Published online 30 December 2011 *Nucleic Acids Research, 2012, Vol. 40, No. 8* ***3403–3418***
doi:10.1093/nar/gkr1290

Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells

Pietro Vella, Iros Barozzi, Alessandro Cuomo, Tiziana Bonaldi and Diego Pasini*

Department of Experimental Oncology, European Institute of Oncology (IEO), Via Adamello 16, 20139 Milan, Italy

Received September 26, 2011; Revised December 3, 2011; Accepted December 14, 2011

ABSTRACT

The Yin Yang 1 (YY1) transcription factor is a master regulator of development, essential for early embryogenesis and adult tissues formation. YY1 is the mammalian orthologue of Pleiohomeotic, one of the transcription factors that binds Polycomb DNA response elements in *Drosophila melanogaster* and mediates Polycomb group proteins (PcG) recruitment to DNA. Despite several publications pointing at YY1 having a similar role in mammalians, others showed features of YY1 that are not compatible with PcG functions. Here, we show that, in mouse Embryonic Stem (ES) cells, YY1 has genome-wide PcG-independent activities while it is still stably associated with the INO80 chromatin-remodeling complex, as well as with novel RNA helicase activities. YY1 binds chromatin in close proximity of the transcription start site of highly expressed genes. Loss of YY1 functions preferentially led to a down-regulation of target genes expression, as well as to an up-regulation of several small non-coding RNAs, suggesting a role for YY1 in regulating small RNA biogenesis. Finally, we found that YY1 is a novel player of Myc-related transcription factors and that its coordinated binding at promoters potentiates gene expression, proposing YY1 as an active component of the Myc transcription network that links ES to cancer cells.

INTRODUCTION

Yin Yang 1 (YY1) is a DNA binding transcription factor discovered 20 years ago as the main binding factor, induced by the adenoviral protein E1, of the adeno associated virus (AAV) promoter region and takes its name from the dual activity of the AAV promoter (1). YY1 is also the mammalian orthologue of pleiohomeotic (pho), one of the DNA binding transcription factors that mediate Polycomb Group (PcG) proteins binding at the Polycomb Response Elements (PRE) of the *Drosophila melanogaster* genome (2).

PcG proteins have a key role in early embryogenesis. They are master regulators of organism development that control cell fate by maintaining repression of their target genes, in part through their ability to modify histone proteins within the surroundings of their binding sites (3). Until now, very few DNA binding factors have been described to have the ability to recruit PcG proteins to specific chromatin sites and YY1 is one of the best candidates (3). In fact, similar to PcG proteins, YY1 activity results essential for mammalian development, as YY1-null embryos die at the peri-implantation stages of embryogenesis (4). YY1 activity is necessary also for adult tissue development: for instance, oligodendrocytes-specific depletion of YY1 causes serious neural defects, mainly due to lack of global nerves myelination (5). Moreover, reduced YY1 expression in heterozygous knock out (KO) mice induces serious growth retardation, proliferative and neurological defects (6). Altogether, these data show the critical role of YY1 in regulating several developmental processes and highlights its similarities with PcG activities.

Several reports proposed YY1 as a potential recruiting factor for Polycomb activities in mammalian cells. For example, YY1 was shown to directly interact with the Polycomb Repressive Complex 2 (PRC2) subunit Eed in Burkitt's lymphoma cells (7), to mediate PcG recruitment during myoblasts differentiation (8) and during muscles regeneration from satellite stem cells (9). Moreover, YY1 binding sites were identified into a putative PRE element isolated in mammalian cells and it was shown that, when the binding sites are mutated, PRE responsiveness is affected (10). Finally, recent data on X-chromosome inactivation proposed YY1 as a DNA–RNA binding factor that links PcG-Xist to the inactive X-chromosome (11).

Despite these observations, several data pointed at YY1 having PcG-independent functions. YY1 was shown to interact to with the INO80 complex in cancer cell lines and proposed to have a positive effect on *Cdc6* expression (12,13). Similarly, YY1 role in nerve myelination was linked to direct YY1 binding at the *Egr2* promoter and to activation of *Egr2* expression in Schwann cells (14). In

*To whom correspondence should be addressed. Tel: +39 02 94375139; Fax: +39 02 94375990; Email: diego.pasini@ifom-ieo-campus.it

line with this, it was shown that YY1 interacts with several other transcription factors often linked with transcriptional activation (6). In addition, YY1 was reported to control p53 levels in a DNA independent manner (15) and to bind the cruciform structure of Holliday junctions, suggesting a role in DNA repair via homologous recombination that is consistent with the genomic instability observed in YY1 deficient fibroblasts (13). Many of these studies are based on *in vitro* or non-physiological observations, often based on experiments made on single genes without determining a direct YY1 association. Thus, these data do not completely clarify YY1 functions and particularly do not fully address the real transcriptional nature of YY1. We therefore believe that a detailed analysis of YY1 activity in a biologically relevant system is needed to define YY1 functions at a genome-wide level. Due to YY1 essential role in early embryogenesis (4) and the high degree in similarities with the phenotypes observed in mutant mice for different PcG proteins (4,16–19), we decided to characterize YY1 functions in mouse embryonic stem (ES) cells.

ES cells are the tissue culture adaptation of the cells that form blastocyst's inner cell mass (20). Mouse ES cells can be expanded through an active BMP and STAT3 signaling that maintains ES pluripotent state by preserving their potential to give rise to all cells of an adult organism (20). Such signaling stimulates the activity of several transcription factors that are required for their maintenance and differentiation (21). The same transcription factors are also actively involved in the reprogramming of committed cells to a pluripotent ES-like state (22), highlighting how the activity of these proteins is essential for ES cell identity. Kim and colleagues (23) recently dissected such transcriptional network and identified three distinct transcription modules: a PcG-module, strictly linked to transcriptional repression; a core-module, made of transcription factors that directly respond to the BMP-STAT signaling pathway (Oct4-Nanog-Sox2); and a Myc module, made of transcription factors such as c-Myc, n-Myc, E2fs and Zfx. In this work, the authors showed that the Myc-module, but not the core module, is responsible for the previously identified ESC-like transcriptional features of cancer cells. Moreover, they proposed that high activity of the ES Myc-module predicts a poor outcome of different kind of human tumors. Consistent with this, Myc, like PcG proteins and YY1, is frequently over-expressed in cancers and several studies demonstrated direct oncogenic effects in mediating normal cell transformation and tumor development (6,24,25).

In the present study, we show at a genome-wide level that YY1 exerts PcG-independent functions in ES cells. Like in cancer cells, YY1 is associated with all components of the INO80 chromatin-remodeling complex, as well as to newly identified partners with RNA helicase activity. YY1 is preferentially associated with hyperacetylated promoters with high transcriptional activity. Loss of YY1 functions in ES cells predominantly diminished mRNAs expression while increased the expression levels of small non-coding RNAs such as small nuclear, nucleolar and micro RNAs. In addition, we identified components of the Myc transcription module as potential cooperating factors at YY1 sites and demonstrated that YY1 binding is prevalently associated with promoters co-occupied by other transcription factors such as c-Myc, n-Myc, Zfx and E2f1 at a genome-wide level. Finally, we show that a coordinated occupancy of YY1 with the Myc transcription module correlates with an increased expression of target genes proposing YY1 as a partner of the Myc network.

MATERIALS AND METHODS

Cell lines generation, manipulation and culturing

All ES cell lines were grown on 0.1% gelatinized tissue culture dishes in DMEM supplemented with 15% Serum (Euroclone), Leukemia Inhibitory Factor (produced in house), Penicillin-Streptomycin (Gibco), non-essential aminoacids (Gibco), Na-Pyruvate (Gibco). BirA expressing ES cell clones were generated from an ES cell line described elsewhere (26) by removing the puromycin selection cassette used for targeting purposes by transient CRE recombinase expression. The expression constructs for Fbio-Ezh2 and Fbio-YY1 were generated by LR recombination of the YY1 and EZH2 coding sequences from a pCR8 Gateway entry vector into a pCAG-Flag-Avi-ires-Puromycin Gateway compatible destination vector using LR recombinase (Invitrogen). Stable cell lines were obtained by transient transfection of FBio empty, YY1 and Ezh2 expression constructs in BirA-ES cells using Lipofectamine 2000 (Invitrogen) and stable selection with 2 μg/ml of puromycin. RNA interference (RNAi) experiments were carried out by transfecting short interfering RNA (siRNA) oligos specific for YY1 (Sigma-Aldrich SASI_Mm01 00125709) or with a control-scrambled (SCR) sequence (Sigma-Aldrich SIC001) using Lipofectamine 2000 at a concentration of 50 nM. Cells were harvested 48 hours post-transfection and RNA isolated by TRIzol (Invitrogen) extraction. Stable shRNA knock down was obtained with ES cell transduction with viral particles produced with the LKO.1 vectors TRCN0000054556 (shYY1) and SHC202 (shSCR) purchased from Sigma-Aldrich.

Antibodies

Western blot analyses were performed using antibodies against: YY1 (Santa Cruz, cat. sc-281); Suz12 (Santa Cruz, Cat. sc-46264), Actr8 (Sigma-Aldrich, Cat.A2107), Ddx5 (Abcam, Cat. ab10261), Ddx3x (Millipore, Cat. #09-860), Vinculin (Sigma-Aldrich, Cat. V9131), HA (Santa Cruz, Cat. sc-805), Oct4 (Abcam, Cat. ab19857), β-Tubulin (Santa Cruz, Cat. sc-9104), Biotin (Pierce, Cat. 31852). Ezh2 and Eed were described elsewere (18). Ruvbl2 was also previously described (27).

Immunoprecipitation and Chromatin-Immunoprecipitation (ChIP) analyses were carried out using antibodies against: YY1 (Santa Cruz, Cat. sc-281), Suz12 (Cell Signaling, Cat. 3737), cMyc (N-262) (Santa Cruz, Cat. sc-764) and E2f1 (C-20) (Santa Cruz, Cat. sc-193). Rabbit IgG (Sigma, Cat. I5006) were used as negative control.

Tandem affinity purification and Mass spectrometry analysis

All protein purifications were carried out on ES cell nuclei prepared by 20 min swelling in nuclear prep buffer (10 mM Tris, 100 mM NaCl, 2 mM $MgCl_2$, 0.3 M Sucrose, 0.25 % v/v Igepal) at 4°C. Nuclei were lysed in high salt buffer (50 mM Tris–HCl pH 7.5, 300 mM NaCl, 10% glycerol, 0.25% Igepal) with fresh addition of a protease inhibitor cocktail (Roche). Direct streptavidin purifications were carried out by over-night (ON) incubation of 25 µl of streptavidin magnetic beads (Invitrogen, Cat. 656-01) for each milligram of protein extract. The tandem affinity purifications were performed by incubating ~20 mg of nuclear protein extract with 200 µl of packed anti-Flag agarose beads (Sigma, Cat. A2220) ON at 4°C on a rotating platform. Beads were washed six times in minimum 10 beads volumes of high salt buffer at 4°C and protein complexes eluted for 30 min with 0.5 mg/ml of flag peptide (DYKDDDDK) in high salt buffer at 20°C four times. Eluates were pulled and further precipitated with 100 µl of streptavidin magnetic beads (Invitrogen cat. 656-01) ON at 4°C. Streptavidin beads were washed six times as before at 4°C and protein complexes eluted with Laemli sample buffer (Invitrogen).

Gel electrophoresis and in-gel digestion

Proteins from both FBio-YY1 and FBio empty vector control purification were separated by 1D SDS–PAGE, using 4–12% NuPAGE® Novex Bis–Tris gels (Invitrogen) and NuPAGE® MES SDS running buffer (Invitrogen) according to manufacturer's instructions. The gel was stained with coomassie Blue using Colloidal Blue Staining Kit (Invitrogen). Samples were digested with trypsin (Promega). Briefly, the gel bands were cut and then washed four times with 50 mM ammonium bicarbonate, 50% ethanol and incubated with 10 mM DTT in 50 mM ammonium bicarbonate for 1 h at 56°C for protein reduction. Alkylation step was performed incubating the sample with 55 mM iodoacetamide in 50 mM ammonium bicarbonate for 1 h at 25°C in the dark. Gel pieces were washed two times with a 50 mM ammonium bicarbonate, 50% acetonitrile solution, dehydrated with 100% ethanol and dried in a vacuum concentrator. Digestion was performed using 12.5 ng/ml trypsin in 50 mM ammonium bicarbonate and incubated for 16 h at 37°C for protein digestion. Supernatant was transferred to fresh tube, and the remaining peptides were extracted by incubating gel pieces two times with 30% acetonitrile (MeCN) in 3% trifluoroacetic acid (TFA), followed by dehydration with 100% acetonitrile. The extracts were combined, reduced in volume in a vacuum concentrator, desalted and concentrated using RP-C18 StageTip columns and the eluted peptides used for mass spectrometric analysis (28).

Mass spectrometry analysis

Peptide mixtures were separated by nano-LC/MSMS using an Agilent 1100 Series nanoflow LC system (Agilent Technologies), interfaced to a 7-Tesla LTQ-FT-Ultra mass spectrometer (ThermoFisher Scientific, Bremen, Germany). The nanoliter flow LC was operated in one column set-up with a 15-cm analytical column (75 µm inner diameter, 350 µm outer diameter) packed with C18 resin (ReproSil, Pur C18AQ 3 µm, Dr Maisch, Germany). Solvent A was 0.1% FA and 5% ACN in ddH_2O and Solvent B was 95% ACN with 0.1% FA. Samples were injected in an aqueous 0.1% TFA solution at a flow rate of 500 nl/min. Peptides were separated with a gradient of 0–40% Solvent B over 90 min followed by a gradient of 40–60% for 10 min and 60–80% over 5 min at a flow rate of 250 nl/min. The mass spectrometer was operated in a data-dependent mode to automatically switch between mass spectrometry (MS) and MS/MS acquisition. In the LTQ-FT full scan MS spectra were acquired in a range of m/z 300–1700 by FTICR with resolution $r = 100\,000$ at m/z 400 with a target value of 1 000 000. The five most intense ions were isolated for fragmentation in the linear ion trap using collision-induced dissociation at a target value of 5000. Singly charged precursor ions were excluded. In the MS/MS method, a dynamic exclusion of 60 s was applied and the total cycle time was ~2 s. The nanoelectrospray ion source (Proxeon, Odense, Denmark) was used with a spray voltage of 2.4 kV. No sheath and auxiliary gases were used and capillary temperature was set to 180°C. Collision gas pressure was 1.3 millitorrs and normalized collision energy using wide band activation mode was 35%. Ion selection threshold was 250 counts with an activation $q = 0.25$. The activation time of 30 ms was applied in MS2 acquisitions.

Data analysis and assigning sequences using MASCOT

The raw data from LTQ-FT Ultra were converted to mgf files using Raw2MSM software (29). The MS/MS peak lists were filtered to contain at most six peaks per 100 Dalton intervals and searched by Daemon (version 2.2.2, Matrix Science) against a concatenated forward and reversed version of IPI mouse database (version 6.63) (56 073 sequences; 25 214 299 residues) (30). This database was complimented with frequently observed contaminants (porcine trypsin and human keratins) and their reversed sequences as well. Search parameters were: an initial MS tolerance of 7 ppm, a MS/MS mass tolerance at 0.5 Da and full trypsin cleavage specificity, allowing for up to two missed cleavages. Carbamidomethylation of cysteine was set as a fixed modification and variable modifications included oxidation on methionine and acetylation on N-terminus of proteins. We accepted peptides and proteins with a false discovery rate (FDR) of <1%, estimated based on the number of accepted reverse hits (31).

ChIP, bioChIP and high-throughput sequencing

ChIP assays were carried out as described previously (32). Briefly, 1% formaldehyde cross-linked chromatin was fragmented by sonication to an average size of 200–350 bp and immunoprecipitated ON with 10 µg of indicated antibodies. For bioChIP, 25 µg of streptavidin beads were added instead of the antibodies, following the protocol described in (33). DNA samples were sequenced on an Illumina Genome Analyzer II. About 36-bp short reads were then mapped onto the mm9 release of the

mouse genome using Bowtie (34). The alignments were performed allowing zero to two mismatches and keeping only the reads that align to unique positions in the genome. The YY1 sample was compared with the control DNA using Model-based Analysis for Chip-Seq [MACS, (35)]. Wiggle tracks for the visualization on the UCSC genome browser (36) were generated using MACS. Gene Interval Notator [GIN, (37)] was then used to annotate peaks over RefSeq mouse genes. A peak was assigned to the transcriptional start site (TSS) of a RefSeq gene when falling into the surrounding 4 kb (±2 kb). Datasets are available for download from NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under accession number GSE31786. Row data from previously published ChIPseq datasets were aligned to the mm9 release following the same criteria. Raw ChIPseq data and relative negative controls were obtained from the following GEO accession numbers: Jarid2 GSE19365; Ezh2 and Ring1b GSE13084; H3K27me3 and H3K4me3 GSE12241; H3K27AC GSE24164; cMyc, nMyc, Zfx, E2f1, [A]Oct4 and Sox2 GSE11431; [B]Oct4 and Nanog GSE11724.

De novo motif discovery

Multiple EM for Motif Elicitation [MEME, (38), version 4.4.0] was used in order to search for highly occurring pattern in the DNA sequence underlying the putative binding sites. The analysis was narrowed to the 50 bp (±25 bp) around the peak summit. All the identified putative binding sites were included. The analysis was run looking on both strands for motifs with zero or one occurrence per sequence (zoops), ranging from 6 to 16 bp in length.

Motif analysis

Position-specific weight matrices (PWMs) were collected from the literature (39–43), and used to build a custom set of 597 models. The YY1 putative binding site identified through *de novo* motif discovery was added to this set. For some analyses, PWMs were clustered using BLiC (44). In this way we could reduce the complexity of our results using a non-redundant set of 229 PWMs.

In order to identify over-represented PWMs in the YY1 putative binding sites proximal (±2.5 kb from a RefSeq TSS) regions were analyzed using Clover (45). The DNA sequences underlying the YY1 peaks were scanned for all the PWMs in the redundant set. Over-representation was statistically evaluated using three independent background sets, namely the entire chromosome 19, all the RefSeq TSSs (±2.5 kb) and all the CpG islands annotated in the mm9 genome. A PWM was retained only when significantly over-represented ($P \leq 0.01$) compared with all of these backgrounds. Clover is available as a standalone tool while results were parsed using a custom Python script.

Regions bound by both c-Myc and n-Myc (now on referred as Myc) were intersected with the YY1 proximal peaks. In this way we defined three sets, namely the YY1-bound Myc-unbound, the YY1-bound Myc-bound and the YY1-unbound Myc-bound. For each different class of genes, we used Pscan (46) to detect statistically significant over-represented PWMs against a background dataset consisting of the three sets pulled together. In this case, the non-redundant set was used. In case a PWM showed $P \leq 0.01$ (two-tailed Welch's *t*-test) it was considered as significantly over-represented. The Pscan source code was modified in order to replace the statistical evaluation step based on the *z*-test with a step based on the *t*-test. The *t*-test is more suitable than the *z*-test when comparing datasets with similar cardinality (46). In order to get a graphical representation of the results, PWMs that were found significant in at least one class were retained. Values were log10-transformed and hierarchically clustered using average linkage and Pearson correlation as distance measure. A heat map was then drawn using this information. Pscan is available as a standalone application, whereas the clustering and the heatmap were performed using R.

Density profile clusters were generated using SeqMINER K-means ranked clustering REF within a 4-kb region centered on peaks' summit. Density values were generated using a 50-bp window.

Microarray and micro RNA analysis

RNA from two independent RNAi experiments was hybridized independently to Mouse Gene 1.0 ST Affymetrix Arrays. Signals were RMA normalized and probeset with a 1.3-fold expression difference and a 95% confidence determined by ANOVA were selected for the analyses. Datasets are available for download from NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under accession number GSE31786.

The micro RNA (miRNA) expression was determined using the TaqMan® Rodent MicroRNA A+B Cards Set v2.0 following manufacturer procedures. The miRNA with a 2-fold expression difference and a 90% *t*-test confidence were selected for the analyses.

Density profiles clusters

Density profiles clusters were generated using SeqMINER (47) K-means ranked clustering REF within a 4-kb region centered on peaks' summit. Density values were generated using a 50-bp window.

Real Time quantitative PCR

RT–qPCRs were carried out using Fast Sybergreen as previously described (48). Primers used for PCRs are listed in Supplementary Table S7.

Functional annotations

mRNA and miRNA functional annotation were generated using Ingenuity Systems Pathway Analysis (IPA; www.ingenuity.com). The miRNA target genes annotation was generated toward the validated miRNA database of IPA.

RESULTS

In order to test if YY1 shares regulatory functions with PcG proteins, we performed ChIP analysis in E14 ES cells using antibodies specific for mouse YY1 and a subunit of the Polycomb Repressive Complex 2 (PRC2), Suz12. As shown in Figure 1A, while Suz12 was strongly enriched at known PcG binding sites, the YY1 antibody did not show any significant enrichment. Such result either indicate that YY1 does not associate with these genomic sites or that the YY1 antibody is not efficient in ChIP assays. In order to bypass these technical issues, we decided to develop a purification strategy that takes advantage of the *in vivo* biotinylation of proteins in mouse ES cells (49). Such technique involves the constitutive expression of proteins of interest bearing a tag (Flag-Avi; FBio) recognized by the biotinylating enzyme BirA (Supplementary Figure S1A) allowing *in vivo* biotinylation of the tagged protein (33). Such a system has been previously used for native protein complexes purification, as well as for ChIP assays (49–51). For this, we took advantage of ES cells that carry the coding sequence of the BirA enzyme knocked-in the Rosa26 locus that drives BirA constitutive expression (BirA-ES) (26). As shown in Supplementary Figure S1B, these cells constitutively express physiological levels of a hemagglutinin-tagged version of the BirA enzyme and normal levels of the pluripotency marker Oct4. Thus, we generated stable cell lines expressing independently a biotinylated (bio) form of YY1 and Ezh2 (bioYY1 and bioEzh2), which is the catalytic subunit of the PRC2 complex (Supplementary Figure S1D and E). With these cells, we performed a streptavidin co-precipitation experiment using optimal extraction conditions (Supplementary Figure S1C) and demonstrated that neither endogenous YY1 nor bioYY1 co-precipitated components of the PRC2 complex, whereas the bioEZH2 protein efficiently co-precipitated endogenous core PRC2 subunits (Eed and

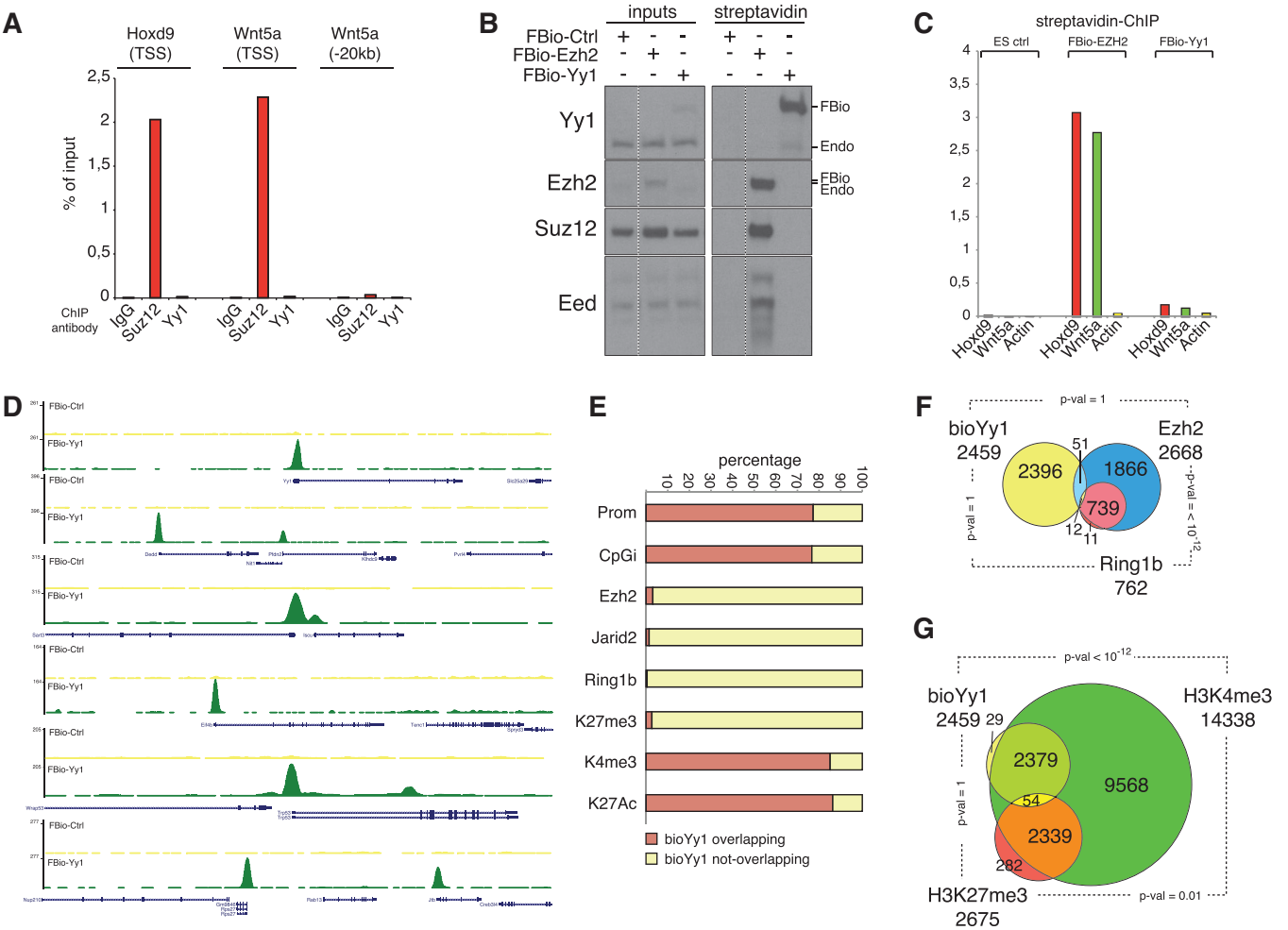


Figure 1. Genome-wide localization of YY1 does not correlate with PcG proteins (**A**) ChIP analysis using qPCR on the indicated genomic loci using the specified antibodies. (**B**) Western blot analyses using the indicated antibodies of streptavidin pull-down assays from protein extracts of ES cell lines independently expressing bioEzh2 and bioYY1. A control ES cell line with BirA expression alone is presented as purification control. Dotted line denotes removal of non-relevant lanes from the original blot. (**C**) BioChIP analysis by qPCR on chromatin prepared from the same cell lines presented in (B). (**D**) Genomic snapshots of the bioChIPseq results for YY1 and control (FBio-Crtl) BirA expressing ES cells. (**E**) Overlap of binding sites between bioYY1 and the indicated ChIPseq datasets. (**F** and **G**) Overlap between target genes of the indicated ChIPseq datasets. Target genes are defined by the presence of at least one peak within ±2 kb from RefSeq genes annotated TSS. *P*-values of the indicated overlaps are determined by hypergeometric distribution.

Suz12), (Figure 1B). Similar results were obtained using antibodies against endogenous proteins, further demonstrating the lack of interaction between YY1 and components of the PRC2 complex in ES cells (Supplementary Figure S1F). Using these cell lines, we also performed streptavidin ChIP (bioChIP) analyses at known PRC2 binding sites. Consistent with Figure 1A data, we found that, although bioEzh2 was efficiently enriched at PcG binding sites, bioYY1 was not (Figure 1C). Overall, these data confirm that YY1 does not interact with the PRC2 complex in ES cells and does not associate with the tested PcG binding sites.

In order to generate a genome-wide map of the bioYY1 binding profile to DNA, we performed high-throughput sequencing of the DNA enriched in bioYY1 ChIP (ChIPseq) (Figure 1E). Enrichment analysis of the bioYY1 ChIPseq, relative to a control BirA-ES cell line, revealed strong enrichment sites of bioYY1 in proximity of TSS (Figure 1D). Indeed, the overlap of all bioYY1 peaks relative to promoter regions, defined as ± 2 kb from TSS, showed that ~80% of bioYY1 peaks were found at genes promoters (Figure 1E). Consistent with this, bioYY1 binding showed a similar degree of overlap with CpG islands, a typical feature of promoter regions (Figure 1E). Conversely, when overlapped with previously generated ChIPseq datasets for PcG proteins in ES cells, bioYY1 binding sites did not overlap (<5%) with components of the PRC2 (Ezh2, Jarid2) and PRC1 (Ring1b) complexes, as well as with regions of accumulation of repressive tri-methylated (me3) histone H3 (H3) lysine (K) 27, which is the product of PRC2 enzymatic activity (Figure 1E) (25). This result is not due to a bias of chromatin accessibility in bio-YY1 BirA-ES cells since both H3K27me3 and unrelated genomic loci such as ES cells specific enhancers or intra-genic and inter-genic regions are efficiently immuno-precipitated with H3K27me and Histone H3-specific antibodies (Supplementary Figure S2A). These data are consistent with the observations presented in Figure 1A–C. Moreover, bioYY1 peaks strongly overlapped (>80%) with H3K4me3 and H3K27 acetylated (ac) regions (Figure 1E): while H3K4me3 was shown to form 'bivalent domain' of poised chromatin with H3K27me3 (52), H3K27ac was demonstrated to be mutually exclusive with H3K27me3 in ES cells (53). Moreover, while Ring1b and Ezh2 shared nearly all their entire set of targets, they did not significantly overlap bioYY1-bound promoters (Figure 1F). Consistent with this, bioYY1 target genes were not enriched for H3K27me3 but strongly enriched for H3K4me3, demonstrating that bioYY1 targets are not bivalent genes and further suggesting an association with actively transcribed promoters (Figure 1G). A more detailed analysis of the distribution profile of bioYY1 binding sites confirmed that bioYY1 is associated preferentially with promoter regions, to a lesser extent to intra-genic regions and, only for a remaining ~16% of binding sites, to inter-genic regions (Figure 2A). Interestingly, most of these intra- and inter-genic bioYY1 binding sites did not overlap significantly with CpG islands or with recently identified ES enhancer regions (<10%, data not shown) (54). Furthermore, the analysis of the density profiles of the distance between the summit of peaks and genes TSS showed that bioYY1 was strongly enriched in close proximity of transcriptional initiation (Figure 2B), in agreement with the examples presented in Figure 1D.

Comparison of ES cell microarray expression analyses with bioYY1 target genes demonstrated that bioYY1 is directly associated with the promoter of genes with high level of transcriptional activity (Figure 2C). Indeed, functional annotation of bioYY1 target genes revealed a strong enrichment in highly expressed genes, like genes encoding for proteins involved in RNA biogenesis, protein synthesis and mitochondrial functions involved particularly in embryonic development (Supplementary Figure S3A–C). Such result is consistent with the presence of non-bivalent H3K4me3 and with the high level of acetylation found on H3K27 at bioYY1 binding sites, suggesting a global activatory role of YY1 in regulating gene transcription. Finally, sequence analysis of bioYY1-bound genomic regions identified a motif that perfectly matches a known YY1 DNA binding site, strongly suggesting that bioYY1 genome-wide association to chromatin is directly mediated by its DNA binding activity (Figure2D).

In order to gain further insight into the functional properties of YY1, we used bioYY1 BirA-ES cells to identify YY1 specific interacting proteins. For this, we performed a tandem purification using the Flag and biotin tag and identified by mass spectrometry (MS) bioYY1-associated proteins in ES cells (Figure 2E). Such analysis revealed that bioYY1 is stably associated with several components of the INO80 chromatin-remodeling complex in ES cells, as previously reported for cancer cells (12,13) (Figure 2E). In addition, we identified novel interacting partners of bioYY1, Ddx5 and Ddx3x, two proteins carrying RNA helicase activity (Figure 2E). Consistent with Figure 1 data, no peptides of PcG proteins were found in the MS analysis. These interactions were further validated in an independent experiment probing the product of a streptavidin purification with specific antibodies against different proteins identified in the MS analysis (Figure 2F). These findings suggest that co-recruitment of the INO80 remodeling complex and the RNA helicase activities could contribute to promote active transcription from promoters bound by YY1.

In order to validate endogenous YY1 binding at target sites and the presence of its interacting partners, we performed ChIP analyses in ES cells using antibodies specific for YY1 and Ruvbl2, a stable component of the INO80 complex. As shown in Figure 2G, both YY1 and Ruvbl2 antibodies produced a significant enrichment over the background signal at several gene promoters, validating YY1 binding at the sites identified by the bioChIPseq analysis and demonstrating the co-recruitment of the INO80 complex at the same genomic regions.

To gain further insights into the role of YY1 in transcriptional regulation, we developed an efficient siRNA mediated down-regulation of YY1 expression in mouse ES cells (Figure 2H). Using total RNA extracted from two independent RNAi experiments (Figure 2H), we determined, by means of Affymetrix microarrays, global gene expression changes upon acute YY1

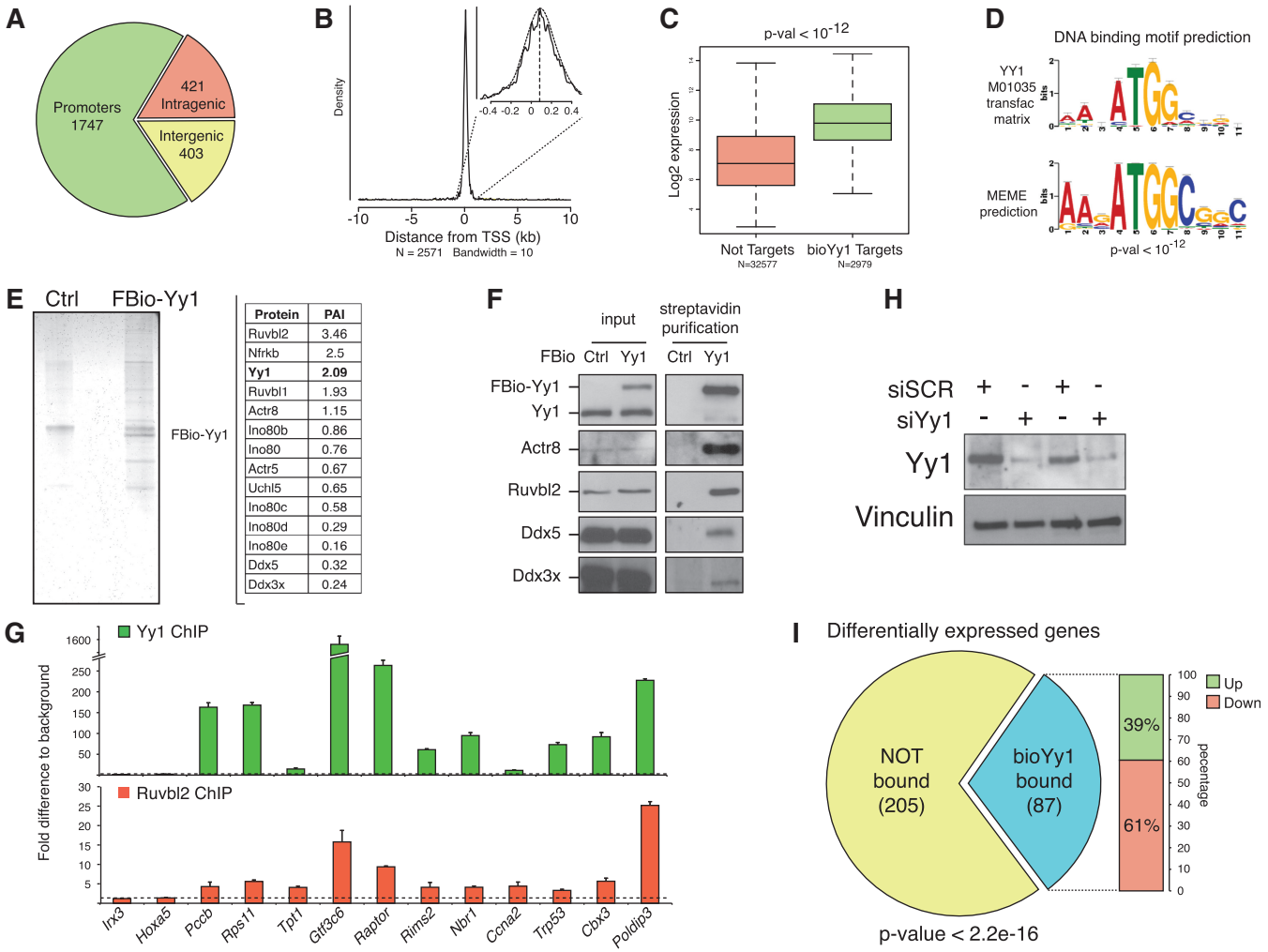


Figure 2. YY1 complex directly regulates active gene expression. (**A**) Distribution of bioYY1 binding sites relative to the gene bodies of RefSeq annotated transcripts. (**B**) Density profile of bioYY1 binding sites relative to TSS. All binding sites within ±10 kb are included in the analysis. TSS distance is measured as the relative base pair distance to peaks' summits. A close up image of a ±500 bp TSS density profile is also presented with an identical band with (10 bp). (**C**) Annotation of gene expression levels between bioYY1 target and non-target genes. *P*-values are determined by Wilcoxon test. (**D**) MEME motif prediction of DNA sequences enriched in bioYY1 ChIPseq. YY1 Transfac matrix is presented for comparison. (**E**) Silver staining of the isolated proteins with a Flag-Streptavidin tandem purification using protein extracts of FBio-YY1 expressing ES cells (left). A purification using BirA expressing ES cells is presented as negative control. A summary table of the mass spectrometry results is presented on the right. Protein Abundance Index (PAI) is indicated as measure of purification efficiency. (**F**) Western blot analyses of streptavidin-purified proteins from nuclear extracts of the indicated ES cell lines using the specified antibodies. Input lanes correspond to 2% of extract used in IPs (**G**) ChIP analysis of E14 ES cells using the indicated antibodies on the specified genes TSS. (**H**) Western blot analysis of ES cell extracts independently transfected with YY1-specific or scrambled (SCR) control siRNA oligos. Vinculin is presented as loading control. (**I**) Distribution of bioYY1 binding at the promoters of differentially expressed genes in YY1 siRNA-treated ES cells. *P*-values are determined with a chi-square test. Stacked columns show the relative distribution of up-regulated (Up) or down-regulated (Down) bioYY1 target genes.

down-regulation. Such analyses identified 292 genes that were differentially expressed, with a 95% confidence, between YY1 and SCR control siRNA-treated ES cells (Supplementary Table S2). Importantly, ~30% of the regulated genes present YY1 binding at their TSS. Such number is significantly higher than expected (chi-squared $P < 2^{-16}$) and strongly suggests a direct activity of YY1 in controlling the expression of its target genes (Figure 2I). Consistent with this, qRT–PCR analysis in cells treated with YY1-specific siRNA or shRNA targeting sequences validated the microarray results (Supplementary Figure S2B). Expression of YY1 target genes was preferentially diminished upon YY1 depletion, in agreement with an activatory role for YY1 (Figure 2I). Nevertheless, several transcripts were also up-regulated, suggesting potential opposing functions for YY1 in transcriptional control (Figure 2I). Interestingly, the most up-regulated transcripts in YY1 depleted ES cells are nuclear and nucleolar small non-coding RNAs (sncRNA) (Supplementary Table S3). A more detailed analysis of the whole microarray data identified ~22 RNA transcripts that were differentially regulated in absence of YY1 (Figure 3A). Nearly all these RNAs (86%) were up-regulated upon YY1 depletion (Figure 3A).

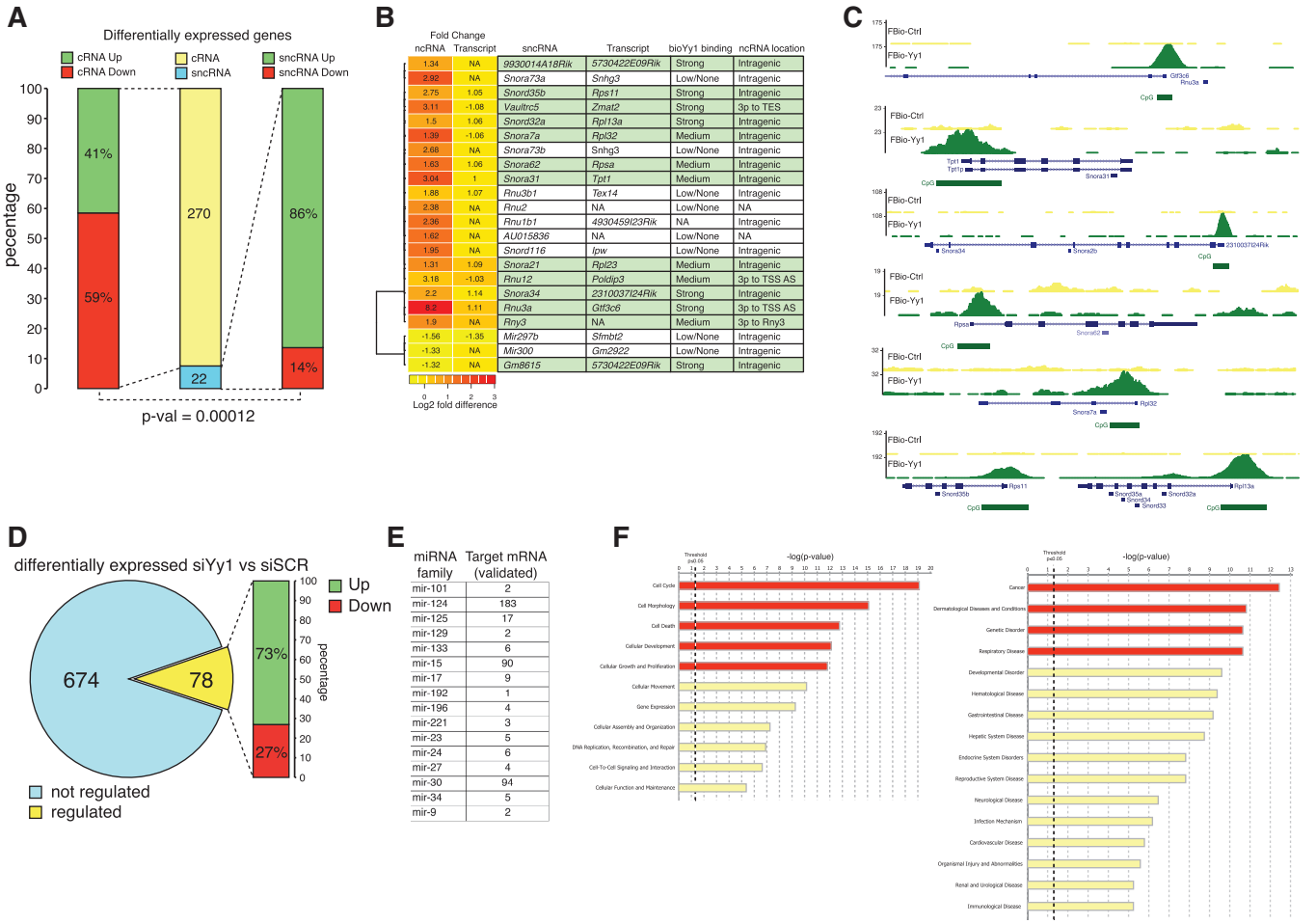


Figure 3. YY1 negatively regulates sncRNAs intra-cellular levels. (**A**) Distribution of the differentially expressed coding RNAs (cRNA) and small non-coding RNAs (sncRNA) present on the Mouse Gene 1.0 ST Affymetrix Array (center column). Differential distributions of up-regulated (Up) or down-regulated (Down) RNA classes are presented in the left and right columns. *P*-value is determined by chi-square test. (**B**) Heat map of the fold change expression values of the indicated sncRNAs and their annotated transcripts in YY1 RNAi-treated ES cells. Linear fold changes are also indicated within the heat map boxes. Left boxes indicate the presence of bioYY1 binding and its relative position with respect to transcripts TSS. The bioYY1 target sncRNAs are highlighted in green. YY1 binding intensity is defined using peaks *P*-value: low >10–40; mediun <10–40 and >10–70; High <10–70. (**C**) Genomic snapshots of the bioChIPseq results for YY1 and control (FBio-Crtl) at the indicated genomic loci. (**D**) Distribution of regulated miRNAs upon YY1 down-regulation. Stacked columns show the relative distribution of up-regulated (Up) or down-regulated (Down) miRNAs. (**E**) Summary of the number of validated target mRNAs of the indicated miRNAs identified in (D). (**F**) Functional annotation of miRNA targets shown in Figure 3E using miRNAs Ingenuity Systems Pathway Analysis. Top scoring pathways are highlighted in red.

Such result was significantly different from that expected (chi-squared $P < 0.001$), strongly suggesting that YY1 could play a negative role in these sncRNA biogenesis.

Since most of these sncRNAs are localized within intra-genic regions and are generated by either independent transcription units or through splicing of longer transcripts (55), we believed that our direct bioYY1 target annotation might have missed out most of this RNA species. Thus, we manually annotated transcript position of the 22 sncRNAs found in our expression analysis and established that most of the regulated sncRNAs mapped within longer transcripts or in close proximity of a TSS (Figure 3B). Importantly, more than 60% of sncRNAs presented bioYY1 binding at the promoter of their associated transcripts (Figure 3C). Intriguingly, although sncRNA expression increased upon YY1 depletion, whole transcript expression was unaffected, suggesting a transcriptional independent role of YY1 in regulating sncRNA levels.

To gain further insights for this observation, we decided to look at the expression levels of another class of sncRNAs, mature miRNAs. Using Applied Biosystem TaqMan technology, we measured the expression of 752 mature miRNA species present in the mouse genome in two independent experiments. Such analyses identified 78 miRNAs that were differentially expressed upon YY1 knock down of which more than 70% resulted up-regulated upon YY1 RNAi, further validating our previous observations (Figure 3D and Supplementary Table S4). Functional analysis identified several mRNA transcripts that had been previously validated to be targeted by these miRNAs (summary list shown in Figure 3E and whole list in Supplementary Table S5), highlighting a significant functional enrichment in cell

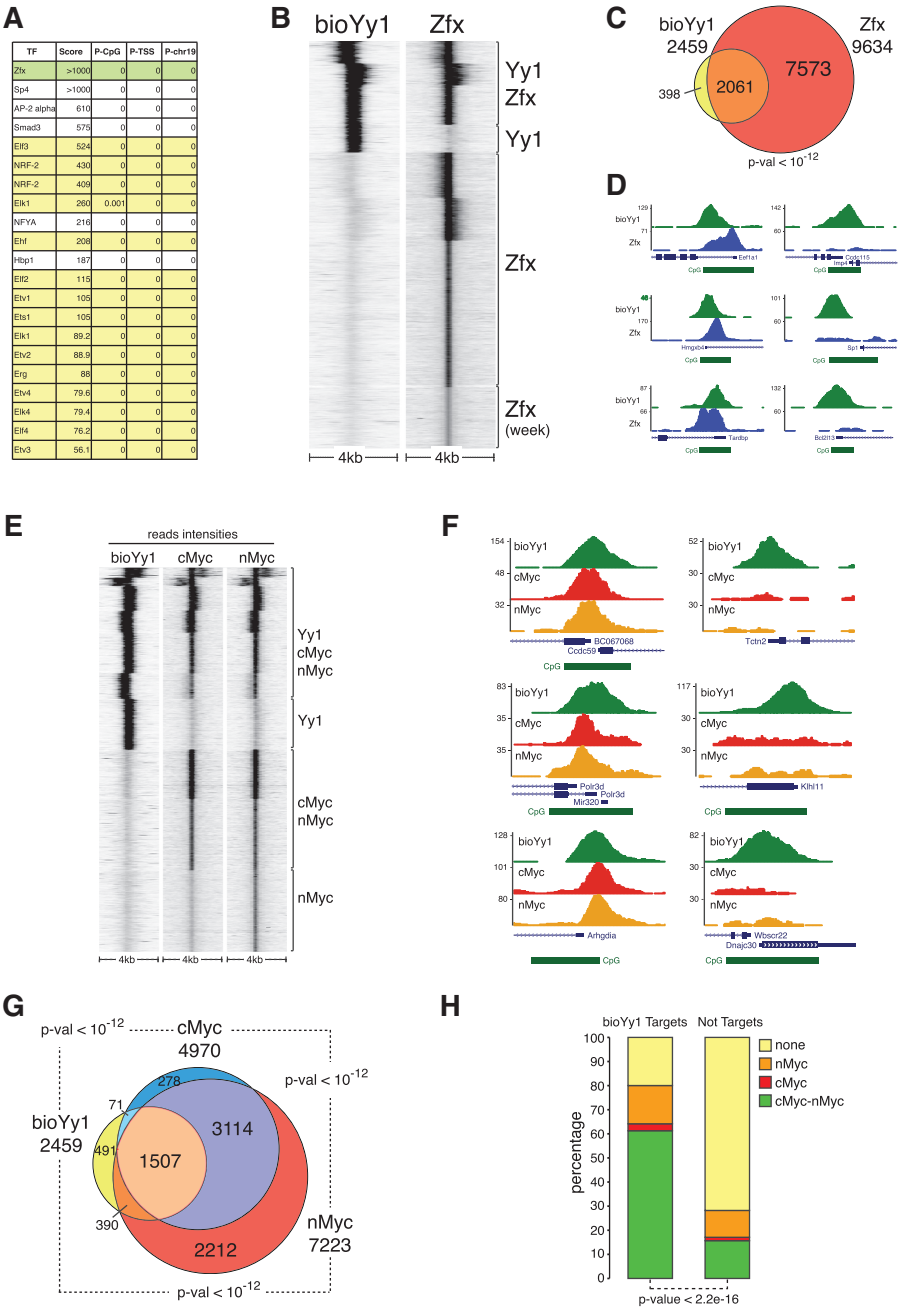
cycle and developmental processes, as well as potential implications with diseases such as cancer and genetic disorders (Figure 3F). Overall, these data show that YY1 has a direct positive action on genes expression but it acts negatively on the accumulation of sncRNAs in ES cells.

These data exclude a functional interaction between YY1 and PcG proteins and put forward YY1 as a positive regulator of gene expression in ES cells. Thus, to gain further insights on YY1 transcriptional activities, we scanned the bioYY1-bound genomic regions looking for the enrichment of known protein–DNA binding motifs. Such analysis identified several DNA elements enriched in proximity of bioYY1 binding sites; in particular, we found an evident over-representation of ETS transcription factors (Figure 4A). In addition, one of the DNA motifs with the highest score in the analysis was the DNA binding site of Zfx (Figure 4A). Zfx is a transcription factor that plays an important role in ES cell self-renewal (56) characterized as part of the Myc module of transcription factors (23,57). To investigate a potential overlap between Zfx and YY1, we analyzed the distribution of bioYY1 and Zfx ChIPseq signals across their binding sites and found that a large part of bioYY1-bound regions were also occupied by Zfx (~70%) (Figure 4B). The density profiles of both transcription factors ChIPseq reads displayed a very similar distribution, demonstrating that regions of association were in proximity with each other and validate the data of the DNA motif analysis (Figure 4A). Analysis of bioYY1 and Zfx direct target genes further confirmed this result, showing a highly significant overlap between bioYY1 and Zfx targets (Figure 4C). Examples of bioYY1 and Zfx binding profiles at target sites are presented in Figure 4D.

Interestingly, two subunits of the INO80 complex that co-purified with bioYY1, Ruvbl1 and Ruvbl2, were also shown to stably interact with a component of the Myc-Max complex, Dmap1 (Figure2 E and F) (23). Moreover, previous reports proposed a connection between YY1 and Myc activities in defined conditions (58–60) that, together with the high degree of overlap between Zfx and bioYY1 genomic association (Figure 4A–C), could suggest a functional overlap between YY1 and Myc functions. To test this, we analyzed the genome-wide binding profile of Myc transcription factors relative to bioYY1 binding. The density profiles of bioYY1 together with c-Myc and n-Myc ChIPseq signals showed a large degree of overlap. Similar to Zfx, both n-Myc and c-Myc bind in proximity of bioYY1 and define different clusters of binding regions either co-occupied by Myc and YY1 or by the proteins alone (Figure 4E). Examples of the binding profiles of these regions are shown in Figure 4F. Similarly for Zfx, analysis of Myc and bioYY1 co-occupancy at target genes revealed a large degree of overlap between bioYY1- and Myc-bound promoters (Figure 4G). Importantly, the largest group of overlapping genes is simultaneously bound by bioYY1, c-Myc and n-Myc (Figure 4H). YY1 and Myc co-occupancy was further validated at endogenous level in both wild-type and BirA-ES cells at several YY1 target genes (Supplementary Figure S4A–C). Consistent with the lack of interaction between Myc and YY1 observed in the MS analyses, loss of YY1 expression did not alter Myc binding from co-occupied promoters (Supplementary Figure S5A).

The discovery of YY1 and Myc co-occupancy induced us to explore the properties of Myc-bound YY1 promoters. We carried out a DNA binding sites prediction of the promoter regions that are co-occupied by bioYY1 and Myc proteins relative to promoters that present the binding of the two transcription factors alone. Such analysis identified three different clusters of significantly enriched DNA binding motifs. Among the DNA binding motifs enriched in the YY1–Myc cluster, together with previously identified Smad and Elk binding sites (Figure 4A), we found that sites for different E2f transcription factors were also over represented in this group (Figure 5A). To extend these findings, we carried out an additional analysis that scanned for DNA binding motifs preferentially associated with promoters co-occupied by Myc and bioYY1 relative to the ones excluded from Myc binding. Such analysis identified, together with the Myc DNA binding site (E-BOX, green box), binding sites for Zfp161, Gmeb1 and different E2f proteins (sky blue boxes) (Figure 5B). In contrast, non-Myc YY1 targets (that represent the smaller fraction of bioYY1 target promoters) were strongly enriched of A-/T-rich DNA binding motifs of which binding sites for homeobox related transcription factors (yellow boxes) were extensively represented (Figure 5B). The specificity of this result is further supported by the preferential association of bioYY1 at CG rich promoters (Figure 2E) and suggests a potential Myc-independent cooperation between homeobox factors and YY1 on a specific set of target genes.

Since E2f factors were previously characterized to be part of the Myc transcription module in ES cells (23,57), since E2f activity was linked to YY1 (61) and since E2f binding sites were always enriched in our DNA motif discovery analyses, we extended our genome-wide analysis to E2f1. Analysis of E2f1 ChIPseq data from ES cells revealed an extensive overlap with bioYY1 and Myc binding sites (Figure 5C). Examples of binding profiles for these data are presented in Figure 5D and validations of endogenous YY1, c-Myc and E2f1 co-occupancy at target sites are shown in Supplementary Figures S4A–C and S5A. All together, these findings strongly suggest an extensive cooperation between YY1 and transcription factors previously characterized within the Myc transcriptional module. An overall analysis of these data is summarized in Figure 6A and shows the existence of different clusters of genomic loci characterized by different combinations of occupancy for these five transcription factors. Examples of the binding profiles at target genes by the different transcription factors are presented in Figure 6B. Importantly, the group co-occupied by all transcription factors represents the largest group of bioYY1 targets (Figure 6C and Supplementary Table S6). Such result is significantly different from expected and supports a transcriptional cooperation between YY1 and Myc related transcription. To test this, we generated a correlation map of the ChIPseq binding profiles between bioYY1, the components of the core pluripotency module (Oct4, Nanog and Sox2) and of Myc module (c-Myc,



TF	Score	P-CpG	P-TSS	P-chr19
Zfx	>1000	0	0	0
Sp4	>1000	0	0	0
AP-2 alpha	610	0	0	0
Smad3	575	0	0	0
Elf3	524	0	0	0
NRF-2	430	0	0	0
NRF-2	409	0	0	0
Elk1	260	0.001	0	0
NFYA	216	0	0	0
Ehf	208	0	0	0
Hbp1	187	0	0	0
Elf2	115	0	0	0
Etv1	105	0	0	0
Ets1	105	0	0	0
Elk1	89.2	0	0	0
Etv2	88.9	0	0	0
Erg	88	0	0	0
Etv4	79.6	0	0	0
Elk4	79.4	0	0	0
Elf4	76.2	0	0	0
Etv3	56.1	0	0	0

Figure 4. YY1 binding sites overlap with Zfx and Myc occupancy. (**A**) Result of Clover analysis on bioYY1 target promoters. Only the motifs with a score >50 are presented in the table. *P*-values of each motif relative to the indicated reference set are presented in addition to the score values. (**B**) K-means clustering of reads intensities in bioYY1 and Zfx ChIPseq data on all bioYY1- and Zfx-associated genomic loci within a 4 kb region centered on peaks' summits. (**C**) Overlap between the target genes of the indicated ChIPseq datasets. Target genes are defined by the presence of at least one ChIPseq peak within ±2 kb from genes annotated TSS. *P*-values of the indicated overlaps are determined by hypergeometric distribution. (**D**) Examples of the Genomic snapshots generated in bioYY1 and Zfx ChIPseq. (**E**) As in (B) using bioYY1, c-Myc- and n-Myc-bound genomic loci. (**F**) Examples of the Genomic snapshots generated in bioYY1, c-Myc and n-Myc ChIPseq. (**G**) As in (C) with bioYY1, c-Myc and n-Myc target genes. (**H**) Distribution of YY1 promoters association relative to c-Myc and n-Myc co-occupancy at target genes. Chi-square test determines the *P*-value.

n-Myc, Zfx and E2f1) (Figure 6D). Such analyses clearly demonstrate that bioYY1 binding has a strong correlation with components of the Myc module but do not correlate in binding with core pluripotency factors, indicating that YY1 is part of the Myc-related transcription network.

In order to correlate transcription factors co-occupancy at YY1-bound promoters with genes transcriptional activity, we compare microarray expression data from ES cells with the different classes of bioYY1 target genes. From this analysis, we concluded that cumulative

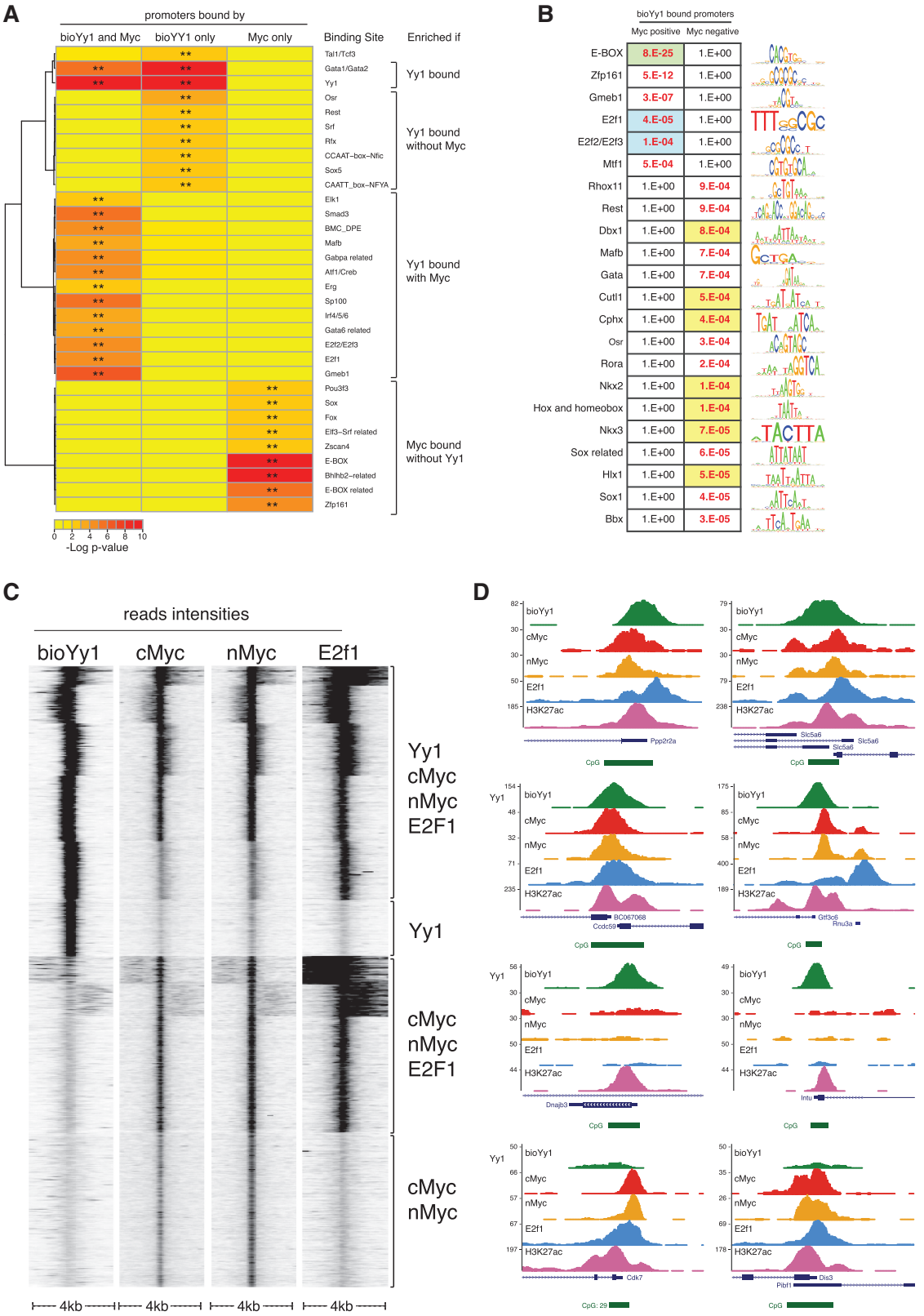


Figure 5. YY1–Myc binding sites are enriched for E2f1 association (**A**) Heat map of the known DNA binding sites enriched in the different clusters of bio YY1- and Myc-bound promoters by Pscan analysis. Double asterisks denote $P < 0.01$. (**B**) As in (A) comparing bioYY1 Myc-bound and not bound target promoters. *P*-values are indicated within the boxes. Myc binding sites are highlighted in green, E2F binding sites in blue, whereas homeobox-related binding sites in yellow. Position-specific weight matrices (PWM) are shown on the left highlighting the CG rich content of Myc related and the A/T rich content of non-Myc YY1 associated PWMs. (**C**) K-means clustering of reads intensities in bioYY1, c-Myc, n-Myc and E2f1 ChIPseq data on all bioYY1 c-Myc ad n-Myc associated genomic loci within a 4-kb region centered on the peaks' summits. (**D**) Examples of the genomic snapshots generated in bioYY1, c-Myc, n-Myc, E2f1 and H3K27ac ChIPseq.

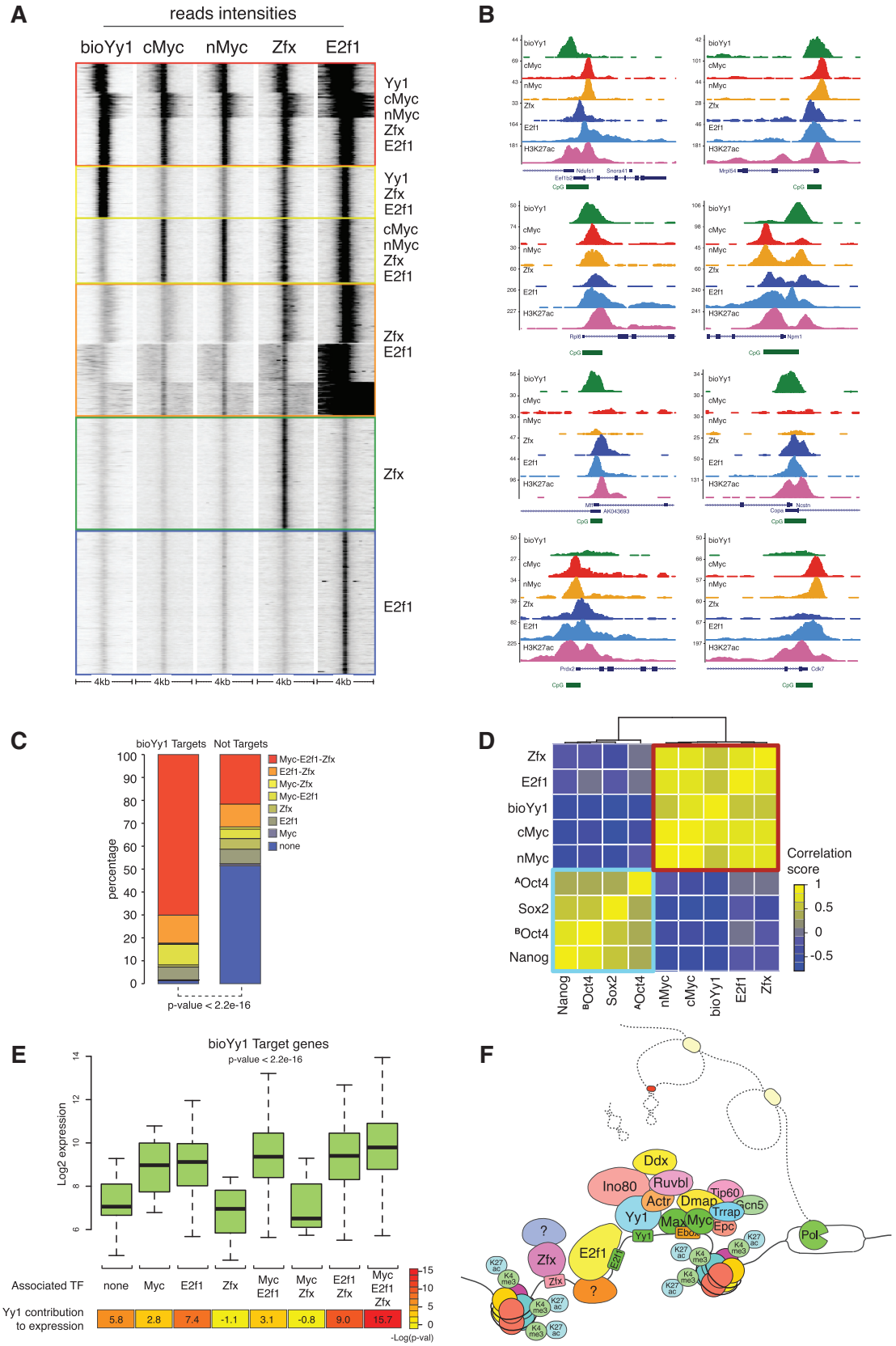


Figure 6. YY1 is part of the Myc TF module and potentiate target genes expression. (**A**) K-means clustering of reads intensities in the indicated ChIPseq data on bioYY1, c-Myc, n-Myc, Zfx and E2f1 associated genomic loci within a 4-kb region centered on each peaks' summit. (**B**) Examples of the genomic snapshots generated in the indicated ChIPseq datasets. (**C**) Distribution of bioYY1 binding relative to c-Myc, n-Myc, Zfx and E2f1 co-occupancy at target genes. Chi-square test determines the *P*-value. (**D**) Heat map of the correlation between the genome-wide associations of the

(continued)

binding of the different transcription factors with YY1 potentiates gene expression (Figure 6E). Interestingly, although Myc and E2f1 binding alone have a positive transcriptional effect, Zfx does not seem to influence gene expression significantly. Nevertheless, cumulative binding of all three classes of transcription factors at bioYY1 target genes correlate with maximal transcriptional activity (Figure 6E, boxplots). A comparative analysis of the same classes of target genes relative to bioYY1 association reveals that the presence of YY1 potentiates gene expression when combined with Myc and or E2f1 and induces maximum expression level when associated simultaneously with all the other transcription factors (Figure 6E, bottom heat map and Supplementary Figure S5B). Simultaneous loss of cMyc and nMyc activity induces loss of ES cells pluripotency and activation of differentiation programs (62). Consistent with this, stable shRNA mediated reduction of YY1 expression in ES cells led to an increased expression of differentiation markers and defects in the activation of some lineage-specific genes upon differentiation in embryoid bodies (EB) (Supplementary Figure S6A and S6B). Importantly, EBs differentiation induced a counter selection for YY1 knock down efficiency (Supplementary Figure S6A) in agreement with a potential role for YY1 in regulating proper ES cell differentiation that is consistent with the severe early implantation defects observed in *YY1* KO embryos. Overall, these data demonstrate a functional role of YY1 in genome-wide Myc transcriptional functions proposing a positive effect on the expression on genes that belong to the Myc transcriptional network.

DISCUSSION

Determination of cellular fate is a complex and still poorly understood process that is controlled at a transcriptional level by the activity of tissue-specific transcription factors (21). ES cells are the cell type with the highest differentiation plasticity (pluripotency) that can be isolated in tissue culture (63). Moreover, several publications have proposed that cancer cells share specific transcriptional features with ES cells (ESC-like signatures) (64,65). This makes the characterization of the transcriptional mechanisms that control ES self-renewal and differentiation important not only to understand their identity and plasticity but also to characterize transcriptional features shared with cancer cells within a physiological context. In line with this, distinct transcriptional modules have been proposed to control ES transcription programs. This involves the repressive PcG module, the Myc module and the core pluripotency factors module (23). Not surprisingly, all components of these networks were shown to play important roles in controlling ES self-renewal and differentiation capabilities (66,67). Moreover, several of these factors are also strongly implicated in tumor development: the best examples are represented by the frequent activation of PcG and Myc functions (25,59).

Here, we have presented a detailed characterization of YY1 functions in mouse ES cells. We show that YY1 does not share physical and functional properties with PcG proteins, while it preferentially associates with the CG-rich promoters of actively transcribed genes that correspond to non-bivalent H3K4me3 enriched and H3K27 hyper-acetylated promoters. This does not exclude that YY1 could play a role in PcG recruitment in particular situations but our data clearly show that this activity, if occurs, must be restricted to very defined circumstances. Consistent with this, like in cancer cells (12,13), YY1 stably associates with components of the INO80 remodeling complex, as well as with newly identified RNA helicase activities. YY1 RNAi experiments further confirm these observations showing that loss of YY1 preferentially leads to a down-regulation of gene expression. Although, only a minority of direct bioYY1 target genes was impaired in expression, this suggests that YY1 binding, together with INO80 and RNA helicase activities, facilitates gene expression in agreement with their co-occupancy at different target promoters. This is in line with the YY1–INO80 co-occupancy observed at *Cdc6* promoter in cancer cells (12). Lack of significant transcriptional changes in the majority of bioYY1 target genes could be due either to the partial knock down that we obtained with both YY1 siRNAs (Figure 2H) or shRNAs treatment (Supplementary Figures S2B, S5A and S6A) or by compensatory effects mediated by the other transcription factors. Moreover, YY1–INO80 interaction has been also proposed to play a role in DNA homologous recombination that occurs in case of DNA damage (13). Due to its ability to associate directly with the DNA cruciform structure of Holliday junctions, YY1 has been proposed to function as a bridge for the recruitment of INO80 remodeling activities at damaged sites (13). Such DNA binding activity does not seem to require YY1 DNA binding motifs in odds with the tight association between the ChIPseq results and the YY1 DNA binding elements (Figure 3D). Nevertheless, the increased genomic instability observed in hypomorphic YY1 KO mouse embryonic fibroblasts supports the hypothesis that YY1 could exert this function (13). However, our knock down studies do not show any evident genomic instability in YY1 depleted ES cells (data not shown), either because of the transient nature of the siRNA-based experiments we performed or because the YY1–DNA repair activities are suppressed by the lack of cell cycle checkpoints and the

Figure 6. Continued
indicated ChIPseq dataset. Two independent Oct4 datasets are included in the analysis. The datasets references are indicated in the 'Materials and Methods' section. (**E**) Box plots of the expression of bioYY1 target genes relative to the co-occupancy of the indicated proteins. *P*-value was determined by Kruskal–Wallis test. Bottom intensity map highlights the significance of the contribution of bioYY1 binding relative to non-YY1 binding to target gene clusters defined by the co-occupancy of the indicated proteins. *P*-values are indicated within the boxes and are determined by Wilcoxon test. (**F**) Model of YY1, Myc, Zfx and E2f1 co-occupancy at target promoters of transcribed genes. The model includes data from previous purifications of Myc complexes from ES cells (23) and speculates on potential transcriptional and post-transcriptional activities of YY1.

high rate of apoptosis present in mouse ES cells (68,69). Finally, the genome-wide combination of expression and localization analyses that we have performed allows strengthening a direct role of YY1 to regulate its target genes expression. Despite a significant portion of transcripts increases expression in absence of YY1 activity, the finding that the most up-regulated transcripts belong to classes of sncRNAs is very surprising and intriguing. It is not clear if YY1 binding plays a role in regulating differential transcriptional units of sncRNAs or if it plays a role in the biogenesis of these mature RNAs. The direct binding of YY1 to the promoters of most of these sncRNAs associated mRNAs might suggest that either YY1 binding suppresses the activation of alternative sncRNA transcription units or that YY1, together with the INO80 remodeling and the RNA helicases activities, can have a negative effect on sncRNA processing, hence altering their stability. It is also possible that sncRNAs up-regulation is an indirect effect of a stress response caused by YY1 down-regulation. However, the direct binding of YY1 in proximity of sncRNA genomic loci could support a direct role of YY1 in regulating their cellular levels.

Both our finding that YY1 do not share functional properties with PcG activities and that instead share a global regulatory functions with the Myc-related transcriptional network are in perfect line with other observations. YY1 was shown to possess the ability, in non-physiological conditions, to interact with Myc (58,59). The lack of Myc protein in our YY1 purification, as well as in others performed in cancer cells (12,13) clearly shows that, if Myc interacts physiologically with YY1, such binding is not stable. Nevertheless, this could be important to stabilize the genome-wide DNA mediated co-occupancy of YY1 and Myc at their shared binding sites. Similar to this, E2f1 was also shown to interact *in vitro* with YY1 (61) but, like Myc, was not found in the YY1 purifications. In addition, YY1 was also shown to directly activate expression from the Myc promoter and to increase, when overexpressed, Myc endogenous transcripts (70,71). In ES cells, we do not observe binding of YY1 at any Myc paralog promoters neither we detect significant expression changes in Myc expression (Supplementary Table S2) suggesting that this regulatory mechanism is not conserved in ES cells.

Our data on the transcriptional cooperation between YY1 and the Myc module of transcription factors are in line with previous observations. For example, YY1 was shown to cooperate with Myc in activating *Surf-1* expression (60) while it was shown to act synergistically with E2F1 in activating the *Cdc6* promoter (61). The cMyc and nMyc KO mice are both early embryonic lethal (72–74) (E10.5 and E11.5, respectively) (72–74) and double Myc KO ES cells activate differentiation genes expression and loose their pluripotency (62). Similarly, reduction of YY1 expression induces the activation of differentiation markers and genes related to embryonic development. Moreover, differentiation of these cells induced a strong counter selection for YY1 knock down efficiency that is suggestive of a potential role for YY1 in differentiation consistent with its essential role early implantation development. Generation of inducible YY1 KO ES cells will therefore be an invaluable tool to extend these observations. Importantly, in addition to the essential function in regulating normal development and differentiation, YY1 and Myc are both activated in NDEA-induced hepatocarcinogenesis (75), as well as in Burkitt Lymphomas (76) and in prostate cancers (77). Interestingly, YY1 and Zfx were identified out of six proteins that scored in a proteomic study aimed to identify proteins expressed in neoplastic nodes of diffuse large B cell and Follicular lymphomas (78), two tumor types that are frequently driven by *Myc* amplifications and overexpression (24). Altogether, our data on the global transcriptional cooperation between YY1 and the Myc transcription network (summarized in Figure 6F) makes a big stem in putting together little pieces of observations disperse over several years of literature and identify a novel component of a keystone transcriptional regulatory network of normal pluripotent and cancer cells.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–7 and Supplementary Figures 1–6.

ACKNOWLEDGEMENTS

We would like to thank Dies Meier for providing the BirA-ES cells, Arianna Sabò and Karin Ferrari for providing reagents and Bruno Amati for the Ruvbl2 antisera, comments and discussion. We thank Gioacchino Natoli and Saverio Minucci for their comments. We also acknowledge Gabriele Bucci, Davide Cittaro, Simone Minardi e Alberto Termanini for their computational support. We thank all members of the Pasini's laboratory for discussion.

FUNDING

The Italian Association of Cancer Research (AIRC) (to D.P.); Italian Ministry of Health (to D.P.); Fellowship of the Italian Cancer Research Foundation (FIRC) (to I.B.); Giovanni Armenise-Harvard Foundation Career Development Program (to T.B.); Association of International Cancer Research (to T.B.); Italian Association for Cancer Research and Cariplo Foundation (to T.B.). Funding for the open access charge: AIRC, Italian Association for Cancer Research and Italian Ministry of Health.

Conflict of interest statement. None declared.

REFERENCES

1. Shi,Y., Seto,E., Chang,L.S. and Shenk,T. (1991) Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell*, **67**, 377–388.
2. Brown,J.L., Mucci,D., Whiteley,M., Dirksen,M.L. and Kassis,J.A. (1998) The *Drosophila* Polycomb group gene pleiohomeotic

encodes a DNA binding protein with homology to the transcription factor YY1. *Mol. Cell*, **1**, 1057–1064.
3. Bracken,A.P. and Helin,K. (2009) Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nat. Rev. Cancer*, **9**, 773–784.
4. Donohoe,M.E., Zhang,X., McGinnis,L., Biggers,J., Li,E. and Shi,Y. (1999) Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol. Cell Biol.*, **19**, 7237–7244.
5. He,Y., Dupree,J., Wang,J., Sandoval,J., Li,J., Liu,H., Shi,Y., Nave,K.A. and Casaccia-Bonnefil,P. (2007) The transcription factor Yin Yang 1 is essential for oligodendrocyte progenitor differentiation. *Neuron*, **55**, 217–230.
6. Gordon,S., Akopyan,G., Garban,H. and Bonavida,B. (2006) Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene*, **25**, 1125–1142.
7. Satijn,D.P., Hamer,K.M., den Blaauwen,J. and Otte,A.P. (2001) The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in Xenopus embryos. *Mol. Cell Biol.*, **21**, 1360–1369.
8. Caretti,G., Di Padova,M., Micales,B., Lyons,G.E. and Sartorelli,V. (2004) The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes Dev.*, **18**, 2627–2638.
9. Palacios,D., Mozzetta,C., Consalvi,S., Caretti,G., Saccone,V., Proserpio,V., Marquez,V.E., Valente,S., Mai,A., Forcales,S.V. *et al.* (2011) TNF/p38alpha/polycomb signaling to Pax7 locus in satellite cells links inflammation to the epigenetic control of muscle regeneration. *Cell Stem Cell*, **7**, 455–469.
10. Woo,C.J., Kharchenko,P.V., Daheron,L., Park,P.J. and Kingston,R.E. (2010) A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell*, **140**, 99–110.
11. Jeon,Y. and Lee,J.T. (2011) YY1 tethers Xist RNA to the inactive X nucleation center. *Cell*, **146**, 119–133.
12. Cai,Y., Jin,J., Yao,T., Gottschalk,A.J., Swanson,S.K., Wu,S., Shi,Y., Washburn,M.P., Florens,L., Conaway,R.C. *et al.* (2007) YY1 functions with INO80 to activate transcription. *Nat. Struct. Mol. Biol.*, **14**, 872–874.
13. Wu,S., Shi,Y., Mulligan,P., Gay,F., Landry,J., Liu,H., Lu,J., Qi,H.H., Wang,W., Nickoloff,J.A. *et al.* (2007) A YY1-INO80 complex regulates genomic stability through homologous recombination-based repair. *Nat. Struct. Mol. Biol.*, **14**, 1165–1172.
14. He,Y., Kim,J.Y., Dupree,J., Tewari,A., Melendez-Vasquez,C., Svaren,J. and Casaccia,P. (2011) YY1 as a molecular link between neuregulin and transcriptional modulation of peripheral myelination. *Nat. Neurosci.*, **13**, 1472–1480.
15. Sui,G., Affar el,B., Shi,Y., Brignone,C., Wall,N.R., Yin,P., Donohoe,M., Luke,M.P., Calvo,D. and Grossman,S.R. (2004) Yin Yang 1 is a negative regulator of p53. *Cell*, **117**, 859–872.
16. Faust,C., Lawson,K.A., Schork,N.J., Thiel,B. and Magnuson,T. (1998) The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development*, **125**, 4495–4506.
17. O'Carroll,D., Erhardt,S., Pagani,M., Barton,S.C., Surani,M.A. and Jenuwein,T. (2001) The polycomb-group gene Ezh2 is required for early mouse development. *Mol. Cell Biol.*, **21**, 4330–4336.
18. Pasini,D., Bracken,A.P., Jensen,M.R., Denchi,E.L. and Helin,K. (2004) Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.*, **23**, 4061–4071.
19. Voncken,J.W., Roelen,B.A., Roefs,M., de Vries,S., Verhoeven,E., Marino,S., Deschamps,J. and van Lohuizen,M. (2003) Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc. Natl Acad. Sci. USA*, **100**, 2468–2473.
20. Wray,J., Kalkan,T. and Smith,A.G. (2010) The ground state of pluripotency. *Biochem. Soc. Trans.*, **38**, 1027–1032.
21. Ng,H.H. and Surani,M.A. (2011) The transcriptional and signalling networks of pluripotency. *Nat. Cell Biol.*, **13**, 490–496.
22. Stadtfeld,M. and Hochedlinger,K. (2010) Induced pluripotency: history, mechanisms, and applications. *Genes Dev.*, **24**, 2239–2263.
23. Kim,J., Woo,A.J., Chu,J., Snow,J.W., Fujiwara,Y., Kim,C.G., Cantor,A.B. and Orkin,S.H. (2010) A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell*, **143**, 313–324.
24. Nesbit,C.E., Tersak,J.M. and Prochownik,E.V. (1999) MYC oncogenes and human neoplastic disease. *Oncogene*, **18**, 3004–3016.
25. Piunti,A. and Pasini,D. (2011) Epigenetic factors in cancer development: polycomb group proteins. *Future Oncol.*, **7**, 57–75.
26. Driegen,S., Ferreira,R., van Zon,A., Strouboulis,J., Jaegle,M., Grosveld,F., Philipsen,S. and Meijer,D. (2005) A generic tool for biotinylation of tagged proteins in transgenic mice. *Transgenic Res.*, **14**, 477–482.
27. Frank,S.R., Parisi,T., Taubert,S., Fernandez,P., Fuchs,M., Chan,H.M., Livingston,D.M. and Amati,B. (2003) MYC recruits the TIP60 histone acetyltransferase complex to chromatin. *EMBO Rep.*, **4**, 575–580.
28. Rappsilber,J., Mann,M. and Ishihama,Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*, **2**, 1896–1906.
29. Olsen,J.V., de Godoy,L.M., Li,G., Macek,B., Mortensen,P., Pesch,R., Makarov,A., Lange,O., Horning,S. and Mann,M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell Proteomics*, **4**, 2010–2021.
30. Kall,L., Storey,J.D., MacCoss,M.J. and Noble,W.S. (2008) Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J. Proteome Res.*, **7**, 29–34.
31. Elias,J.E., Haas,W., Faherty,B.K. and Gygi,S.P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods*, **2**, 667–675.
32. Pasini,D., Cloos,P.A., Walfridsson,J., Olsson,L., Bukowski,J.P., Johansen,J.V., Bak,M., Tommerup,N., Rappsilber,J. and Helin,K. (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature*, **464**, 306–310.
33. Kim,J., Cantor,A.B., Orkin,S.H. and Wang,J. (2009) Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nat. Protoc.*, **4**, 506–517.
34. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
35. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
36. Fujita,P.A., Rhead,B., Zweig,A.S., Hinrichs,A.S., Karolchik,D., Cline,M.S., Goldman,M., Barber,G.P., Clawson,H., Coelho,A. *et al.* (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.*, **39**, D876–D882.
37. Cesaroni,M., Cittaro,D., Brozzi,A., Pelicci,P.G. and Luzi,L. (2008) CARPET: a web-based package for the analysis of ChIP-chip and expression tiling data. *Bioinformatics*, **24**, 2918–2920.
38. Bailey,T.L. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28–36.
39. Jolma,A., Kivioja,T., Toivonen,J., Cheng,L., Wei,G., Enge,M., Taipale,M., Vaquerizas,J.M., Yan,J., Sillanpaa,M.J. *et al.* (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res.*, **20**, 861–873.
40. Portales-Casamar,E., Thongjuea,S., Kwon,A.T., Arenillas,D., Zhao,X., Valen,E., Yusuf,D., Lenhard,B., Wasserman,W.W. and Sandelin,A. (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **38**, D105–D110.
41. Bucher,P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.*, **212**, 563–578.
42. Berger,M.F., Badis,G., Gehrke,A.R., Talukder,S., Philippakis,A.A., Pena-Castillo,L., Alleyne,T.M., Mnaimneh,S., Botvinnik,O.B., Chan,E.T. *et al.* (2008) Variation in

homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*, **133**, 1266–1276.
43. Badis,G., Berger,M.F., Philippakis,A.A., Talukder,S., Gehrke,A.R., Jaeger,S.A., Chan,E.T., Metzler,G., Vedenko,A., Chen,X. *et al.* (2009) Diversity and complexity in DNA recognition by transcription factors. *Science*, **324**, 1720–1723.
44. Habib,N., Kaplan,T., Margalit,H. and Friedman,N. (2008) A novel Bayesian DNA motif comparison method for clustering and retrieval. *PLoS Comput. Biol.*, **4**, e1000010.
45. Frith,M.C., Fu,Y., Yu,L., Chen,J.F., Hansen,U. and Weng,Z. (2004) Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res.*, **32**, 1372–1381.
46. Zambelli,F., Pesole,G. and Pavesi,G. (2009) Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res.*, **37**, W247–W252.
47. Ye,T., Krebs,A.R., Choukrallah,M.A., Keime,C., Plewniak,F., Davidson,I. and Tora,L. (2010) seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.*, **39**, e35.
48. Pasini,D., Hansen,K.H., Christensen,J., Agger,K., Cloos,P.A. and Helin,K. (2008) Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev.*, **22**, 1345–1355.
49. de Boer,E., Rodriguez,P., Bonte,E., Krijgsveld,J., Katsantoni,E., Heck,A., Grosveld,F. and Strouboulis,J. (2003) Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl Acad. Sci. USA*, **100**, 7480–7485.
50. Wang,J., Rao,S., Chu,J., Shen,X., Levasseur,D.N., Theunissen,T.W. and Orkin,S.H. (2006) A protein interaction network for pluripotency of embryonic stem cells. *Nature*, **444**, 364–368.
51. Kim,J., Chu,J., Shen,X., Wang,J. and Orkin,S.H. (2008) An extended transcriptional network for pluripotency of embryonic stem cells. *Cell*, **132**, 1049–1061.
52. Bernstein,B.E., Mikkelsen,T.S., Xie,X., Kamal,M., Huebert,D.J., Cuff,J., Fry,B., Meissner,A., Wernig,M., Plath,K. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315–326.
53. Pasini,D., Malatesta,M., Jung,H.R., Walfridsson,J., Willer,A., Olsson,L., Skotte,J., Wutz,A., Porse,B., Jensen,O.N. *et al.* (2010) Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res.*, **38**, 4958–4969.
54. Creyghton,M.P., Cheng,A.W., Welstead,G.G., Kooistra,T., Carey,B.W., Steine,E.J., Hanna,J., Lodato,M.A., Frampton,G.M., Sharp,P.A. *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA*, **107**, 21931–21936.
55. Mattick,J.S. and Makunin,I.V. (2006) Non-coding RNA. *Hum. Mol. Genet.*, **15 (Spec No 1)**, R17–R29.
56. Galan-Caridad,J.M., Harel,S., Arenzana,T.L., Hou,Z.E., Doetsch,F.K., Mirny,L.A. and Reizis,B. (2007) Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell*, **129**, 345–357.
57. Chen,X., Xu,H., Yuan,P., Fang,F., Huss,M., Vega,V.B., Wong,E., Orlov,Y.L., Zhang,W., Jiang,J. *et al.* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, **133**, 1106–1117.
58. Shrivastava,A., Saleque,S., Kalpana,G.V., Artandi,S., Goff,S.P. and Calame,K. (1993) Inhibition of transcriptional regulator Yin-Yang-1 by association with c-Myc. *Science*, **262**, 1889–1892.
59. Shrivastava,A., Yu,J., Artandi,S. and Calame,K. (1996) YY1 and c-Myc associate in vivo in a manner that depends on c-Myc levels. *Proc. Natl Acad. Sci. USA*, **93**, 10638–10641.
60. Vernon,E.G. and Gaston,K. (2000) Myc and YY1 mediate activation of the Surf-1 promoter in response to serum growth factors. *Biochim. Biophys. Acta*, **1492**, 172–179.
61. Schlisio,S., Halperin,T., Vidal,M. and Nevins,J.R. (2002) Interaction of YY1 with E2Fs, mediated by RYBP, provides a mechanism for specificity of E2F function. *EMBO J.*, **21**, 5775–5786.
62. Varlakhanova,N.V., Cotterman,R.F., deVries,W.N., Morgan,J., Donahue,L.R., Murray,S., Knowles,B.B. and Knoepfler,P.S. (2010) myc maintains embryonic stem cell pluripotency and self-renewal. *Differentiation*, **80**, 9–19.
63. Blair,K., Wray,J. and Smith,A. (2011) The liberation of embryonic stem cells. *PLoS Genet.*, **7**, e1002019.
64. Ben-Porath,I., Thomson,M.W., Carey,V.J., Ge,R., Bell,G.W., Regev,A. and Weinberg,R.A. (2008) An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.*, **40**, 499–507.
65. Wong,D.J., Liu,H., Ridky,T.W., Cassarino,D., Segal,E. and Chang,H.Y. (2008) Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell*, **2**, 333–344.
66. Christophersen,N.S. and Helin,K. (2010) Epigenetic control of embryonic stem cell fate. *J. Exp. Med.*, **207**, 2287–2295.
67. Orkin,S.H. and Hochedlinger,K. (2011) Chromatin connections to pluripotency and cellular reprogramming. *Cell*, **145**, 835–850.
68. Corbet,S.W., Clarke,A.R., Gledhill,S. and Wyllie,A.H. (1999) P53-dependent and -independent links between DNA-damage, apoptosis and mutation frequency in ES cells. *Oncogene*, **18**, 1537–1544.
69. Heyer,B.S., MacAuley,A., Behrendtsen,O. and Werb,Z. (2000) Hypersensitivity to DNA damage leads to increased apoptosis during early mouse development. *Genes Dev.*, **14**, 2072–2084.
70. Riggs,K.J., Saleque,S., Wong,K.K., Merrell,K.T., Lee,J.S., Shi,Y. and Calame,K. (1993) Yin-yang 1 activates the c-myc promoter. *Mol. Cell Biol.*, **13**, 7487–7495.
71. Lee,T.C., Zhang,Y. and Schwartz,R.J. (1994) Bifunctional transcriptional properties of YY1 in regulating muscle actin and c-myc gene expression during myogenesis. *Oncogene*, **9**, 1047–1052.
72. Davis,A.C., Wims,M., Spotts,G.D., Hann,S.R. and Bradley,A. (1993) A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes Dev.*, **7**, 671–682.
73. Charron,J., Malynn,B.A., Fisher,P., Stewart,V., Jeannotte,L., Goff,S.P., Robertson,E.J. and Alt,F.W. (1992) Embryonic lethality in mice homozygous for a targeted disruption of the N-myc gene. *Genes Dev.*, **6**, 2248–2257.
74. Stanton,B.R., Perkins,A.S., Tessarollo,L., Sassoon,D.A. and Parada,L.F. (1992) Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. *Genes Dev.*, **6**, 2235–2247.
75. Parija,T. and Das,B.R. (2003) Involvement of YY1 and its correlation with c-myc in NDEA induced hepatocarcinogenesis, its prevention by d-limonene. *Mol. Biol. Rep.*, **30**, 41–46.
76. Hu,H.M., Kanda,K., Zhang,L. and Boxer,L.M. (2007) Activation of the c-myc p1 promoter in Burkitt's lymphoma by the hs3 immunoglobulin heavy-chain gene enhancer. *Leukemia*, **21**, 747–753.
77. Sun,C., Song,C., Ma,Z., Xu,K., Zhang,Y., Jin,H., Tong,S., Ding,W., Xia,G. and Ding,Q. (2011) Periostin identified as a potential biomarker of prostate cancer by iTRAQ-proteomics analysis of prostate biopsy. *Proteome Sci.*, **9**, 22.
78. Sakhinia,E., Glennie,C., Hoyland,J.A., Menasce,L.P., Brady,G., Miller,C., Radford,J.A. and Byers,R.J. (2007) Clinical quantitation of diagnostic and predictive gene expression levels in follicular and diffuse large B-cell lymphoma by RT-PCR gene expression profiling. *Blood*, **109**, 3922–3928.