nature structural & molecular biology

Multilayered chromatin analysis reveals E2f, Smad and Zfx as transcriptional regulators of histones

David Gokhman, Ilana Livyatan, Badi Sri Sailaja, Shai Melcer & Eran Meshorer

Histones, the building blocks of eukaryotic chromatin, are essential for genome packaging, function and regulation. However, little is known about their transcriptional regulation. Here we conducted a comprehensive computational analysis, based on chromatin immunoprecipitation–sequencing and –microarray analysis (ChIP-seq and ChIP-chip) data of over 50 transcription factors and histone modifications in mouse embryonic stem cells. Enrichment scores supported by gene expression data from gene knockout studies identified E2f1 and E2f4 as master regulators of histone genes, CTCF and Zfx as repressors of core and linker histones, respectively, and Smad1, Smad2, YY1 and Ep300 as restricted or cell type–specific regulators. We propose that histone gene regulation is substantially more complex than previously thought, and that a combination of factors orchestrate histone gene regulation, from strict synchronization with S phase to targeted regulation of specific histone subtypes.

Histones are encoded by more than 80 genes in mouse and humans. Histone proteins are essential for DNA packaging but can be toxic to cells at high levels. This fact necessitates tight regulation of histone biogenesis with DNA synthesis during S phase. Disruption of this coupling can result in loss of chromosomes, developmental arrest and DNA damage[1–4]. To meet the high demands for new histones during S phase, the process of histone biogenesis must be faster and more efficient than that of most other proteins. This is achieved through several bypasses of conventional expression pathways. The first is through extensive clustering of genes; more than 80% of the replication-dependent histone genes in the mouse are located in two main loci on chromosomes 13 and 3 (ref. 4). This clustering has persisted throughout evolution, is present in organisms from yeast to human and is presumed to facilitate common transcriptional regulation[5]. Additionally, histone genes lack introns and their mRNAs are non-polyadenylated, avoiding the time-consuming processes of splicing and polyadenylation[6]. Apart from rapid production, the process of histone synthesis must be tightly synchronized with S phase when histone gene transcription is elevated 3–5-fold[7,8] compared to other cell cycle stages. This synchronization strongly implies the involvement of DNA-binding transcription factors (TFs) that are part of the general regulation system in S phase.

To date, several factors, including OTF1, HiNF, YY1, AP1, AP2, H1TF1, HITF2, HIRA, BZAP45, TBP and TTF2, have been associated with histone gene regulation. However, none of these have been shown to regulate all subtypes of histone genes and coordinate their transcription with S phase. OTF1 was shown to bind to histone H2b genes in HeLa cells[9,10]; HiNF has been found to associate with core histone genes, mostly histones H2b and H4 in human cells[11]; NPAT, although does not bind DNA directly, is essential for histone gene transcription[12]; YY1, AP1 and AP2 have been shown to regulate the histone H3.2 variant in hamster fibroblasts[13–15], and YY1 also has been shown to interact with histone H4 gene promoters in HeLa cells[16]; H1TF1 and H1TF2 bind histone H1 promoters in human[17,18]; HIRA has been shown to repress expression of histones H1, H2a, H2b and H3 in human[19]; BZAP45 stimulates histone H4 gene transcription in human[20], and TTF2 and TBP have been shown to target linker and core histones genes, respectively, in *Drosophila melanogaster*[21].

In addition, several histone acetyltransferases have been shown to regulate acetylation of histone proteins bound to histone gene promoters, thereby activating their expression. This group includes CBP, Ep300 (also known as p300) and TIP60 that regulate histone H2b and H4 promoters through an NPAT-dependent pathway in human cells[22,23].

Despite the discovery of these factors, how they all cooperate to regulate transcription of histone genes during the cell cycle remains largely unknown[4]. In addition, most of these studies were conducted before the high-throughput sequencing era, centering mostly on a handful of histone genes, with a strong bias toward histone H4. This made it difficult to determine the extent of regulation by each TF, and to identify master and subtype-specific or cell type–specific regulators of histone gene expression.

E2f, which is a major S-phase regulator[24], is a good candidate factor for synchronizing histone transcription with S phase. However, it has only been linked to a few histone genes so far[25,26] through analysis of consensus motifs and binding assays, but no comprehensive study has been performed to determine whether it serves as a bona fide histone gene S-phase regulator.

In addition to strict S-phase regulation, histones exhibit differential expression patterns that cannot be explained by general S-phase

Department of Genetics, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. Correspondence should be addressed to E.M. (meshorer@huji.ac.il).

Received 3 June; accepted 15 October; published online 9 December 2012; doi:10.1038/nsmb.2448

regulators. Different expression of histone genes has been observed between cell types, between tissues[27], during development[28] and even between individual histone genes of the same subtype[29]. Such differing expression must be explained through regulators that act differently between cell types and/or through subtype-specific regulation. Histone H1–specific regulation is of special interest, as the stoichiometry of histone H1 to core histones differs between cell types[30]. This suggests the existence of a TF that is both specific to histone H1 gene and is developmental stage–associated or cell type–associated. However, none of the known factors have been shown to possess both of these traits. Moreover, previous studies showed activity of a TF in a specific cell type. Therefore, activity in different cell types remains undetermined.

Mouse embryonic stem cells (ESCs) are an excellent platform to study histone gene transcriptional regulation because they have very short G1 and G2 cell cycle phases and thus relative to all other cell types, mouse ESCs spend the highest fraction of the cell cycle in S phase. It is estimated that up to 60% of cells in a given ESC population are in S phase at any given moment[31]. In addition, mouse ESCs have been extensively characterized for genome-wide binding of chromatin proteins and histone modifications. These experiments lay the foundations for a comprehensive and unbiased genome-wide analysis.

Here we sought to shed light on three of the main questions regarding histone gene regulation: (i) what are the master regulators of S phase–coupled histone genes, (ii) which factors are responsible for the differential regulation of core and linker histones and (iii) which factors regulate cell type–specific or tissue type–specific expression?

Using a multilayered computational analysis, we identified nine transcriptional regulators that bind and regulate histone genes in ESCs. We propose that E2f1 and E2f4 are master regulators of histone genes, that CTCF and Zfx are repressors of core and linker histones, respectively, and that Smad1, Smad2, YY1 and Ep300 are restricted or cell type–specific regulators. Taken together, our data suggest a complex multilayered regulation of histone gene expression in mammalian cells that is likely crucial for their function as the basic building blocks of mammalian chromatin.

RESULTS

E2f1 and E2f4 target and regulate histone genes

To investigate the transcription factors that are involved in the regulation of histone genes, we analyzed data from previous ChIP-seq and ChIP-chip experiments of over 50 ESC-associated DNA-binding factors and histone modifications (**Fig. 1** and **Supplementary Table 1**). We determined TF binding (either direct or indirect) by locating peaks, via the MACS algorithm[32], that were located within 1 kilobase (kb) upstream to 1 kb downstream of transcription start sites (TSSs). Read-density maps for several TFs are shown in **Figure 1**. Among the transcriptional regulators tested, we found a strong, significant enrichment (hypergeometric test) for Smad1, Smad2, p300, E2f1, E2f4, Med1, Med12, Smc1, Smc3, Nipbl and Tet1 in histone genes in ESCs, and a strong enrichment for YY1 in ESC-derived neuronal progenitor cells (NPCs; **Fig. 2**).

Smad1, p300, Med1 and Smad2 were all enriched 15-fold or higher above a random distribution; Med12 was enriched over sevenfold; E2f4 was enriched 3.9-fold; Smc3, E2f1, Nipbl, Tet1 and Smc1 were all enriched ~2-fold; and YY1 was enriched 2.6-fold in NPCs but notably was completely absent in histone gene promoters in ESCs. We validated the enrichment of E2f1, Smad1 and Zfx (which was exclusively enriched on linker histone genes, see below) on selected histone gene promoters in ESCs (**Supplementary Fig. 1** and **Supplementary Table 2**).

Previous studies have described the regulation of only a few histone genes by E2f1, namely *Hist1h3f, Hist1h1d* and a few histone H2a genes[25,26]. These studies have also identified the E2f binding motifs close to a few other histone genes, suggesting a potential regulation by E2f proteins in a limited number of histone genes, while the rest were defined as "E2f-independent"[20,33]. Nevertheless, in our study, we found 86% (55/64) of the replication-dependent histone genes within 1 kb of an E2f1 peak. In accordance with targeting of E2f to promoters[34], the majority of these peaks (>90%) were precisely on the TSS. E2f4 binding to histone genes was even more substantial: 92.2% (59/64) of the histone genes had an E2f4 peak within 1 kb of their TSS (with 54 of these exhibiting a peak precisely on the TSS; **Fig. 3**). Union of the two groups reveals that 100% of the replication-dependent histones (64/64) have either an E2f1 or an E2f4 peak precisely on their TSS. We also detected significant enrichment of E2f4 at histone genes in five additional tissues, namely adipocyte, liver, myoblast, B-cell lymphoma and erythroleukemia cells ($P < 10^{-15}$ for B cell, myoblast and leukemia, $P = 1.3 \times 10^{-3}$ for adipocyte, $P = 1.7 \times 10^{-4}$ for liver; **Supplementary Fig. 2**). This suggests general rather than tissue-specific regulation. As all these studies were conducted on unsynchronized populations of cells, it is possible that E2f1 and E2f4 bind histone genes at different stages of the cell cycle. Despite the vast evidence of E2f1 and E2f4 binding

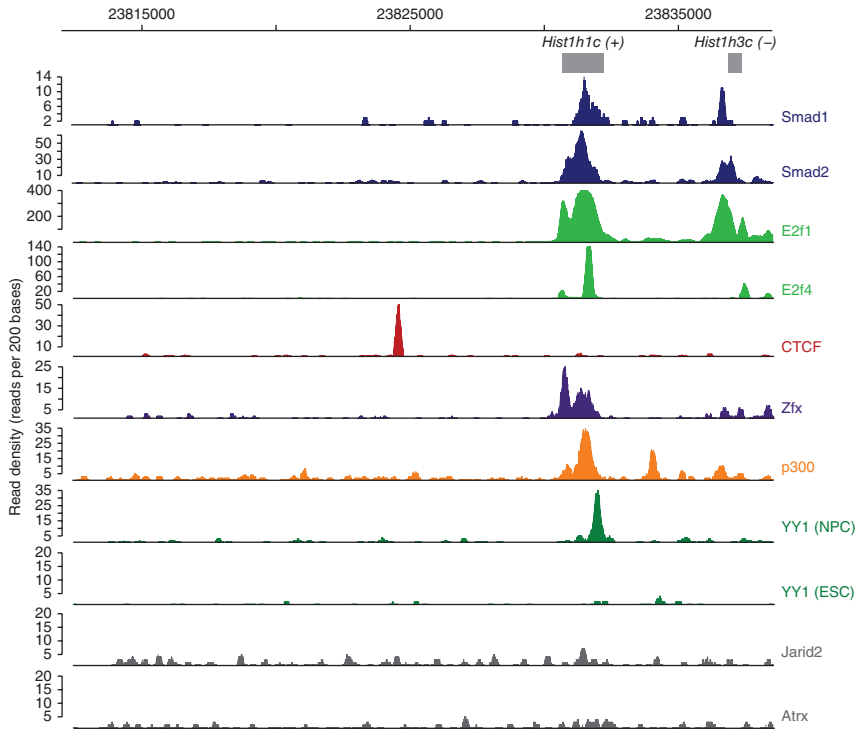

Figure 1 Multifactorial regulation of histone genes. Read density maps of two histone genes in chromosome 13: 23812000–23838500. Jarid2 and Atrx are shown as two (of many) examples where no peaks are detected in the vicinity of histone genes.

to promoters of histone genes, the majority of histone genes lack a classic E2f binding consensus motif[35], which coincides with the fact that most E2f-bound regions *in vivo* do not contain an E2f binding motif[34]. Read density of E2f1 ChIP-seq data in MCF7 breast cancer cells showed similar targeting by E2F1 around promoters of histone genes in human cells[36] (**Fig. 3c**). Furthermore, experimental work on *Arabidopsis thaliana* revealed that at least *CENH3*, a histone H3 variant, is regulated by E2f[37]. This suggests a deep evolutionary conservation of histone regulation by E2f proteins.

To our surprise, histone H4 genes, which are regulated through the HiNF signaling pathway and have been considered to be E2f-independent[20,33], also exhibited a high level of E2f binding (**Fig. 3**), with ~92% (12/13) of all histone H4 genes bound by E2f1 and 100% (13/13) bound by E2f4. HiNF alone has been shown to be insufficient for maximal transcription of a histone H4 gene[38], suggesting that E2f binding may be necessary for complete histone H4 activation.

To investigate the regulatory role of E2f4 in histone genes, we analyzed histone gene expression in wild-type versus *E2f4*$^{-/-}$ mouse embryonic fibroblasts (MEFs)[39]. The analysis revealed a decrease in expression in 17 of 19 relevant probe sets in the *E2f4*$^{-/-}$ cells. The decrease averaged 20% and reached ~50% in some genes (**Fig. 3d**). These results, although predicted by our analyses, are somewhat surprising, considering the compensatory nature of E2f proteins[40]. This decrease in expression is probably not the result of a change in cell cycle, as *E2f4* knockout in MEFs has been shown to have no effect on proliferation or re-entry into cell cycle[41].

Taken together, these data suggest that E2f1 and E2f4 may serve as master regulators of the histone multigene family synchronizing its expression with S phase and coordinating the expression of its individual members.

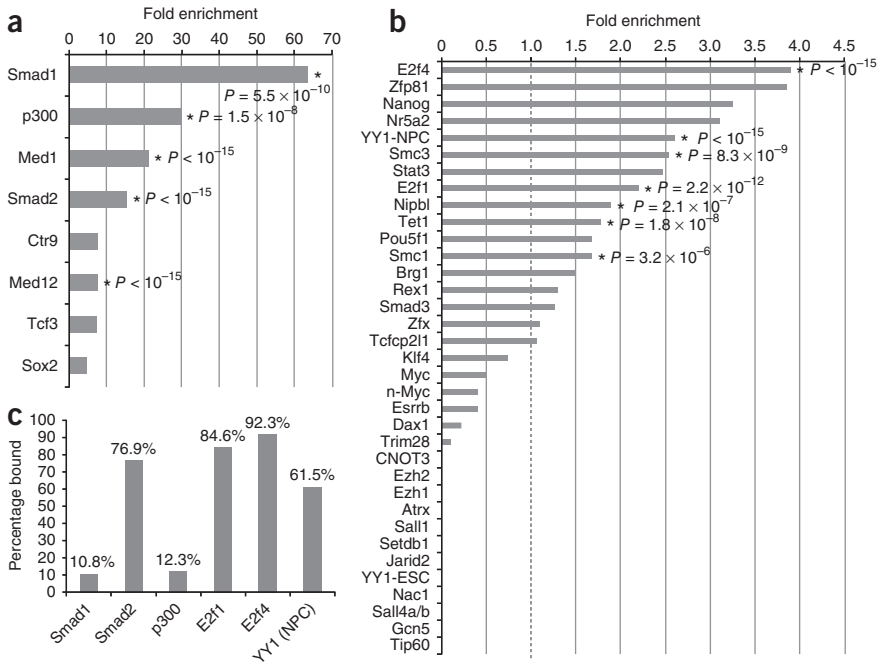

Figure 2 Enrichment analysis of multiple TFs. (**a**,**b**) Enrichment of various DNA-binding factors. Fraction of genes bound by a factor served as the probability of a factor to bind a gene in a random distribution. *, significant enrichment ($P < 10^{-5}$, hypergeometric test). Dashed line indicates no enrichment. (**c**) Percentage of histone genes bound by selected enriched factors. All values are from data in ESCs, except YY1 (NPCs).

Histone genes are bound and regulated by Smad1 and Smad2

Histones constitute only ~0.2% of mouse genes, yet they comprise 16.3% of genes targeted by the TGF-β pathway signal transducer Smad1 (**Fig. 4**), more than 60 times the expected ratio in a random distribution ($P < 10^{-9}$, hypergeometric test). Not only were histone genes enriched for Smad1, but histone genes actually constituted one of the main groups that Smad1 bound in ESCs. Smad1 bound only a small portion of genes in each subtype (H1, 2/5; H2a, 2/17; H2b, 1/18; H3, 3/11; and H4, 4/12 < 2.5 kb from TSS). This suggests that Smad1 regulation of histone genes is not necessarily subtype-oriented and may help explain the different expression of genes within the same subtype.

Smad2, on the contrary, was predominant in histone genes, with 50/64 (78%) of histone genes bound within 1 kb of the TSS. The percentage of histone genes that were regulated by Smad proteins may be even higher than this, owing to the clustering architecture of histone genes, which may indicate that they are co-regulated by similar factors[6].

Unlike the binding of E2f proteins, which correlates with the expression levels of the histone genes they bind, Smad proteins did not display such a trend. To examine the potential regulatory effect of Smad2, we analyzed previously determined gene expression changes after *Smad2* knockout, after Smad2 activation and after Smad2 inhibition[42,43]. TGF-β–stimulated *Smad2*$^{+/+}$ mouse T cells compared with TGF-β–stimulated *Smad2*$^{-/-}$ mouse T cells, exhibited significantly lower expression of histone genes ($P = 0.048$, binomial test, **Supplementary Table 3**). We also observed a significant decrease in expression when comparing TGF-β–stimulated to nonstimulated T cells ($P = 3.81 \times 10^{-6}$, binomial test; **Supplementary Table 3**) and the same trend in *Smad2*$^{-/-}$ T cells ($P = 0.015$, binomial test; **Supplementary Table 3**). In agreement, in mouse ESCs, Smad2 activation induced a decrease in the expression of histone genes ($P = 0.046$, binomial test; **Supplementary Table 4**), whereas Smad2 inhibition did not induce any notable trend. Together, these studies support the hypothesis that TGF-β signaling in general, and Smad2 in particular, inhibit expression of histone genes in mouse cells.

Linker histones are negatively regulated by Zfx

Histone H1 subtypes, unlike the core histones, are not part of the nucleosome octamer, but instead they bind linker DNA between two adjacent nucleosomes. Stoichiometry of histone H1 to core histones differs between cell types[30]. This suggests the existence of a transcriptional regulator that is both histone H1–specific and developmental stage–associated or cell type–associated. To date, two factors, namely H1TF1 and H1TF2, have been identified[17,18], although genome-wide binding data for these factors does not exist and non–histone H1 binding has also been detected for these factors[18].

Our analysis revealed that Zfx was significantly enriched in linker histone genes, binding four of the five histone H1 genes (**Fig. 5a**,**b**), whereas in core histone genes, its binding was akin to a random distribution. Zfx expression is greater in stem cells compared to differentiated cells, and Zfx has been shown to be a major regulator of

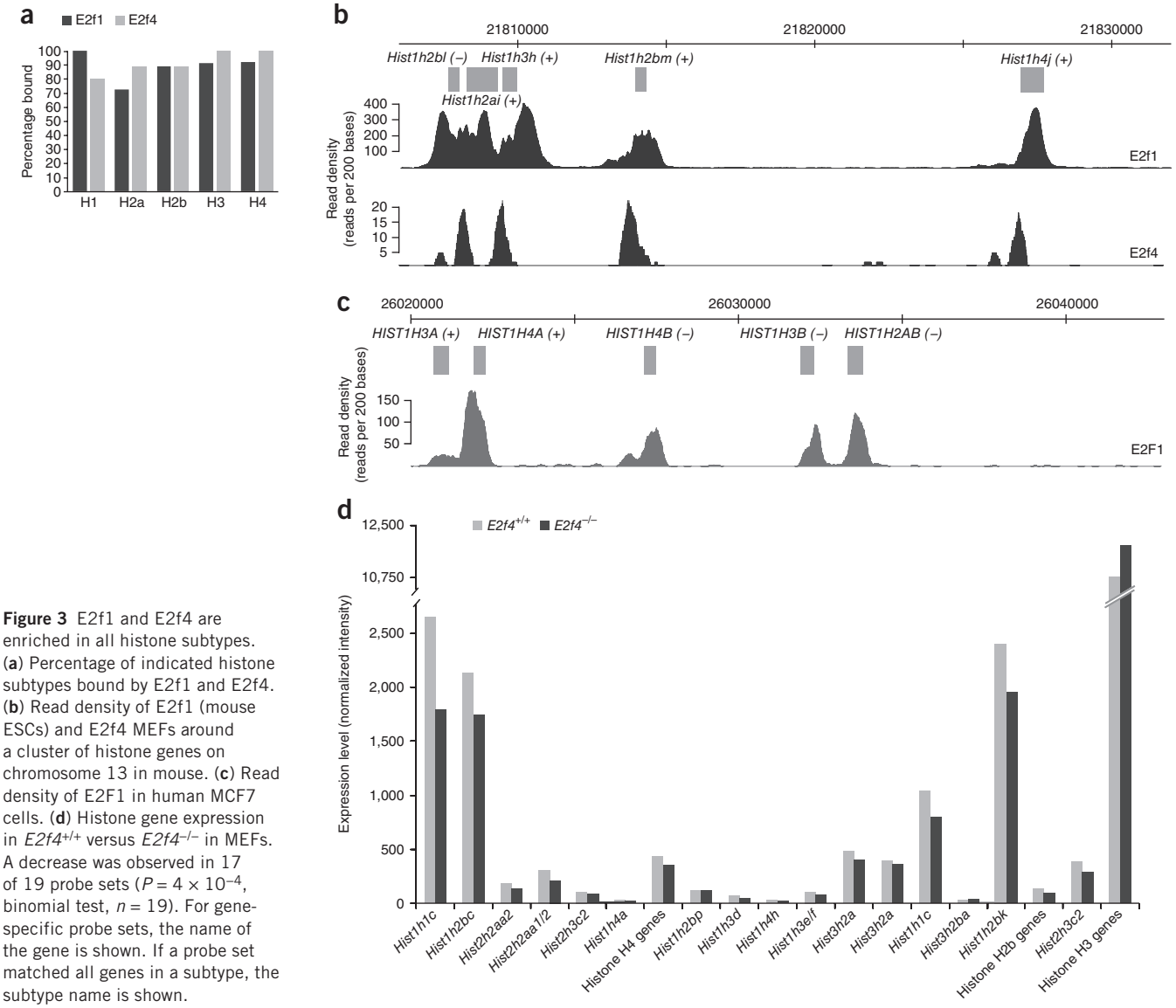

Figure 3 E2f1 and E2f4 are enriched in all histone subtypes. (**a**) Percentage of indicated histone subtypes bound by E2f1 and E2f4. (**b**) Read density of E2f1 (mouse ESCs) and E2f4 MEFs around a cluster of histone genes on chromosome 13 in mouse. (**c**) Read density of E2F1 in human MCF7 cells. (**d**) Histone gene expression in *E2f4*$^{+/+}$ versus *E2f4*$^{-/-}$ in MEFs. A decrease was observed in 17 of 19 probe sets ($P = 4 \times 10^{-4}$, binomial test, $n = 19$). For gene-specific probe sets, the name of the gene is shown. If a probe set matched all genes in a subtype, the subtype name is shown.

self-renewal in ESCs[44]. Histone H1 to core histone stoichiometry has been observed to correlate with the extent of cell differentiation, with the lowest ratio in ESCs and the highest ratio in quiescent cells[30]. This suggests that Zfx may act as a histone H1–specific transcriptional regulator, adding a unique layer of regulation for linker histones that is not observed with core histone genes.

As the ratio of linker to core histones increases with differentiation but Zfx amounts decrease with differentiation, we expected to see a repressive effect of Zfx on histone gene expression. To test this hypothesis, we analyzed previously generated gene expression data in *Zfx*$^{+/+}$ versus *Zfx*$^{-/-}$ mouse ESCs[44]. As predicted, the analysis revealed a significant increase in histone H1 gene expression in the absence of Zfx. For probe sets with a detectable *P* value ($P < 0.01$), we observed a ~50% increase in histone H1 expression, whereas core histone gene expression was unchanged in wild-type versus knockout cells (**Fig. 5c**; $P > 0.25$).

Binding of CTCF to histone genes predicts expression

Our analysis revealed a notable negative correlation between the binding of the CCCTC-binding factor CTCF and expression of both replication-dependent and -independent histones (**Fig. 6a**). This finding suggests that CTCF has a repressive effect on histone genes. Despite this remarkable correlation, we observed no detectable trend when we zoomed in on specific genes. This corresponds with the intricate way in which CTCF regulates expression, which involves DNA looping, insulation and other indirect effects. The strong, yet enigmatic effect that CTCF has on histone genes requires additional investigation.

YY1 is a cell type–specific regulator of histone genes

Despite previous studies that showed that the transcriptional regulator YY1 binds and regulates histone H3.2 genes in hamster fibroblasts[15] and HeLa cells[16], our analysis showed no binding of YY1 within histone H3.2 genes in mouse ESCs. In fact, mouse histone H3.2 genes (*Hist2h3c1, Hist2h3b* and pseudogene *Hist2h3c2*) showed very low YY1 read density (averaging less than eight reads per 1 kb in the vicinity of these genes). The closest YY1 peak was more than 50,000 base pairs away, making YY1 highly unlikely to regulate H3.2 genes in mouse ESCs. Notably, the other 62 replication-dependent histone genes were not bound by YY1. In contrast, in NPCs, we detected significant enrichment of YY1 in histone genes, where YY1 binds 66% (39/59) of core histone promoters (**Fig. 6b**). Our data combined

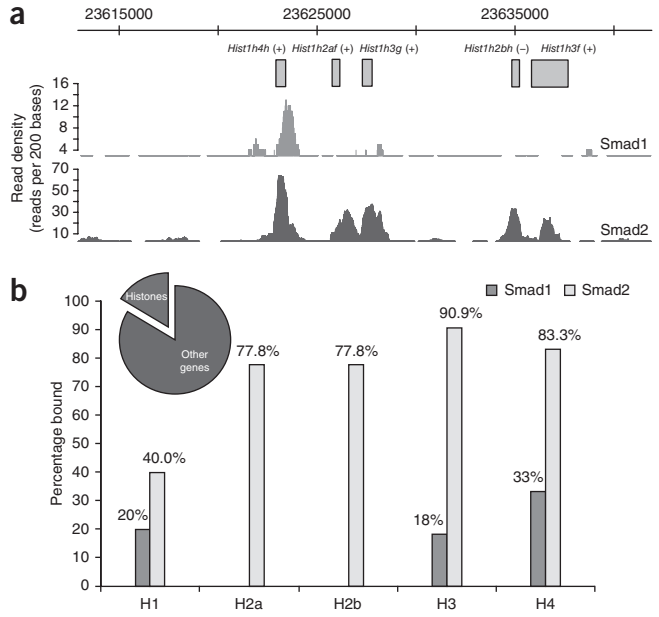

Figure 4 Histone genes are a main target of Smad proteins. (**a**) Read density of Smad1 and Smad2 around a histone cluster on chromosome 13. (**b**) Percentage of genes encoding indicated histone subtype bound by Smad1 and Smad2. The pie chart represents the percentage of Smad1-bound histone genes (16.3%) among all other Smad1-bound genes.

with previous studies strongly suggest that YY1 binding to histone gene promoters is restricted to differentiated cells, and is completely absent in ESCs.

YY1 can form a complex with the histone acetyltransferase p300 (ref. 45). Our analysis revealed a strong enrichment of p300 in the core histone gene family, in accordance with a previous study that showed regulation of histone H4 and H2b genes by p300 (ref. 22). But in addition, we found p300 enrichment around additional histone subtypes, including histones H2A and H1. Notably, enrichment of p300 at ESC histone genes occurred despite a complete absence of YY1, whereas in NPCs, the same genes that are bound by p300 in ESCs were significantly enriched with YY1 ($P = 0.015$). This may suggest that p300 induces binding of YY1 in later developmental stages but not in ESCs[45]. Additional analysis of p300 binding in other cell types revealed that none of the histone genes were bound by p300 in midbrain, forebrain or limb bud, perhaps owing to the high demand for histone genes in ESCs (**Supplementary Fig. 3**). Analyses of the different available histone modifications around the histone genes revealed significant enrichment of H3K9ac and H3K27ac ($P < 10^{-15}$ in both cases), in accordance with p300 action as an acetyltransferase (**Supplementary Fig. 4**).

YY1 also has been shown to be capable of interacting with E2f proteins[46]. However, we observed a completely different behavior for YY1 and E2f proteins, with the latter being enriched in all histone genes in ESCs, whereas YY1 bound none of the histone gene promoters in these cells.

Clustering analysis reveals subtype-oriented regulation

To examine the extent to which histone regulation is subtype-oriented, we created transcriptional signatures for each of the 79 histone genes. This signature was represented as a 43-digit binary code, where every position represents whether a gene is bound by a specific TF (1) or not (0). Next, we calculated a multidimensional distance matrix using Jaccard similarity coefficient, which ignores 0–0 similarities and will take into account only those events in which at least one of the two genes in question is bound by a TF, and therefore is most suitable for this kind of comparison. We embedded the multidimensional distance matrix to two dimensions using classical multidimensional scaling. Finally, we used these data to generate a scatter plot in which the Euclidian distance between two genes represents the extent of similarity of combinatorial factor binding (**Fig. 7**).

In addition to the 64 replication-dependent histone genes in mouse, there are more than a dozen histone variants, which are usually found outside of histone gene clusters, are replication-independent, undergo splicing and are polyadenylated. These variants have unique roles in chromatin and in some cases are cell type–specific[4,47]. The analysis reveals distinct clustering of these variants, which cluster both apart from the replication-dependent histone genes, and apart from one another (**Fig. 7a**). The distinct regulation is evident by the fact that these variants are relatively depleted of E2f1, E2f4, Smad2, Med1, Med12 and YY1 (in NPCs). Oocyte- and testis-specific histones cluster farthest from replication-dependent histones, likely because these genes are not expressed in ESCs. Eight replication-dependent histones clustered away from the rest of the replication-dependent histone genes and closely with the 15 histone variants (**Fig. 7a**). Two of the genes were the sperm-specific *Hist1h2ba* and *Hist1h2aa* genes, but to our surprise, these genes were strongly bound by RNA polymerase II and exhibited high expression in ESCs. Five of the eight genes were in the same genomic locus (chromosome 13: 21898817–21929318), which contains only these five genes. Similarly to the pattern observed for histone variants, this 30-kb locus, was depleted of E2f1, Smad2, Med1 and Med12, accounting for its unique clustering pattern.

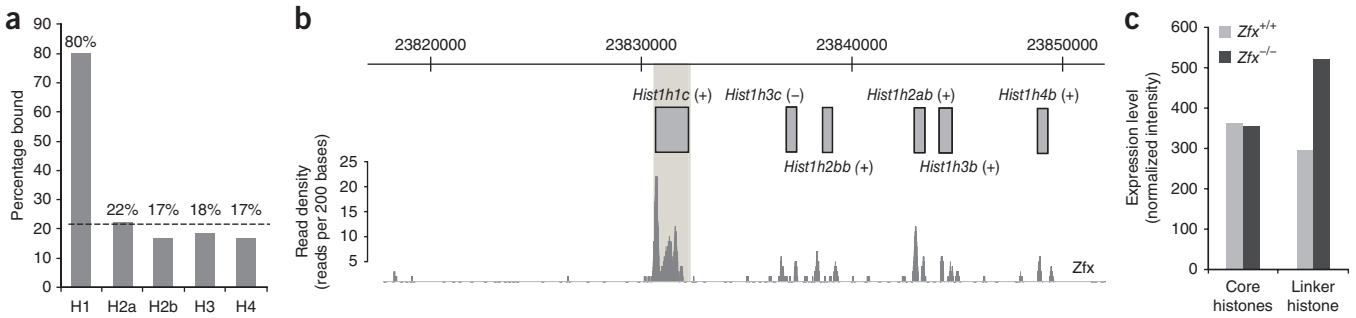

Figure 5 Zfx negatively regulates linker histones. (**a**) Percentage of histone genes bound by Zfx. Linker histones, but not core histones, were significantly enriched with Zfx ($P < 0.01$, hypergeometric test). Horizontal line represents the expected percentage of bound histones (21.5%) in a random distribution. (**b**) Map of Zfx read density with highest read density found around histone H1. (**c**) $Zfx^{+/+}$ versus $Zfx^{-/-}$ expression in core histone genes and linker histone *Hist1h1c* (from ref. 44).

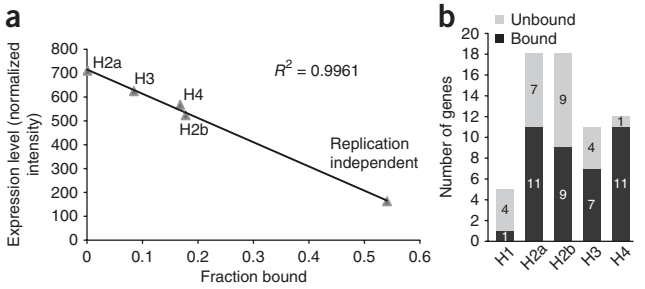

Figure 6 CTCF and YY1 binding to histone genes. (**a**) The average expression for each histone subtype versus the fraction of genes that are bound by CTCF. For replication-dependent histones, expression of non-polyadenylated transcripts is presented. For replication-independent histone variants, expression of polyadenylated transcripts is presented. Distance threshold for binding was determined as 2.5 kb. Solid line indicates linear regression ($R^2 = 0.996$, $P = 10^{-4}$). (**b**) YY1 binding of histone genes in NPCs. In ESCs, no YY1 binding at histone genes was detected.

Another characteristic that sets this locus apart from other replication-dependent histone genes is its histone modification pattern. In general, histone genes are highly enriched with acetylation marks (see above) and with H3K4me2/3 modifications, which mark active promoters (enrichment of 1.3-fold ($P = 0.009$) and 1.4-fold ($P < 10^{-5}$), respectively). This five-gene locus, however, was completely depleted of any active histone modification marks. Outside this locus, acetylation marks were present in more than 90% of replication-dependent histone genes and H3K4me2 peaks were present in all replication-dependent histone genes. Despite the depletion of active histone modification marks and the absence of many of the TFs that predict transcription, genes in this locus were strongly bound by RNA polymerase II and were highly expressed in ESCs.

Additional analysis revealed that genes in the histone H1 subtype clustered closely together and apart from the core histone genes ($P = 1.26 \times 10^{-4}$; Online Methods and **Fig. 7**), suggesting that at least some factors have a tendency to bind histone H1 genes more frequently compared to the core histone genes or vice versa. Considering the different stoichiometry of histone H1 to core histones and the histone H1–specific regulation by Zfx, this comes as no surprise. We observed the same trend for the histone H4 subtype ($P = 0.033$; Online Methods and **Fig. 7b**): 11 of the 12 histone H4 genes clustered closely together. One histone H4 gene (*Hist1h4n*), however, was found in the E2f1-, Smad2- and Med-depleted locus mentioned above, which positions it apart from the other histone H4 genes. Taken together, these clustering analyses suggest that histones can be selectively regulated in a subtype-specific manner.

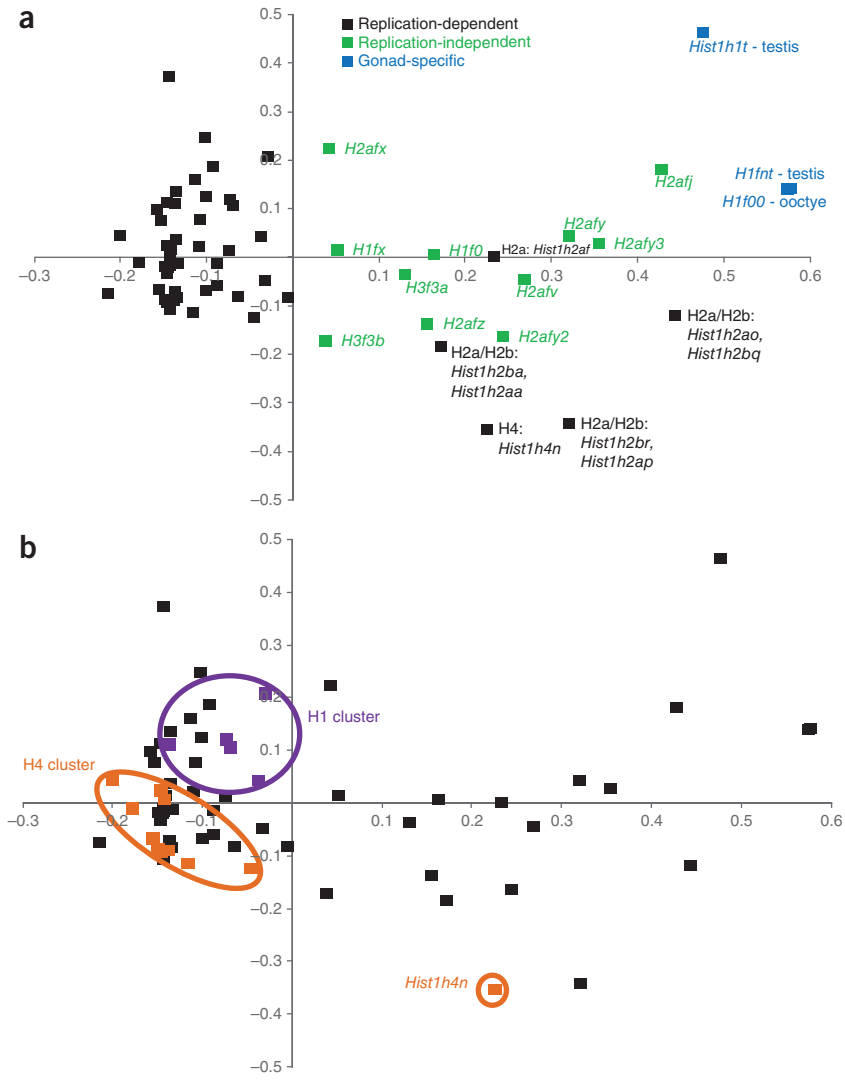

Figure 7 Pair-wise distance between TF binding signatures reveals distinct transcriptional regulation for different histones. (**a**,**b**) Euclidian distance between two genes represents a two-dimensional embedding of the Jaccard distance between TF binding signatures of histone genes. The *x* and *y* axes represent the first and second dimensions, respectively, of the multidimensional scaling. Axes units are arbitrary. Note that the gonad-specific histones cluster farthest apart from replication-dependent histones and relatively close to each other. Circles highlight histone H1 genes and histone H4 genes.

DISCUSSION

Using computational analyses of previously published ChIP-seq and ChIP-chip experiments in mouse ESCs, we identified several potential transcriptional regulators of histone genes, most of which we implicated, to our knowledge for the first time, in regulating histone gene expression. Using gene expression studies in three knockout cell systems of *E2f4, SMAD2* and *Zfx*, we find strong support for our predictions.

Despite their role as major cell-cycle coordinators, E2f proteins were regarded as having a limited role in the regulation of histone gene expression. The surprising finding was not only that the previously considered E2f-independent[20,33] histone H4 genes were vastly bound by E2f, but that essentially all histone genes were bound by either E2f1 or E2f4. The high predictive power of E2f1 for the expression of histone genes, together with previous experimental work that

we analyzed here, strongly suggests that E2f1 and E2f4 are master regulators of the histone gene family. We note that as the analysis of this work was restricted to available data sets, it is possible that additional master regulators of histone gene expression are present in the mammalian genome.

An additional factor that we found to be highly and significantly enriched at histone gene promoters is Smad2, an important factor regulating stem cell differentiation[48]. Most of the knockout and activation or inhibition studies we examined support the hypothesis that TGF-β signaling in general, and Smad2 in particular, inhibits the expression of histone genes in mouse cells. This hypothesis also relies on the general repressive role of Smad2 in previous studies[49]. However, when we examined the effects of a knockout of *Smad2* in nonstimulated cells, we detected a decrease in expression rather than an increase ($P = 0.048$, binomial test; **Supplementary Table 3**). As the effects of the TGF-β superfamily of proteins depend both on their concentration and on the presence of additional factors[50], it is therefore likely that the presence of Smad2 alone would not be sufficient to fully predict its effect.

We also detected high occupancy of histone genes by Tet1 (**Supplementary Fig. 5**). Tet1, which converts 5-methylcytosine to 5-hydroxymethylcytosine, has been previously shown to target promoters of highly transcribed genes[51]. This finding sits well with the fact that histone genes are highly expressed in ESCs.

We also detected high enrichment of Med1, Med12, Nipbl, Smc1 and Smc3 in histone genes (**Fig. 1** and **Supplementary Fig. 6**). These are central factors in DNA-loop formation and cohesin loading, which affect gene expression and have been shown to localize together at both enhancers and core promoters of expressed genes[52]. As histone genes are highly expressed in ESCs, this comes as no surprise. Our analysis also revealed four active histone modification marks that were prevalent among histone genes. These modifications include H3K4me2, H3K4me3, H3K9ac and H3K27ac. Somewhat surprising was the finding that although these active histone marks (as well as several TFs) are enriched in all histone genes, a particular five-gene locus (chromosome 13: 21898817–21929318) was depleted of these active modifications (despite high expression of the genes in this locus).

The ability of ESCs to self-renew and their high proliferation rate requires elevated levels of histone biogenesis. These traits are also common to cancer cells. Notably, all of the factors we identified in this study have been extensively associated with cancer[53–56]. As histones are essential for proliferation, additional study of histone-related TFs in cancer cells might shed light on the mechanisms that allow their self-renewal and high proliferation rates.

To summarize, in this study we combined existing ChIP-seq and ChIP-chip databases in ESCs to systematically identify potential regulators of histone genes in ESCs. Although we found several new factors potentially regulating histone gene expression, which have not been previously implicated in this process, there are likely additional factors for which ChIP-seq data are not available currently. This multifactorial analysis enabled us to identify master versus specific regulators. We propose that the combination of the different factors, whether suppressive (CTCF, Zfx and Smad2), neutral (Smad1) or active (E2f1/4), orchestrate and regulate proper histone gene expression and can explain the intricate modes of regulation, from strict synchronization with S phase to differential expression in cell types and exclusive expression of various histone subtypes.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank L. Carmel and Y. Aaronson for help with statistical analyses, and H. Margalit and O. Rando for critical comments. This work was supported by the Israel Science Foundation, the Abisch Frenkel Foundation, the Israel Cancer Research Foundation and the European Research Council (ERC-281781).

AUTHOR CONTRIBUTIONS

D.G. and E.M. conceived the idea; D.G. and I.L. analyzed the data; B.S.S. and S.M. performed the experiments; D.G. and E.M. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nsmb.2448.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Zhao, J. Coordination of DNA synthesis and histone gene expression during normal cell cycle progression and after DNA damage. *Cell Cycle* **3**, 695–697 (2004).
2. Meeks-Wagner, D. & Hartwell, L.H. Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. *Cell* **44**, 43–52 (1986).
3. Ye, X. *et al.* Defective S phase chromatin assembly causes DNA damage, activation of the S phase checkpoint, and S phase arrest. *Mol. Cell* **11**, 341–351 (2003).
4. Marzluff, W.F. & Duronio, R.J. Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Curr. Opin. Cell Biol.* **14**, 692–699 (2002).
5. Osley, M.A. The regulation of histone synthesis in the cell cycle. *Annu. Rev. Biochem.* **60**, 827–861 (1991).
6. Jaeger, S., Barends, S., Giege, R., Eriani, G. & Martin, F. Expression of metazoan replication-dependent histone genes. *Biochimie* **87**, 827–834 (2005).
7. DeLisle, A.J., Graves, R.A., Marzluff, W.F. & Johnson, L.F. Regulation of histone mRNA production and stability in serum-stimulated mouse 3T6 fibroblasts. *Mol. Cell Biol.* **3**, 1920–1929 (1983).
8. Heintz, N., Sive, H.L. & Roeder, R.G. Regulation of human histone gene expression: kinetics of accumulation and changes in the rate of synthesis and in the half-lives of individual histone mRNAs during the HeLa cell cycle. *Mol. Cell Biol.* **3**, 539–550 (1983).
9. Segil, N., Roberts, S.B. & Heintz, N. Mitotic phosphorylation of the Oct-1 homeodomain and regulation of Oct-1 DNA binding activity. *Science* **254**, 1814–1816 (1991).
10. Fletcher, C., Heintz, N. & Roeder, R.G. Purification and characterization of OTF-1, a transcription factor regulating cell cycle expression of a human histone H2b gene. *Cell* **51**, 773–781 (1987).
11. Mitra, P. *et al.* Identification of HiNF-P, a key activator of cell cycle-controlled histone H4 genes at the onset of S phase. *Mol. Cell Biol.* **23**, 8110–8123 (2003).
12. Zhao, J. *et al.* NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. *Genes Dev.* **14**, 2283–2297 (2000).
13. Artishevsky, A., Wooden, S., Sharma, A., Resendez, E. Jr. & Lee, A.S. Cell-cycle regulatory sequences in a hamster histone promoter and their interactions with cellular factors. *Nature* **328**, 823–827 (1987).
14. Wu, F. & Lee, A.S. Identification of AP-2 as an interactive target of Rb and a regulator of the G1/S control element of the hamster histone H3.2 promoter. *Nucleic Acids Res.* **26**, 4837–4845 (1998).
15. Wu, F. & Lee, A.S. YY1 as a regulator of replication-dependent hamster histone H3.2 promoter and an interactive partner of AP-2. *J. Biol. Chem.* **276**, 28–34 (2001).
16. Last, T.J., van Wijnen, A.J., Birnbaum, M.J., Stein, G.S. & Stein, J.L. Multiple interactions of the transcription factor YY1 with human histone H4 gene regulatory elements. *J. Cell. Biochem.* **72**, 507–516 (1999).
17. La Bella, F., Gallinari, P., McKinney, J. & Heintz, N. Histone H1 subtype-specific consensus elements mediate cell cycle-regulated transcription in vitro. *Genes Dev.* **3**, 1982–1990 (1989).
18. Yagi, H. *et al.* Regulation of the mouse histone H2A.X gene promoter by the transcription factor E2F and CCAAT binding protein. *J. Biol. Chem.* **270**, 18759–18765 (1995).
19. Nelson, D.M. *et al.* Coupling of DNA synthesis and histone synthesis in S phase independent of cyclin/cdk2 activity. *Mol. Cell Biol.* **22**, 7459–7472 (2002).
20. Mitra, P., Vaughan, P.S., Stein, J.L., Stein, G.S. & van Wijnen, A.J. Purification and functional analysis of a novel leucine-zipper/nucleotide-fold protein, BZAP45, stimulating cell cycle regulated histone H4 gene transcription. *Biochemistry* **40**, 10693–10699 (2001).
21. Isogai, Y., Keles, S., Prestel, M., Hochheimer, A. & Tjian, R. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* **21**, 2936–2949 (2007).

22. He, H., Yu, F.X., Sun, C. & Luo, Y. CBP/p300 and SIRT1 are involved in transcriptional regulation of S-phase specific histone genes. *PLoS ONE* **6**, e22088 (2011).
23. DeRan, M., Pulvino, M., Greene, E., Su, C. & Zhao, J. Transcriptional activation of histone genes requires NPAT-dependent recruitment of TRRAP-Tip60 complex to histone promoters during the G1/S phase transition. *Mol. Cell Biol.* **28**, 435–447 (2008).
24. Marti, A., Wirbelauer, C., Scheffner, M. & Krek, W. Interaction between ubiquitin-protein ligase SCFSKP2 and E2F–1 underlies the regulation of E2F–1 degradation. *Nat. Cell Biol.* **1**, 14–19 (1999).
25. Oswald, F., Dobner, T. & Lipp, M. The E2F transcription factor activates a replication-dependent human H2A gene in early S phase of the cell cycle. *Mol. Cell Biol.* **16**, 1889–1895 (1996).
26. Rabinovich, A., Jin, V.X., Rabinovich, R., Xu, X. & Farnham, P.J. E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. *Genome Res.* **18**, 1763–1777 (2008).
27. Seyedin, S.M., Cole, R.D. & Kistler, W.S. H1 histones from mammalian testes. The widespread occurrence of H1t. *Exp. Cell Res.* **136**, 399–405 (1981).
28. Yang, L., Duff, M.O., Graveley, B.R., Carmichael, G.G. & Chen, L.L. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol.* **12**, R16 (2011).
29. Levine, B.J., Liu, T.J., Marzluff, W.F. & Skoultchi, A.I. Differential expression of individual members of the histone multigene family due to sequences in the 5′ and 3′ regions of the genes. *Mol. Cell Biol.* **8**, 1887–1895 (1988).
30. Fan, Y. *et al.* Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123**, 1199–1212 (2005).
31. Stead, E. *et al.* Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene* **21**, 8320–8333 (2002).
32. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
33. van Wijnen, A.J. *et al.* CDP/cut is the DNA-binding subunit of histone gene transcription factor HiNF-D: a mechanism for gene regulation at the G1/S phase cell cycle transition point independent of transcription factor E2F. *Proc. Natl. Acad. Sci. USA* **93**, 11516–11521 (1996).
34. Bieda, M., Xu, X., Singer, M.A., Green, R. & Farnham, P.J. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res.* **16**, 595–605 (2006).
35. Antonio Giordano, A.G. & Soprano, K.J. *Cell Cycle Inhibitors in Cancer Therapy: Current Strategies*, (Humana Press inc., Totowa, New Jersey, 2003).
36. Cao, A.R. *et al.* Genome-wide analysis of transcription factor E2F1 mutant proteins reveals that N- and C-terminal protein interaction domains do not participate in targeting E2F1 to the human genome. *J. Biol. Chem.* **286**, 11985–11996 (2011).
37. Heckmann, S. *et al.* The E2F transcription factor family regulates CENH3 expression in *Arabidopsis* thaliana. *Plant J.* **68**, 646–656 (2011).
38. Holmes, W.F. *et al.* Coordinate control and selective expression of the full complement of replication-dependent histone H4 genes in normal and cancer cells. *J. Biol. Chem.* **280**, 37400–37407 (2005).
39. Andrechek, E.R., Mori, S., Rempel, R.E., Chang, J.T. & Nevins, J.R. Patterns of cell signaling pathway activation that characterize mammary development. *Development* **135**, 2403–2413 (2008).
40. Xu, X. *et al.* A comprehensive ChIP-chip analysis of E2F1, E2F4, and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members. *Genome Res.* **17**, 1550–1561 (2007).
41. Gaubatz, S. *et al.* E2F4 and E2F5 play an essential role in pocket protein-mediated G1 control. *Mol. Cell* **6**, 729–735 (2000).
42. Takimoto, T. *et al.* Smad2 and Smad3 are redundantly essential for the TGF-beta-mediated regulation of regulatory T plasticity and Th1 development. *J. Immunol.* **185**, 842–855 (2010).
43. Galvin, K.E., Travis, E.D., Yee, D., Magnuson, T. & Vivian, J.L. Nodal signaling regulates the bone morphogenic protein pluripotency pathway in mouse embryonic stem cells. *J. Biol. Chem.* **285**, 19747–19756 (2010).
44. Galan-Caridad, J.M. *et al.* Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell* **129**, 345–357 (2007).
45. Lee, J.S. *et al.* Relief of YY1 transcriptional repression by adenovirus E1A is mediated by E1A-associated protein p300. *Genes Dev.* **9**, 1188–1198 (1995).
46. Schlisio, S., Halperin, T., Vidal, M. & Nevins, J.R. Interaction of YY1 with E2Fs, mediated by RYBP, provides a mechanism for specificity of E2F function. *EMBO J.* **21**, 5775–5786 (2002).
47. Marzluff, W.F., Wagner, E.J. & Duronio, R.J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* **9**, 843–854 (2008).
48. Heldin, C.H., Miyazono, K. & ten Dijke, P. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**, 465–471 (1997).
49. Ten Dijke, P., Miyazono, K. & Heldin, C.H. Signaling inputs converge on nuclear effectors in TGF-beta signaling. *Trends Biochem. Sci.* **25**, 64–70 (2000).
50. Pucéat, M. TGFbeta in the differentiation of embryonic stem cells. *Cardiovasc. Res.* **74**, 256–261 (2007).
51. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* **25**, 2436–2452 (2011).
52. Kagey, M.H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
53. Chen, H.Z., Tsai, S.Y. & Leone, G. Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat. Rev. Cancer* **9**, 785–797 (2009).
54. Nikpour, P., Emadi-Baygi, M., Mohammad-Hashem, F., Maracy, M.R. & Haghjooy-Javanmard, S. Differential expression of ZFX gene in gastric cancer. *J. Biosci.* **37**, 85–90 (2012).
55. Nicholson, S., Whitehouse, H., Naidoo, K. & Byers, R.J. Yin Yang 1 in human cancer. *Crit. Rev. Oncog.* **16**, 245–260 (2011).
56. Fiorentino, F.P. & Giordano, A. The tumor suppressor role of CTCF. *J. Cell Physiol.* **227**, 479–492 (2012).

ONLINE METHODS

ChIP-seq and ChIP-chip data were analyzed for the following TFs: Nanog, Pou5f1, Sox2, Smad1, Smad2, Smad3, Smad4, E2f1, E2f4, Tcfcp2l1, Zfx, Stat3, Klf4, Esrrb, n-Myc c-Myc, CTCF, Brg1, p300, Tet1, Cnot3, Ezh1, Ezh2, Atrx, Sall1, Setdb1, Tcf3, Jarid2, YY1, Ctr9, Nac1, Nr5a2, Rex1, Nr0b (also known as Dax1), Sall4a, Sall4b, Tip60, Gcn5, Trim28, Zfp281, Smc1, Smc3, Nipbl, Med1 and Med12 (for references, see **Supplementary Table 1**). E2f1 and Smad1 reads were downloaded from the Gene Expression Omnibus (GEO) under accession number GSE11431, converted to mm9 using UCSC Lift-Over and peaks were called using MACS[32] with *P*-value cutoff of 10^{-5}. Peaks for Nanog, Pou5f1, Sox2, Tcfcp2l1, Zfx, Stat3, Klf4, Esrrb, n-Myc c-Myc and CTCF were downloaded from GEO under accession number GSE11431. Peaks for E2f4 (ESC) were downloaded from GEO accession number GSE20551 and converted to mm9. Peaks for E2f4 (B-cell lymphoma, myoblast and erythroleukemia were downloaded from the ENCODE[57]. Peaks for E2f4 (liver, adipocyte) were downloaded from GEO accession number GSM427091. Peaks for Smad2 were downloaded from GEO accession number GSM578475. Peaks for the following TFs were downloaded from hmChIP[58] with GEO accession numbers: GSM551138 (Atrx), GSE25523 (Sall1), GSE18371 (Setdb1), GSE11724 (Tcf3), GSE19167 (Jarid2), GSE11329 (c-Myc, Nac1, Rex1, Zfp281, Dax1), GSE19019 (Nr5a2), GSM526869 (Sall4a, Sall4b), GSE20551 (Tip60, Gcn5), GSE12283 (Trim28), GSE14654 (Ctr9), GSE12283 (Cnot3), GSE15388 (Ezh1) and GSE19167 (Ezh2). Peaks were converted from mm 8 to mm 9 using UCSC Lift-Over. p300 peaks (ESCs) (middle threshold) were downloaded from GEO accession number GSM558675 and converted to mm 9 using UCSC Lift-Over. Midbrain, forebrain and limb bud p300 peaks were downloaded from GEO accession number GSE13845. Brg1 raw reads were downloaded from GEO accession number GSM359413. Peaks were generated using MACS with *P*-value cutoff of 10^{-4} and were converted from mm 8 to mm 9 using UCSC Lift-Over. Reads for E2f4 ChIP-seq in secondary MEFs (**Figs. 2b** and **6**) were taken from GEO accession number GSM602777. Raw reads for Tet1, Smad3, YY1 (ESC and NPC), Smc1, Smc3, Med1, Med12 and Nipbl were downloaded from GEO accession numbers GSM706672, GSM539541, GSM628031, GSM628032 and GSE22557 respectively. Reads were aligned to mm9 genome assembly using BOWTIE[59], allowing up to 2 mismatches. MACS was used to create peaks (*P*-value cutoff of 10^{-4}). For H3K9ac histone modification, previously generated peaks were used. For H3K27ac, raw reads were downloaded from GEO under accession number GSE24164, converted to mm9 and MACS was used using threshold of $P < 10^{-4}$.

Unless otherwise specified, a gene was considered bound by a TF if a peak was located within 1,000 base pairs upstream or downstream of the TSS. Genes were taken from NBCI build 37. Genes were divided in a binary manner to either bound or unbound, with no consideration of the number of peaks on a gene.

Enrichment level was calculated as

$$\frac{\text{bound histones}}{\text{all histones}} \div \frac{\text{bound genes}}{\text{all genes}}.$$

Statistical significance of enrichment was calculated using hypergeometric distribution with a Bonferroni correction. Nonpolyadenylated histone gene expression (tiling array and RNA-seq) in ESCs and NPCs were taken from an unpublished dataset (I.L. and E.M.; unpublished data).

The statistic for histone H1 clustering was selected as the mean distance from the center of mass of the replication-dependent histone genes. The statistic for histone H4 was selected as the Fisher ratio det(Sb)/det(Sw), where Sb is the between-group covariance matrix and Sw is the within-group covariance matrix. Statistical significance was measured by comparing the statistic to 1,000,000 repetitions using randomized grouping of the histones.

ChIP was performed as previously described[60]. Antibodies to the following proteins (5 μg per ChIP reaction) were used: Smad1 (Santa Cruz, sc-7965), E2f1 (Upstate Millipore, 05-379) and Zfx (custom-made; ref. 61).

57. The ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
58. Chen, L., Wu, G. & Ji, H. hmChIP: a database and web server for exploring publicly available human and mouse ChIP-seq and ChIP-chip data. *Bioinformatics* **27**, 1447–1448 (2011).
59. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
60. Sailaja, B.S., Takizawa, T. & Meshorer, E. Chromatin immunoprecipitation in mouse hippocampal cells and tissues. *Methods Mol. Biol.* **809**, 353–364 (2012).
61. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).