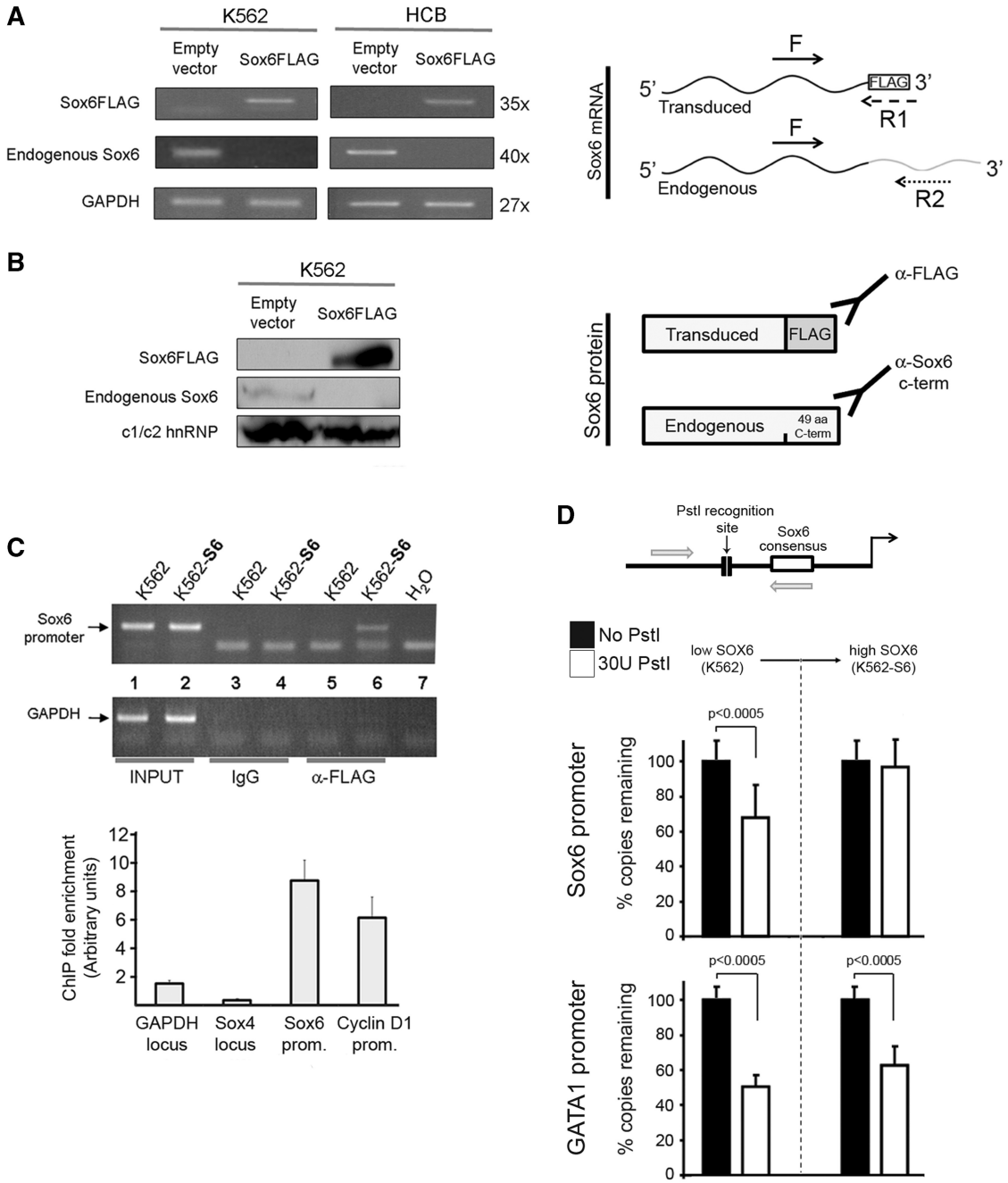

Figure 6. Overexpression of exogenous Sox6 represses the endogenous Sox6 transcription. (**A**) Semiquantitative RT–PCR on cDNA from K562 and HCB-derived progenitors cells transduced with either the Sox6 overexpressing vector or the corresponding empty vector. GAPDH was used to normalize for cDNAs loading. The number of amplification cycles used for each set of primers is indicated on the right of the figure. In Real Time quantification, the endogenous Sox6 transcript—in K562 and HCB cells overexpressing the Sox6FLAG protein—was scored as undetectable (not shown). Right panel: primers combinations used to discriminate between endogenous versus exogenous Sox6 transcript. (**B**) Western blot analysis. Nuclear extracts from 7×10^5 cells were loaded in each lane. Right panel: antibodies used to discriminate between endogenous versus exogenous Sox6 protein: the anti-FLAG antibody detects exogenous SOX6, while the endogenous protein is revealed by anti-SOX6 antibody raised again the C-terminal portion of the protein, absent in the exogenous protein. The anti c1/c2 hnRNP antibody was used to normalize for protein loading. (**C**) ChIP experiments. The anti-FLAG or rabbit IgG antibodies were used to immunoprecipitate chromatin from either K562 (transduced with the empty vector) or K562 overexpressing Sox6FLAG (K562-S6) cells. Lanes 1 and 2: input chromatins. Lanes 3 and 4: normal rabbit IgG. Lane 5 and 6: anti-FLAG antibody. Lane 7: water. Upper gel: Sox6 promoter region. Lower gel: GAPDH genomic locus was used as a negative control. Lower panel: Real Time quantification of the Sox6 promoter enrichment in chromatin immunoprecipitated with the anti-FLAG antibody. GAPDH and SOX4: negative control loci, cyclin D1: positive control. Real Time analysis was performed twice in triplicate. (**D**) Restriction enzyme accessibility assay. Schematic representation (not in scale) of the position of the PstI endonuclease restriction site relative to the Sox6 binding sites within the Sox6 promoter. Grey arrows indicate primers used to amplify the uncut copies of this region. In presence of low Sox6 level, PstI accesses the Sox6 promoter region (30% of cut copies, with *P*-value <0.0005) while in the presence of high Sox6 level it does not. A PstI recognition site within the human GATA1 promoter is used as a positive control.

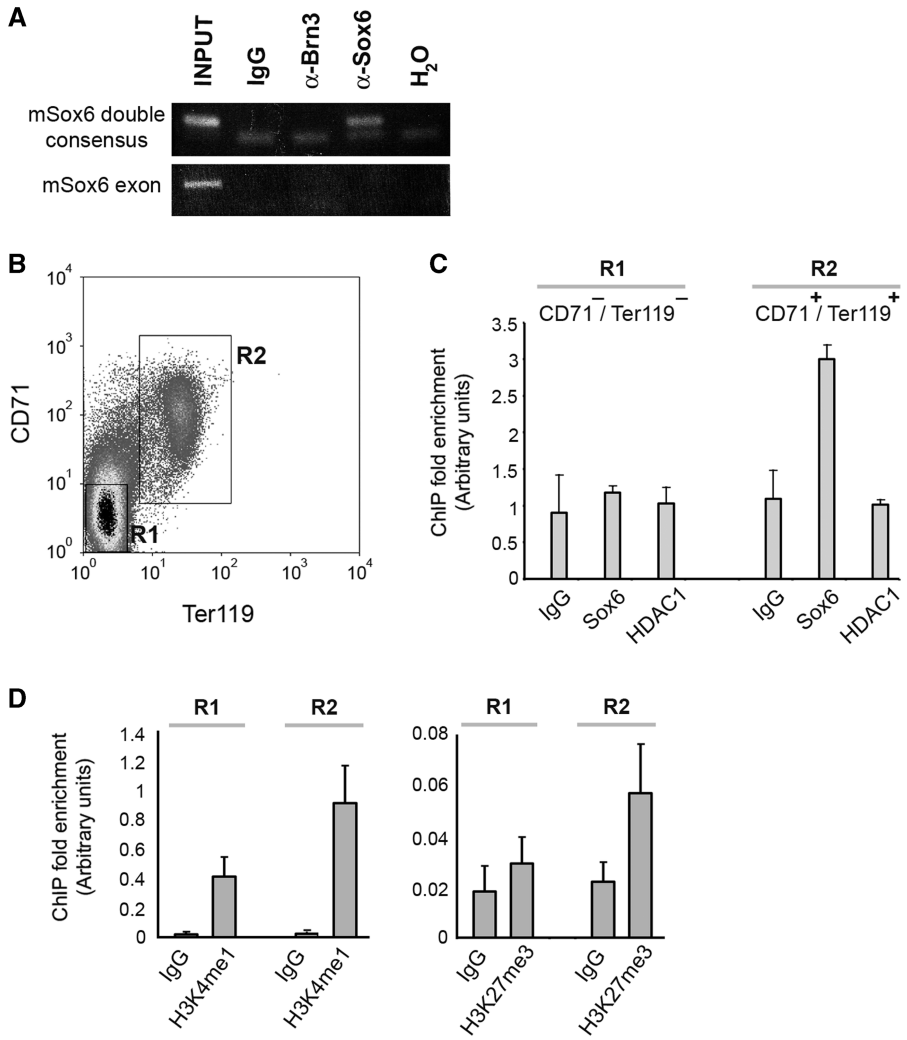

Figure 7. Sox6 binds the murine homologous Sox6 consensus region in bone marrow erythroid cells. (**A**) ChIP experiments. The anti-Sox6, rabbit IgG, and the unrelated anti-Brn3 antibodies were used to immunoprecipitate chromatin from mouse total bone marrow. Upper gel: double Sox6 consensus region. Lower gel: region within the Sox6 exon 16, used as a negative control. (**B**) FACS analysis and sorting. Two cells populations were sorted from murine bone marrow: erythroid $CD71^{+}Ter119^{+}$ (R2) and $CD71^{-}Ter119^{-}$ (R1). (**C**) Chromatins from R1 and R2 cells populations were immunoprecipitated using rabbit IgG, anti-Sox6 and anti-HDAC1 antibodies. (**D**) The same chromatins were immunoprecipitated using rabbit IgG and anti-H3K4me1 (left) and anti-H3K27me3 (right) antibodies.

suggesting that the repression on this element likely relies on different molecular mechanisms.

To get insight into the chromatin conformation of this region in primary cells we tested the chromatin accessibility of this region by looking at the presence of histone modifications generally associated to either transcriptionally active (H3K4me1) or transcriptionally repressed (H3K27me3) chromatin (26). As shown in Figure 7D, in the more immature cell population where the level of Sox6 is very low (see Figure 1D), the double Sox6 region is Immunoprecipitated by the anti H3K4me1 but not by the anti H3K27me3 antibody. In the more mature population, in correspondence with the higher Sox6 concentration, a significant enrichment in H3K27me3-modified chromatin is observed together with H3K4me1 enrichment on the double Sox6 consensus region. These data are in agreement with the hypothesis that in the presence of high Sox6 concentrations this region is involved in the establishment of a repressed chromatin domain.

DISCUSSION

Sox6 plays critical roles in determining cell fate and differentiation in different systems, including erythropoiesis; however, the molecular mechanisms underlying its function are very poorly understood. In particular, despite its profound effect on red cell survival, proliferation and differentiation (11), the only Sox6 target known in erythropoietic cells is the 36 nt sequence within the mouse εy-globin promoter (13). The direct binding of Sox6 to this sequence has been proposed to repress εy-globin gene expression in adult erythropoietic cells, thus suggesting the Sox6 requirement for proper embryonic-adult globin switching to occur. The εy site is

composed of two single Sox6 consensus sites in opposite orientation, spaced by 8 nt. The prevalence of paired consensi on the few known Sox6 target genes (1–3), suggests that two sites are likely required for Sox6 function, although very little is known about their relative configuration and their spacing.

In the attempt to uncover new potential Sox6 targets in erythroid cells, we undertook a genome-wide search of evolutionarily conserved double Sox6 binding sites, taking as a model the Sox6 consensus lying within the mouse εy-globin gene promoter. By combining the bioinformatic search with a data set of genes differentially expressed during erythoid differentiation, we identified seven potential targets of Sox6 regulation (Figure 2D), the Sox6 gene itself included.

The double Sox6 binding site mapping within the proximal promoter of the human Sox6 gene itself is composed, according to the 'paired sites module' often observed in Sox6 binding sites, by two sites in opposite orientation and identical in their sequence (5′AACAAA G3′ and 5′CTTTGTT3′, respectively), spaced by 3 nt. As in the case of the Sox6 sites on the εy-globin promoter, the newly identified Sox6 target site studied in this article is repressed by Sox6 binding (see below). In recent literature, Sox6 has been proposed to work as a repressor, through interactions with a variety of partner factors, in several cell types (13,25,27,28). On the other hand, Sox6, together with the highly related Sox5 and Sox9 genes, was originally identified as a master gene in chondrogenic differentiation, where this 'Sox trio' activates the expression of chondrogenic specific genes, such as type II collagen (*Col2a1*), aggrecan (*Agc*), cartilage link protein (*Crtl1*) and matrilin (3,4,29,30). On these targets, Sox5, Sox6 and Sox9 are thought to secure each other to their binding sites to activate transcription (note that, while Sox6 and Sox5 do not have a transactivation domain, Sox9 possesses it, thus working as a typical transcription factor) (31). Altogether, these data suggest that interactions with various protein partners and/or different DNA arrangements of Sox6 target sites might underlie the different transcriptional outputs observed (activation or repression) in different systems. The repression of Sox6 on the Sox6 promoter does not seem to be mediated by HDAC1 recruitment, as shown in Figure 7C. The observation of different histone modifications and of a varied accessibility (Figures 6C and 7) on the double Sox6 consensus region *in vivo* suggests the differential recruitment of chromatin remodelling/histones-modifying factors in cells at different degrees of erythroid differentiation.

Sox6 downregulates its own expression by direct binding to an upstream negative element

The Sox6 gene structure is extremely complex, as it involves different promoters and possibly enhancers which are spread over hundreds of kilobases (8,9), and it is quite likely that the expression of Sox6 in different tissues and during development relies for its active/negative regulation on multiple regulatory elements, so far poorly characterized. Our results identify a double Sox6 consensus binding site, located −775/−759 nt upstream to one of the identified human Sox6 transcriptional start sites (24); we propose that this site may participate in the regulation of Sox6 expression during erythroid maturation, by negatively modulating the transcription from this promoter in erythroid cells. Indeed, this site is directly bound by Sox6 both *in vitro* (Figure 4; EMSA experiments) and *in vivo* (ChIP, Figures 5E and 6C) and is required for Sox6-dependent repression of reporter constructs in both K562 (and P19) cells lines as well as in erythroblasts obtained from primary $CD34^+$ cells from Peripheral and Cord Blood (Figures 5B and C). The binding of exogenous Sox6 to the double binding site within the Sox6 promoter (Figure 6C) correlates with a change in chromatin accessibility (Figure 6D), further indicating that the binding of Sox6 to this region is indeed relevant to this repression. According to this hypothesis, the increase of Sox6 in bone marrow erythroid cells is accompanied by the appearance of repressive histone marks in this region (Figure 7D). Moreover, overexpression of exogenous Sox6 causes a strong downregulation of the endogenous Sox6 transcript and protein, both in K562 cells and in primary erythroid cultures (Figure 6A and B).

Recently, Ikeda *et al*. (32) identified an alternative Sox6 promoter active in embryonic tissues and located nearly 128-kb upstream to the sequence investigated in the present study. The corresponding transcript is far less abundant in K562 cells than the main transcript studied in this paper, and decreases, as the main transcript, upon Sox6 overexpression (as shown in Figure 6A, where the primers used in RT–PCR detect both transcripts). Whether this transcript is also directly repressed by Sox6 in erythroid cells is presently not known.

Sox6 autoregulation during erythroid differentiation

Genetic circuits of autoregulation of key transcription factors, affecting the rate of their synthesis by influencing the rate of transcription (either by autoactivation or autorepression), are known to govern many processes involving progressive commitment of pluripotent progenitors into lineage-restricted cells (33,34) where they are thought to reinforce cell-fate decisions.

Having shown that Sox6 overexpression is able to repress endogenous Sox6 transcription, we asked whether the kynetics of Sox6 expression during normal erythropoiesis may reflect possible Sox6 roles in autoregulation. Our experiments show that, following a sharp increase of Sox6 expression during the transition from undifferentiated progenitors to early erythroblasts, Sox6 is progressively downregulated during late stages of erythropoiesis (Figure 1). This observation is consistent with a negative modulatory role of Sox6 on its own promoter, although it is possible that changes in other transcription factors (dependent or independent from Sox6) as well as other sequences within the Sox6 locus might also contribute to the regulation of Sox6 transcription.

Many examples of autoregulation are known within the hematopoietic system: Gata-1 is expressed at low levels in multipotent progenitors, but becomes abundant in

committed erythroid precursors where its transcription is sustained by a positive feedback loop (35,36 and references therein). Later on, Gata-1 expression declines at advanced stages of erythroid differentiation (36,37). Interestingly, the initial upregulation of Gata-1 in erythroid differentiation is also linked to repression of another member of the family, Gata-2, which predominates in early progenitors, and may bind to a subset of Gata1-binding sites, often eliciting opposite effects to those of Gata-1 (38,39). In this regard, it would be of great interest to test whether other Sox family members can share common targets with Sox6, either at different stages of erythroid differentiation or in other tissues in which Sox6 is important for cell differentiation. A candidate for such a role in hematopoiesis is Sox4, which is expressed in early progenitors prior to Sox6 and is progressively repressed during differentiation (unpublished results), with a kinetics opposite to that of Sox6 induction.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Drs. Thamar van Dijk, Sjaak Philipsen, Michiko Hamada-Kanazawa, Roberto Mantovani, Cristina Fugazza and Tariq Enver for reagents and advice for lentiviral work, Maria Rosa Lidonnici for help with nucleofection experiments and Stefania Citterio for FACS analyses.

FUNDING

This work was supported by PRIN 2007 (to M.D.C.), by the Italian Telethon Foundation (to G.F.), Fondazione CARIPLO 2005 (to S.O.) and PRIN 2006 to (A.R.). Funding for open access charge: Ronchi FAR Unimib.

Conflict of interest statement. None declared.

REFERENCES

1. Guth,S.I. and Wegner,M. (2008) Having it both ways: Sox protein function between conservation and innovation. *Cell Mol. Life Sci.*, **65**, 3000–3018.
2. Kiefer,J.C. (2007) Back to basics: Sox genes. *Dev. Dyn.*, **236**, 2356–2366.
3. Lefebvre,V., Dumitriu,B., Penzo-Mendez,A., Han,Y. and Pallavi,B. (2007) Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *Int. J. Biochem. Cell Biol.*, **39**, 2195–2214.
4. Lefebvre,V., Li,P. and de Crombrugghe,B. (1998) A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J.*, **17**, 5718–5733.
5. Hagiwara,N., Klewer,S.E., Samson,R.A., Erickson,D.T., Lyon,M.F. and Brilliant,M.H. (2000) Sox6 is a candidate gene for p100H myopathy, heart block, and sudden neonatal death. *Proc. Natl Acad. Sci. USA*, **97**, 4180–4185.
6. Connor,F., Wright,E., Denny,P., Koopman,P. and Ashworth,A. (1995) The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse. *Nucleic Acids Res.*, **23**, 3365–3372.
7. Stolt,C.C., Schlierf,A., Lommes,P., Hillgartner,S., Werner,T., Kosian,T., Sock,E., Kessaris,N., Richardson,W.D., Lefebvre,V. *et al.* (2006) SoxD proteins influence multiple stages of oligodendrocyte development and modulate SoxE protein function. *Dev. Cell*, **11**, 697–709.
8. Ikeda,T., Kamekura,S., Mabuchi,A., Kou,I., Seki,S., Takato,T., Nakamura,K., Kawaguchi,H., Ikegawa,S. and Chung,U.I. (2004) The combination of SOX5, SOX6, and SOX9 (the SOX trio) provides signals sufficient for induction of permanent cartilage. *Arthritis Rheum.*, **50**, 3561–3573.
9. Cohen-Barak,O., Yi,Z., Hagiwara,N., Monzen,K., Komuro,I. and Brilliant,M.H. (2003) Sox6 regulation of cardiac myocyte development. *Nucleic Acids Res.*, **31**, 5941–5948.
10. Hagiwara,N., Yeh,M. and Liu,A. (2007) Sox6 is required for normal fiber type differentiation of fetal skeletal muscle in mice. *Dev. Dyn.*, **236**, 2062–2076.
11. Dumitriu,B., Patrick,M.R., Petschek,J.P., Cherukuri,S., Klingmuller,U., Fox,P.L. and Lefebvre,V. (2006) Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development. *Blood*, **108**, 1198–1207.
12. Cohen-Barak,O., Erickson,D.T., Badowski,M.S., Fuchs,D.A., Klassen,C.L., Harris,D.T. and Brilliant,M.H. (2007) Stem cell transplantation demonstrates that Sox6 represses epsilon y globin expression in definitive erythropoiesis of adult mice. *Exp. Hematol.*, **35**, 358–367.
13. Yi,Z., Cohen-Barak,O., Hagiwara,N., Kingsley,P.D., Fuchs,D.A., Erickson,D.T., Epner,E.M., Palis,J. and Brilliant,M.H. (2006) Sox6 directly silences epsilon globin expression in definitive erythropoiesis. *PLoS Genet.*, **2**, e14.
14. Xu,J., Sankaran,V.G., Ni,M., Menne,T.F., Puram,R.V., Kim,W. and Orkin,S.H. (2010) Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev.*, **24**, 783–798.
15. Migliaccio,G., Di Pietro,R., Digiacomo,V., Di Baldassarre,A., Migliaccio,A.R., Maccioni,L., Galanello,R. and Papayannopoulou,T. (2002) In vitro mass production of human erythroid cells from the blood of normal donors and of thalassemic patients. *Blood Cells Mol. Dis.*, **28**, 169–180.
16. Ronzoni,L., Bonara,P., Rusconi,D., Frugoni,C., Libani,I. and Cappellini,M.D. (2008) Erythroid differentiation and maturation from peripheral $CD34^+$ cells in liquid culture: cellular and molecular characterization. *Blood Cells Mol. Dis.*, **40**, 148–155.
17. Schreiber,E., Matthias,P., Muller,M.M. and Schaffner,W. (1989) Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells. *Nucleic Acids Res.*, **17**, 6419.
18. Bose,F., Fugazza,C., Casalgrandi,M., Capelli,A., Cunningham,J.M., Zhao,Q., Jane,S.M., Ottolenghi,S. and Ronchi,A. (2006) Functional interaction of CP2 with GATA-1 in the regulation of erythroid promoters. *Mol. Cell. Biol.*, **26**, 3942–3954.
19. Hamada-Kanazawa,M., Ishikawa,K., Nomoto,K., Uozumi,T., Kawai,Y., Narahara,M. and Miyake,M. (2004) Sox6 overexpression causes cellular aggregation and the neuronal differentiation of P19 embryonic carcinoma cells in the absence of retinoic acid. *FEBS Lett.*, **560**, 192–198.
20. Demaison,C., Parsley,K., Brouns,G., Scherr,M., Battmer,K., Kinnon,C., Grez,M. and Thrasher,A.J. (2002) High-level transduction and gene expression in hematopoietic repopulating cells using a human immunodeficiency [correction of imunodeficiency] virus type 1-based lentiviral vector containing an internal spleen focus forming virus promoter. *Hum. Gene Ther.*, **13**, 803–813.
21. Ronchi,A., Cirò,M., Cairns,L., Basilico,L., Corbella,P., Ricciardi-Castagnoli,P., Cross,M., Ghysdael,J. and Ottolenghi,S. (1997) Molecular heterogeneity of regulatory elements of the mouse GATA-1 gene. *Genes Funct.*, **1**, 245–258.
22. McArthur,M., Gerum,S. and Stamatoyannopoulos,G. (2001) Quantification of DNaseI- sensitivity by Real-time PCR: Quantitative Analysis of DNaseI-hypersensitivity of the Mouse β-globin LCR. *J. Mol. Biol.*, **313**, 27–34.

23. Donaldson,I.J. and Gottgens,B. (2006) TFBScluster web server for the identification of mammalian composite regulatory elements. *Nucleic Acids Res.*, **34**, W524–W528.
24. Cohen-Barak,O., Hagiwara,N., Arlt,M.F., Horton,J.P. and Brilliant,M.H. (2001) Cloning, characterization and chromosome mapping of the human SOX6 gene. *Gene*, **265**, 157–164.
25. Iguchi,H., Urashima,Y., Inagaki,Y., Ikeda,Y., Okamura,M., Tanaka,T., Uchida,A., Yamamoto,T.T., Kodama,T. and Sakai,J. (2007) SOX6 suppresses cyclin D1 promoter activity by interacting with beta-catenin and histone deacetylase 1, and its down-regulation induces pancreatic beta-cell proliferation. *J. Biol. Chem.*, **282**, 19052–19061.
26. Li,B., Carey,M. and Workman,J.L. (2007) The role of chromatin during transcription. *Cell*, **128**, 707–719.
27. Murakami,A., Ishida,S., Thurlow,J., Revest,J.M. and Dickson,C. (2001) SOX6 binds CtBP2 to repress transcription from the Fgf-3 promoter. *Nucleic Acids Res.*, **29**, 3347–3355.
28. von Hofsten,J., Elworthy,S., Gilchrist,M.J., Smith,J.C., Wardle,F.C. and Ingham,P.W. (2008) Prdm1- and Sox6-mediated transcriptional repression specifies muscle fibre type in the zebrafish embryo. *EMBO Rep.*, **9**, 683–689.
29. Rentsendorj,O., Nagy,A., Sinko,I., Daraba,A., Barta,E. and Kiss,I. (2005) Highly conserved proximal promoter element harbouring paired Sox9-binding sites contributes to the tissue- and developmental stage-specific activity of the matrilin-1 gene. *Biochem. J.*, **389**, 705–716.
30. Han,Y. and Lefebvre,V. (2008) L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer. *Mol. Cell. Biol.*, **28**, 4999–5013.
31. Lefebvre,V. (2002) Toward understanding the functions of the two highly related Sox5 and Sox6 genes. *J. Bone Miner. Metab*, **20**, 121–130.
32. Ikeda,T., Saito,T., Ushita,M., Yano,F., Kan,A., Itaka,K., Moro,T., Nakamura,K., Kawaguchi,H. and Chung,U.I. (2007) Identification and characterization of the human SOX6 promoter. *Biochem. Biophys. Res. Commun.*, **357**, 383–390.
33. Akashi,K., Traver,D., Miyamoto,T. and Weissman,I.L. (2000) A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, **404**, 193–197.
34. Orkin,S.H. (2000) Diversification of haematopoietic stem cells to specific lineages. *Nat. Rev. Genet.*, **1**, 57–64.
35. Ferreira,R., Ohneda,K., Yamamoto,M. and Philipsen,S. (2005) GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol. Cell. Biol.*, **25**, 1215–1227.
36. Shimizu,R. and Yamamoto,M. (2005) Gene expression regulation and domain function of hematopoietic GATA factors. *Semin. Cell Dev. Biol.*, **16**, 129–136.
37. Tanabe,O., Shen,Y., Liu,Q., Campbell,A.D., Kuroha,T., Yamamoto,M. and Engel,J.D. (2007) The TR2 and TR4 orphan nuclear receptors repress Gata1 transcription. *Genes Dev.*, **21**, 2832–2844.
38. Whyatt,D., Lindeboom,F., Karis,A., Ferreira,R., Milot,E., Hendriks,R., de Bruijn,M., Langeveld,A., Gribnau,J., Grosveld,F. *et al.* (2000) An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells. *Nature*, 519–524.
39. Bresnick,E.H., Martowicz,M.L., Pal,S. and Johnson,K.D. (2005) Developmental control via GATA factor interplay at chromatin domains. *J. Cell Physiol.*, **205**, 1–9.