Coassembly of REST and its cofactors at sites of gene repression in embryonic stem cells

Hong-Bing Yu, Rory Johnson, Galih Kunarso, et al.

Genome Res. 2011 21: 1284-1293 originally published online June 1, 2011
Access the most recent version at doi:10.1101/gr.114488.110

Coassembly of REST and its cofactors at sites of gene repression in embryonic stem cells

Hong-Bing Yu,[1,2,3] Rory Johnson,[1,3,4] Galih Kunarso,[1] and Lawrence W. Stanton[1,2,5]

[1]*Stem Cell and Developmental Biology, Genome Institute of Singapore, 138672 Singapore;* [2]*Department of Biological Sciences, National University of Singapore, 117543 Singapore*

The differentiation of pluripotent embryonic stem cells is regulated by networks of activating and repressing transcription factors that orchestrate determinate patterns of gene expression. With the recent mapping of target sites for many transcription factors, it has been a conundrum that so few of the genes directly targeted by these factors are transcriptionally responsive to the binding of that factor. To address this, we generated genome-wide maps of the transcriptional repressor REST and five of its corepressors in mouse embryonic stem cells. Combining these binding-site maps with comprehensive gene-expression profiling, we show that REST is functionally heterogeneous. Approximately half of its binding sites apparently are nonfunctional, having weaker binding of REST and low recruitment of corepressors. In contrast, the other sites strongly recruit REST and corepressor complexes with varying numbers of components. Strikingly, the latter sites account for almost all observed gene regulation. These data support a model where productive gene repression by REST requires assembly of a multimeric "repressosome" complex, whereas weak recruitment of REST and its cofactors is insufficient to repress gene expression.

[Supplemental material is available for this article.]

The pluripotent state of embryonic stem cells (ESC) is maintained by a highly interconnected network of regulatory transcription factors (TF). These factors serve as a DNA sequence-specific platform for the recruitment of various positively and negatively acting cofactors that maintain high expression of self-renewal and propagation genes, while repressing those necessary for differentiation (Boyer et al. 2006). The genome-wide binding profiles of pluripotency TFs have been extensively mapped using chromatin immunoprecipitation (ChIP), showing that most TF are recruited to many thousands of genes throughout the genome (Boyer et al. 2005; Loh et al. 2006; Chen et al. 2008). However, initial optimism that these maps might afford a complete genetic map of pluripotency has been tempered by the realization that only a small fraction of genomic recruitment actually results in direct gene regulation (Li et al. 2008). Therefore, TF-binding data alone are not sufficient to specify gene-regulatory relationships in ESC. Previously, we proposed that differential cofactor recruitment may help explain this heterogeneity in TF function (Johnson et al. 2008), but the absence of comprehensive genome-wide studies on cofactor recruitment has left this hypothesis untested.

Pluripotency of ESC is maintained by a group of TFs, including NANOG, POU5F1 (also known as OCT4), and SOX2, as well as several other "pluripotency factors" (Chen et al. 2008; Kim et al. 2008). These TFs may be individually recruited to target genes, or may be found at common DNA enhancer elements where they form large, multifactor complexes that recruit the P300 (also known as EP300) coactivator complex and drive gene transcription (Chen et al. 2008). Surprisingly, these studies have consistently shown that only a small fraction of genes that are bound by any given TF is actually transcriptionally regulated by it (Johnson et al. 2008; Li et al. 2008).

[3]These authors contributed equally to this work.
[4]Present address: Bioinformatics and Genomics Group, Centre for Genomic Regulation, Barcelona, Catalonia, Spain.
[5]Corresponding author.
E-mail stantonl@gis.a-star.edu.sg.
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.114488.110.

Negative gene regulation is also important in specifying the ESC phenotype. Possibly the best understood ESC repressor is REST—the RE1-silencing transcription factor (Ballas et al. 2005). This zinc finger TF appears to be a key repressor of neural gene expression in ESC and the developing nervous system (Chen et al. 1998). The role of REST in ESC is controversial: While one report suggested that the presence of REST was necessary for pluripotency in mouse ESC (Singh et al. 2008), many other groups have depleted REST from ESC with no apparent change in differentiation status (Loh et al. 2006; Buckley et al. 2009; Jorgensen et al. 2009a; Yamada et al. 2010). Regardless, the high levels of REST in ESC (Ballas et al. 2005; Sun et al. 2005), its evolutionarily conserved targeting by POU5F1-NANOG (Boyer et al. 2005; Loh et al. 2006), and its ability to regulate neuronal gene expression (Chen et al. 1998) strongly suggest that REST is important for suppression of the neuronal gene-expression program in ESC. Most recently, convincing evidence points to a role for REST in shutting down the pluripotency program in the early stages of differentiation (Yamada et al. 2010), which explains its recruitment to, but weak repression of pluripotency genes such as *Lin28* and *Nanog* in mESC (Johnson et al. 2008).

REST is specifically recruited by an unusually long DNA element, the repressor element 1 (RE1). Having little intrinsic repressive activity, REST serves to recruit corepressor complexes through distinct N- and C-terminal repressor domains (Ooi and Wood 2007). Amongst the best-characterized REST cofactors are the SIN3A/B complexes, containing various histone deacetylases (HDAC) (Roopra et al. 2000), and the CoREST complex comprising the various CoREST paralogs (RCOR1, RCOR2, RCOR3) (Andres et al. 1999). Many other corepressor partners of REST have also been identified, including KDM5C (Tahiliani et al. 2007), CTDSP1 (Yeo et al. 2005), KDM1A (Shi et al. 2004), and SMARCA4 (Battaglioli et al. 2002). The significance of so many possible binding partners is unclear and the genome-wide recruitment of transcriptional cofactors by REST or any other factor has never been systematically studied.

Previously, we made the striking observation that only a minority of those genes bound by REST are functional targets (Johnson et al. 2008). Thus, the genomic recruitment of a transcriptional

21:1284–1293 © 2011 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/11; www.genome.org

repressor is not sufficient for repression. This finding is echoed in studies elsewhere on human (Birney et al. 2007), mouse (Visel et al. 2009), and fly (Li et al. 2008). One explanation for this is that REST binding sites do not have uniform recruitment of transcriptional cofactors, and that those genes where REST is capable of robustly recruiting cofactors are those that are strongly transcriptionally repressed. By distinguishing those genes where REST is effectively recruiting cofactors, we might develop more refined predictions of REST target genes in a given cell.

In the present study, we have carried out a comprehensive, genome-wide study of multiple corepressor complexes in embryonic stem cells. We demonstrate that recruitment of REST and its cofactors is not uniform, but rather that some sites efficiently recruit large corepressor complexes, while others are less capable of doing so. The recruitment of cofactors appears to be a function of the intensity of REST recruitment, and the latter also accounts for some of the differences in repression observed between genes.

Results

A tagging approach to genome-wide mapping of REST cofactors

Chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq) has been successfully applied to generate genome-wide maps of TF recruitment (Johnson et al. 2007). We attempted ChIP with various commercial antibodies to REST and its cofactors RCOR1, RCOR2, RCOR3, SIN3A, and SIN3B. With the exception of SIN3A and REST, we were unable to find any antibody yielding ChIP enrichment of sufficient quality for ChIP-seq. Thus, we resorted to a tagging strategy similar to those used previously (Kim et al. 2008). Lentiviral constructs containing cDNAs encoding RCOR1, RCOR2, RCOR3, and SIN3B fused to the V5 epitope tag were used to infect mouse ESC. Using a GFP reporter gene coexpressed from the lentiviral constructs, we specifically selected cell populations expressing tagged transgenes comparable to endogenous levels. Subsequent validation by quantitative reverse transcriptase polymerase chain reaction (qRT–PCR) and Western blotting showed that the ratio of tagged transgene to endogenous gene expression was less than 2:1 in all cases (Fig. 1A; Supplemental Fig. S1). Furthermore, we showed that the expression of tagged cofactor transgenes did not lead to any loss of pluripotency in ESC: (1) In all cases these cells had characteristic morphology of undifferentiated ESC, indistinguishable from wild-type cells (Fig. 1B); and (2) qRT–PCR measurements indicated that levels of core pluripotency gene mRNAs were unaffected (Fig. 1A). Finally, in ChIP experiments using a V5 antibody, we found that tagged transgenes are specifically recruited to known REST binding sites (Johnson et al. 2008), and could not be detected at control loci that do not bind REST (Fig. 1C). In summary, our tagging and ChIP strategy is not likely to perturb endogenous ESC transcriptional networks, while allowing sensitive and specific detection of genuine cofactor-binding loci.

Genome-wide mapping of transcriptional repressor complexes in ESC

We sought to reconstruct the genomic repressor network that controls gene transcription in ESC. ChIP DNAs prepared from REST and five of its cofactors were subjected to high-throughput sequencing. For each experiment 11.2–15.8 million sequence reads were uniquely mapped to the mouse genome. In order to find sequence reads indicative of cofactor recruitment, we tested several peak-finding algorithms (Supplemental Fig. S2). Due to its sensitivity and selectivity in analysis of REST peaks, we used the SISSRs algorithm (Jothi et al. 2008). Despite similar sequencing depths, we observed large differences in the number of peaks amongst the samples, from 5318 for SIN3B, to 84 for RCOR1 (Fig. 2A). We expect that this is a result of differential ChIP efficiencies of the various cofactors, even though they are tagged with the same epitope. Although in principle it is possible to relax the peak-finding stringency for weaker ChIP-seq datasets in order to increase the number of peaks, we found that this introduces a disproportionate amount of spurious peaks. Consequently, we pursued our analysis on the various cofactor peaks that had been identified using identical levels of stringency.

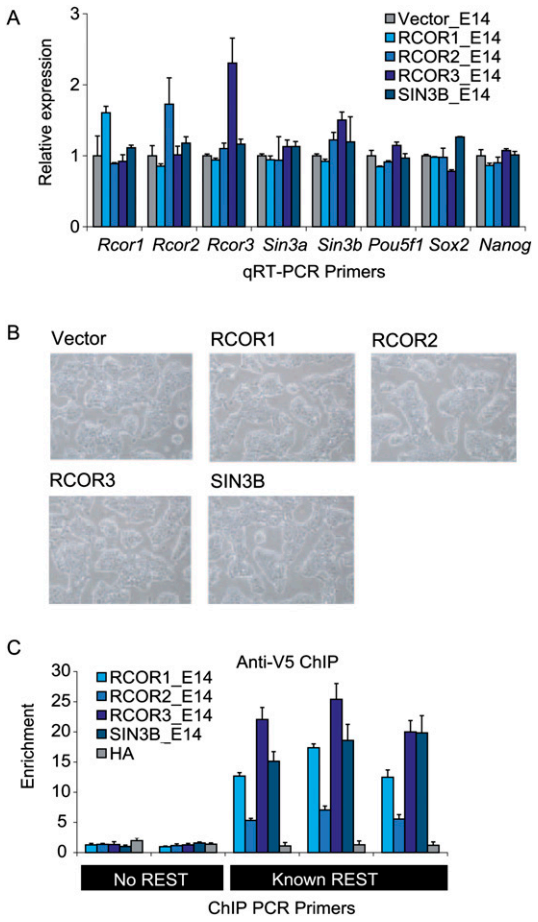

Figure 1. Low-level expression of tagged cofactor transgenes in mouse embryonic stem cells. (*A*) Quantitation of mRNA levels in ESC lines by qRT–PCR. cDNA samples from control ("Vector") and corepressor-expressing cells ("RCOR1_E14" etc.) were interrogated using primers against corepressor mRNAs and pluripotency genes. Raw data were normalized to the housekeeping gene *Actb*, then displayed as a fold difference compared with Vector control. Error bars indicate the standard error of the mean for triplicate biological experiments. (*B*) Bright field microscopy of mouse embryonic stem cell lines used in this study. FACS-sorted and replated cells display normal, undifferentiated morphology. (*C*) Chromatin immunoprecipitations of V5-tagged cofactor proteins were carried out using anti-V5. ChIP DNA was interrogated in qPCR with primers for three known REST target loci (Johnson et al. 2008) as positive controls, and two loci with no evidence for REST recruitment as negative controls. A sham HA antibody was used to immunoprecipitate chromatin from the RCOR1_E14 line as a further control. Enrichments were calculated with reference to a non-REST background PCR amplicon. ChIP experiments were carried out on at least three independent cell cultures.

A

Factor	Reads ($x10^6$)	Peaks	Overlap REST
REST	15.8	3104	
RCOR1	12.5	84	61
RCOR2	12.4	459	43
RCOR3	11.2	3744	885
SIN3A	12.8	1314	230
SIN3B	11.3	5318	1064

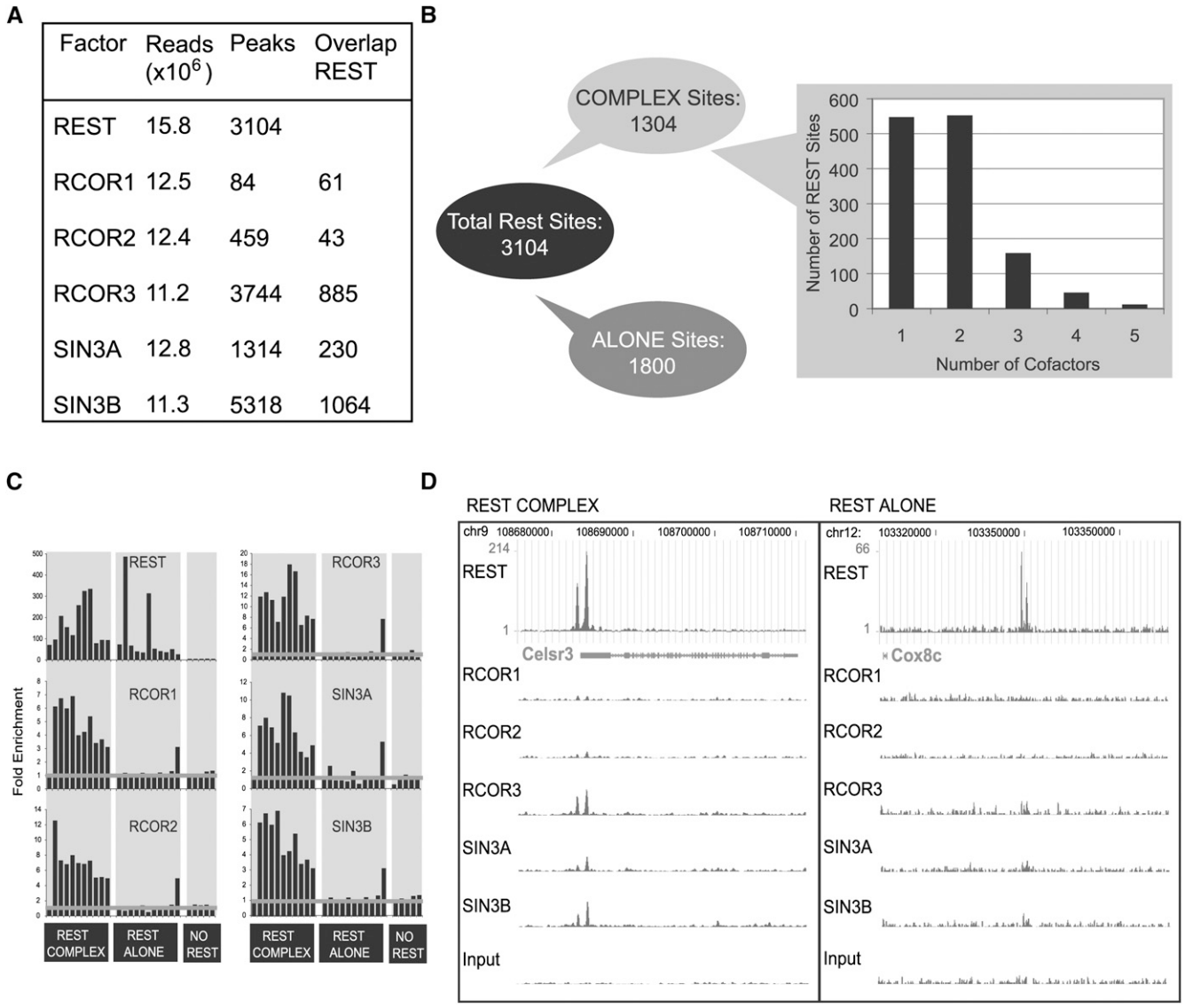

Figure 2. Genomic mapping of cofactor recruitment in ESC by ChIP-seq. (*A*) ChIP-seq statistics for REST and five cofactors. "Reads" indicates the number of millions of uniquely mappable sequence reads. "Peaks" of ChIP-seq reads, indicative of genomic binding loci, were identified using the SISSRs program, as described in Methods. The number of cofactor peaks falling within 100 bp of a REST peak is also indicated. (*B*) Breakdown of REST binding sites, based on cofactor recruitment. REST "Complex" sites are defined as those overlapping at least one of the five cofactors, and "Alone" sites bind only REST. (*C*) Conventional ChIP–qPCR was used to validate REST Alone and Complex sites. The enrichment of cofactors was measured for randomly selected REST binding sites from the REST Complex (five Factors) and the REST Alone sets. Enrichment was also measured at five control loci with no evidence for recruitment of REST. The horizontal gray line denotes background. ChIP was carried out using anti-V5 antibody, with the exception of REST and SIN3A, where endogenous protein was measured. (*D*) Examples of REST Complex (*left*) and REST Alone (*right*) loci. The green bars indicate the density of ChIP-seq reads across the genome. Exons of the proximal genes (*Celsr3* and *Cox8c*) are shown as blue boxes.

Although the corepressors in this study are all known partners of REST, it is likely that they interact with other transcriptional repressors in ESC. Motif analysis showed that the characteristic RE1 motif responsible for genomic recruitment of REST was significantly enriched at the binding sites of each cofactor (Supplemental Fig. S3). We concentrated on cofactor recruitment mediated by REST. By overlapping the genomic peaks of all cofactor libraries with that of REST, we found two distinct populations of REST binding sites: those where REST recruited one or more cofactors, and those where REST was present but did not recruit detectable levels of any cofactor (Fig. 2B). We termed these "REST Complex" and "REST Alone" sites, respectively. The REST Alone sites slightly outnumber the REST Complex sites (1800 vs. 1304, respectively). The majority of the latter contain either one or two cofactors, with a smaller number of REST sites recruiting three, four, or five cofactors. To validate these findings, we randomly selected several examples of REST Complex (containing five cofactors) and REST Alone sites, as well as non-REST binding loci, and measured the enrichment of all factors using conventional ChIP–PCR (Fig. 2C). As expected, we detected REST at all of the sites. Generally, REST recruitment was broadly lower at REST Alone sites, with exceptions. In agreement with ChIP-seq, cofactor recruitment was almost exclusively detected at REST Complex sites. Examples of raw sequencing reads across REST Complex (*Celsr3*) and REST Alone

(*Cox8c*) loci are shown in Figure 2D. These data suggest that the ability of REST binding sites to recruit transcriptional cofactors is not uniform.

The repressosome complex in ESC

Activating TFs in ESC bind target genes as part of at least two large, corecruitment complexes termed MTL (multiple TF loci) (Chen et al. 2008). These clusters were discovered by applying unsupervised, hierarchical clustering to genomic maps of multiple TF. Most strikingly, the pluripotency factors NANOG, POU5F1, and SOX2 are all members of a P300 histone acetyltransferase-recruiting MTL, or "enhanceosome," which maintains high expression of pluripotency genes in ESC (Chen et al. 2008; Kim et al. 2008). A second, distinct MTL comprises MYC proteins, in addition to ZFX, ESRRB, and E2F—the "MYC cluster" (Chen et al. 2008). To establish the independence or not of REST and its corepressors in this framework, we repeated hierarchical clustering with new and existing genomic location data (Fig. 3A). This analysis clearly delineated the NANOG-POU5F1-SOX2 and MYC MTLs. In addition, we found that REST and REST corepressors strongly cocluster in a new "Repressosome" MTL. This MTL is clearly distinguishable from the other two. Closer inspection suggests that, in addition to binding REST, subpopulations of SIN3A and SIN3B sites are recruited to the MYC cluster MTL—9.8% of MYC MTLs show evidence for binding of either SIN3A or SIN3B. In addition, 21% and 20% of RCOR2 and RCOR3, respectively, colocalize with CTCF, suggesting that these factors may play a role in boundary element specification. Mindful of recent studies, we looked for significant recruitment of SIN3A by POU5F1 (Liang et al. 2008). Just a small fraction (1.5%) of POU5F1 regions colocalizes with SIN3A (representing 4% of SIN3A sites), in contrast to 3.5% that recruit the coactivator complex, P300. We investigated the functional impact on gene expression of differential cofactor recruitment to POU5F1 loci (Supplemental Fig. S4). Surprisingly, those genes proximal to POU5F1–SIN3A complexes showed no evidence for activation or repression, in contrast to POU5F1–P300 proximal genes, which were strongly activated. Thus, our data are not consistent with genome-wide functional interaction between POU5F1 and SIN3A in ESC.

Much work has been carried out on the epigenetic consequences of REST recruitment to genomic loci (Ooi et al. 2006; Greenway et al. 2007; Zheng et al. 2009). It has been proposed that the direct deposition of chromatin modifications at RE1 sites may be an important factor in productive gene repression. To test this, we performed similar clustering analysis using available genome-wide epigenetic maps derived from ESC (Fig. 3B). We included maps of histone modifications associated with active enhancers (H3K4me1, H3K4me2), active promoters (H3K4me3), repressed chromatin (H3K9me3), transcriptional elongation (H3K36me3), and heterochromatin (H4K20me3) (Mikkelsen et al. 2007; Ku et al. 2008; Meissner et al. 2008). We also included a map of the nucleosome remodeling factor SMARCA4, previously shown by biochemical methods to be recruited by REST (Battaglioli et al. 2002). Activating histone methylations on H3K4 strongly colocalize, along with SMARCA4, SIN3A, and SIN3B. H3K27me3 also belongs to this cluster, indicative of "bivalent" domains characteristic of embryonic stem cells (Bernstein et al. 2006). Given that REST does not appear to colocalize with these regions, we surmise that SIN3A and SIN3B are recruited by other transcription-factor complexes at these sites. In general, however, REST and its cofactors strongly clustered together. Surprisingly, we observed little association between REST binding sites and any specific histone modification, indicating that there is no universal signature of local histone modification at REST binding sites. Furthermore, this analysis provided scant evidence for interaction of SMARCA4 and REST in ESC.

The Polycomb complex is thought to maintain cellular identity by silencing large numbers of lineage-specific genes through the deposition of methylation marks on histone H3K27 (Ku et al. 2008). Clustering analysis showed that REST and its cofactors are almost completely excluded from Polycomb-bound regions (Fig. 3C).

In summary, clustering analysis shows that REST and its cofactors form a largely distinct molecular repressosome complex in ESC. The repressosome does not deposit a consistent chromatin mark at its binding sites. Our data broadly support a model of REST binding and repressing transcriptionally active regions of the genome, and not binding preferentially to epigenetically silenced regions as has been proposed previously (Lunyak et al. 2002).

REST Alone and Complex sites differ in their ability to recruit REST and its cofactors

We hypothesized that the apparent presence or absence of REST cofactors may be determined by the magnitude of REST recruitment to a particular locus. In other words, loci where REST is strongly bound recruit high levels of cofactors that are detected by ChIP-seq, while more weakly bound REST sites also have lower levels of cofactors that are not detected above the ChIP-seq threshold. Analysis of the heights of REST binding peaks for REST Alone and REST Complex sites showed that this may be the case: The REST Alone set represents a well-defined population of weak binding sites with low peaks, while REST Complex sites extend over a wide range of high peak heights (Fig. 4A). Furthermore, manual inspection of ChIP-seq read densities across some REST Alone sites suggests that weak cofactor recruitment does take place, but cannot be detected at the default sensitivity settings of the SISSRS peak-finding algorithm. We checked this systematically by examining the density of ChIP-seq reads from cofactors across both REST Alone and Complex loci (Supplemental Fig. S5). This analysis clearly showed that REST Alone sites do recruit appreciable amounts of all of the cofactors, but at levels far below that of REST Complex sites. In fact, cofactor recruitment has a clear positive correlation to the intensity of REST recruitment, for both REST Alone and REST Complex sites (Supplemental Fig. S6). Thus, REST Alone and REST Complex sites reflect the populations with different levels of REST recruitment from a single underlying population of sites.

We next investigated whether REST peak height has an impact on gene repression. To discover REST regulated genes, we used RNA interference to knock down REST mRNA in ESC, and measured resultant changes in gene expression by microarray. In response to shRNA-mediated REST knock-down, expression of 721 and 352 genes increased or decreased, respectively, at a corrected significance value of $P < 0.01$. If we assume that "direct" target genes of REST are those that have a binding site within 10 kb of the transcription start site and are up-regulated in response to REST knock-down, then 5% of REST Alone target genes are regulated, compared with 9% of REST Complex target genes (Fig. 4B). The increased repressive capability of REST Complex sites was further demonstrated when we examined the magnitude of gene repression. We ranked genes by fold change in response to REST knock-down, and then used a sliding window to examine the density of direct targets of REST (Fig. 4C). As shown previously (Johnson et al. 2008), this method clearly shows that REST direct targets are significantly derepressed by shRNA treatment, while no evidence exists for any activator function of REST. Likewise, plotting

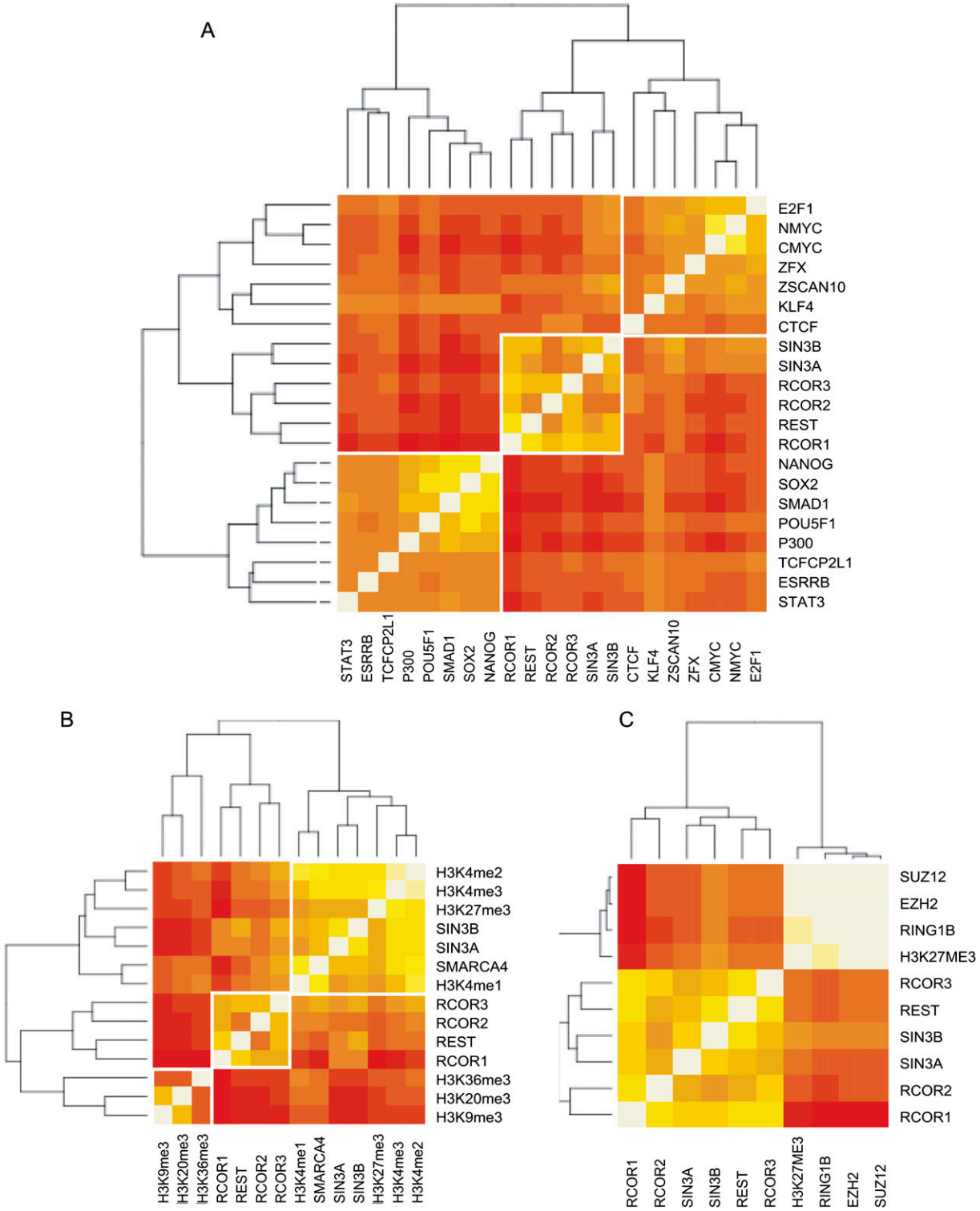

Figure 3. Discerning regulatory complexes by clustering. The heatmap represents the degree of overlap of each ChIP-seq set with every other. Color (yellow as strongest and red as weakest) reflects the intensity of overlap. The data were clustered by Pearson correlation. REST and cofactors were clustered with coordinates of (*A*) 13 pluripotency-associated TF, and P300 (Chen et al. 2008), (*B*) various covalent histone modifications (Mikkelsen et al. 2007), and (*C*) the Polycomb silencing complex, with associated repressive histone modifications (Ku et al. 2008).

genes targeted by the individual cofactors showed that REST responsive genes were similarly enriched for those targeted by cofactors (Fig. 4C).

We reanalyzed shRNA gene-expression profiles using the sliding-window method, this time distinguishing between REST Alone and REST Complex target genes (Fig. 4D). In order to

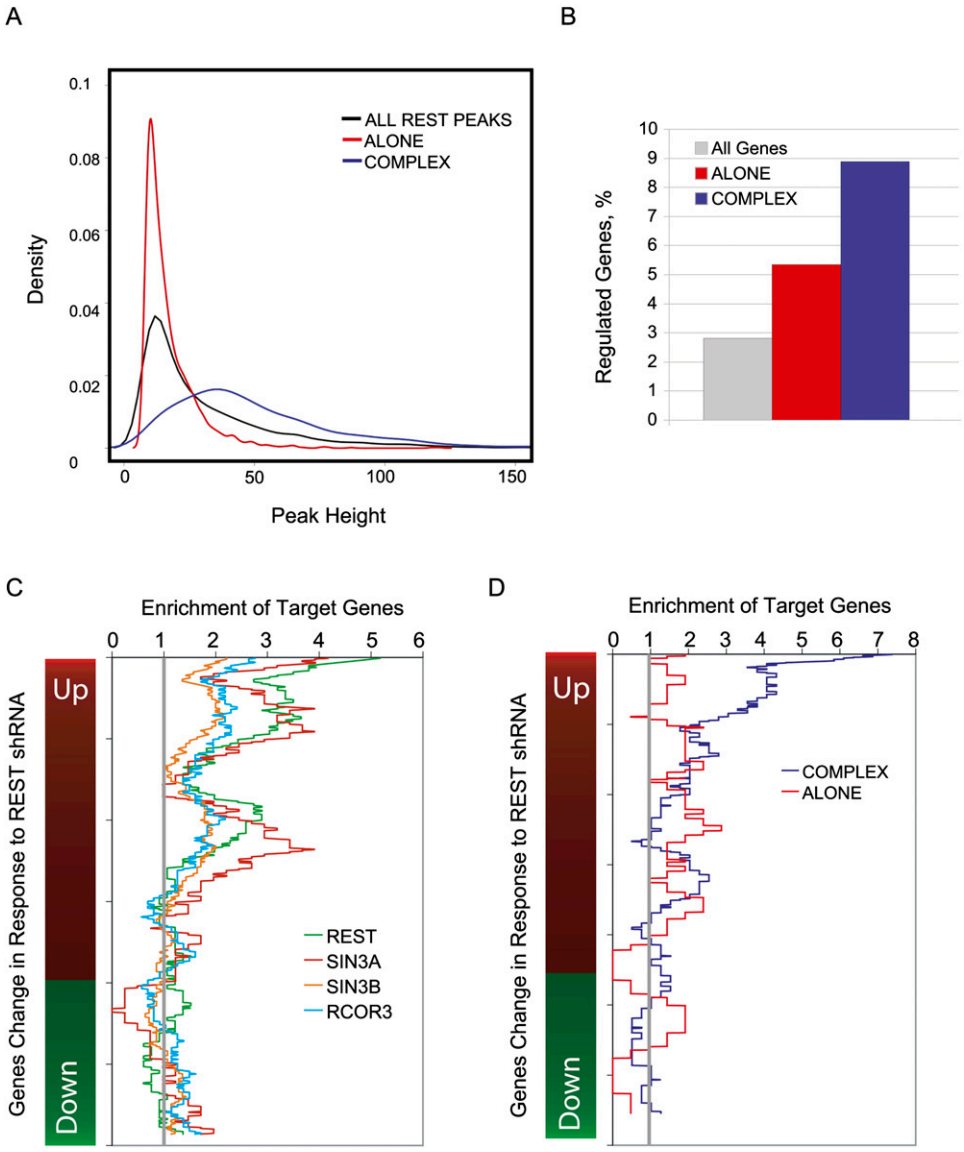

Figure 4. Contribution of genomic binding sites to direct gene regulation. (*A*) Distributions of reported peak heights for all REST peaks (black line), REST Complex peaks (blue), and REST Alone peaks (red). (*B*) The percentage of genes significantly regulated ($P < 0.01$) in response to shREST treatment are shown for all genes on the microarray (gray), all genes having a REST Alone peak within 10 kb of their TSS (red), and all genes having a REST Complex peak within 10 kb of their TSS (blue). ChIP-seq loci were combined with REST-shRNA microarray gene expression to estimate direct gene regulation. Genes were defined as "bound" if a ChIP-seq peak was detected in a 40-kb window centered on the annotated transcription start site. All genes changing significantly (up or down) in response to REST shRNA were ranked based on their fold change. A sliding window was moved down the ranked gene list, and the number of bound genes was calculated and expressed as fold enrichment over background. Background was defined to be the average density of bound genes, for all genes in the genome. (*C*) The density of genes bound by REST and cofactors is shown. (*D*) REST ChIP-seq data was divided into REST Alone and REST Complex categories. In order to avoid any bias of spurious noise sites toward the former, we only included those peaks containing an RE1 motif.

remove any possibility that errors in the ChIP-seq method might introduce unequal bias between REST Alone and REST Complex sets, we only considered sites that contained an RE1 motif. Strikingly, targets of the REST Complex sites clustered amongst the most strongly repressed genes. REST Alone target genes were generally more weakly repressed, although still at a level exceeding that expected by random chance. Therefore, the majority of direct gene repression is mediated by strongly bound REST Complex sites, while the weak REST Alone loci appear to make a lesser contribution to the cell's transcriptional regulation.

REST Alone and Complex sites represent populations of binding sites with distinct features

REST binding sites can be distinguished on the basis of a number of interrelated features, including binding strength, target gene proximity, evolutionary age, and variation amongst cell types (Bruce et al. 2009; Johnson et al. 2009). It is possible that these features distinguish REST Alone and REST Complex loci. We first examined the sequence basis of REST recruitment. The REST recognition motif, the RE1, can be identified using PSSM and assigned a score based on its "quality" with respect to an idealized

binding motif (Johnson et al. 2006). By scanning the sequence around the REST peaks, we found, as expected, that both REST Alone and REST Complex sets contain far stronger RE1 motifs than expected by chance (Fig. 5A). However, REST Alone sites, while nonrandom in nature, contain significantly weaker RE1 motifs. We hypothesized that cryptic signals may exist in RE1 motifs (particularly in the more weakly constrained central or peripheral nucleotides) to specify the recruitment, or not, of REST cofactors. However, we found no evidence for this, nor did we find significant differences in the frequency or length of non-canonical alternate spacer elements that separate the two half sites of the RE1 (Supplemental Fig. S3).

REST Alone sites were found more distally to target gene transcriptional start sites (Fig. 5B). Likely as a result of this, REST Complex sites were more than twice as likely to lie within regions of active epigenetic marks, di- or trimethylated histone H3K4 (Fig. 5C). Perhaps most significant was the finding that for those sites containing a RE1 motif, the REST Alone sites displayed significantly reduced evolutionary conservation (Fig. 5D; Supplemental Fig. S7).

REST peak height as a predictor of gene repression

We previously found that the distance of a REST binding site from a target gene's promoter strongly influences the degree of transcriptional repression (Johnson et al. 2008). We observed the same phenomenon with our present data (Fig. 6A): Genes having a REST ChIP-seq peak within ~2 kb of their transcriptional start site are far more likely to be repressed than those more distal. Nevertheless, only a minority of genes having a proximal REST peak are repressed. We looked for evidence that epigenetic status or transcriptional activity might influence the ability of proximal REST recruitment to repress a gene. First, we found no difference in the mRNA levels between repressed and nonrepressed genes, ruling out absolute gene activity in predicting repression (data not shown). Second, we investigated whether epigenetic status of genes has any impact on REST repression by using available genome-wide, chromatin methylation data (Mikkelsen et al. 2007). We anticipated that "poised" genes might be preferentially targeted by REST, since they tend to be switched on during the course of development. However, we could find no relationship between a gene having "bivalent" chromatin (having both H3K4 and H3K27 methylation) at its promoter, and its probability of being repressed by REST. Finally, we hypothesized that simply the magnitude of REST binding, as measured by peak height, might influence gene repression. Thus, we divided REST ChIP-seq peaks according to their height (low, medium, high, and very high) and plotted the fraction of genes that are repressed in each category (Fig. 6B). This analysis showed that REST peak height correlated with repression, and that genes having a strong REST peak in their promoter-proximal region were more likely to be repressed than those with a weak peak. Surprisingly, this trend was not observed when considering the very most strongly bound genes ("Very High Peak"). Furthermore, even in the best case, only 13% of bound genes were, in fact, repressed. In summary, while the degree of REST recruitment is no doubt important, other factors downstream of REST and cofactor recruitment determine whether or not a gene is to be repressed or not.

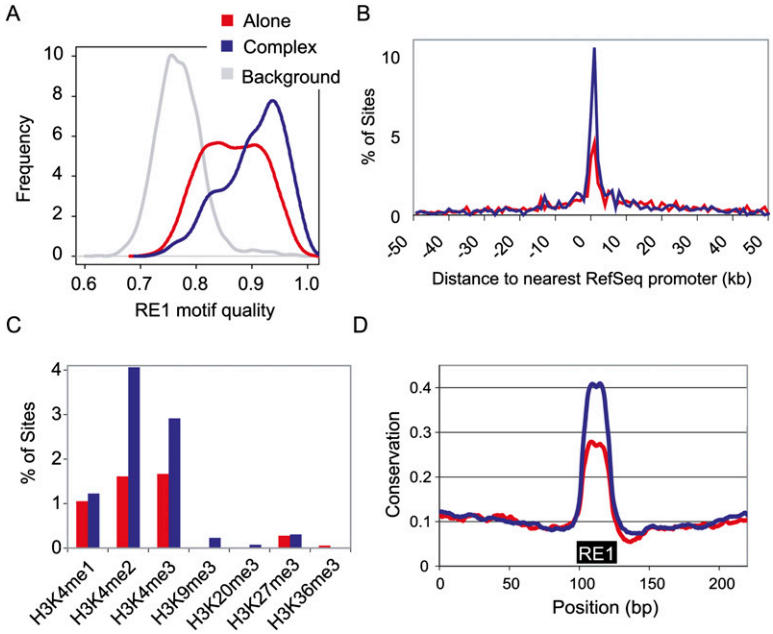

Figure 5. Comparing REST Alone and Complex binding sites. (*A*) REST Alone sites have weaker RE1 motifs compared with Complex. We scanned all peaks for RE1 motifs. Motifs were scored based on their similarity to the ideal, where 1 indicates a perfect match. For every REST Alone site, we extracted an equally sized genomic region immediately adjacent to it, and scanned these for RE1s in the same way ("Background"). (*B*) Proximity of REST sites to gene TSS. The distance of all REST binding sites with respect to the nearest RefSeq gene TSS was calculated. Negative values indicate upstream of TSS. (*C*) REST binding sites were overlapped with the coordinates of the various histone modifications indicated (Mikkelsen et al. 2007). (*D*) Evolutionary conservation of the RE1 motif was greater for Complex REST sites than Alone sites. Degree of conservation is displayed over 200 bp of genomic sequence spanning RE1 sites.

Discussion

TFs are typically recruited to many thousands of distinct binding sites in the genome, but in only a small fraction of cases is the occupancy of a TF at the target gene sufficient to activate or repress the expression of that gene. For example, we identified from more than 3000 genes that are targeted by REST in ESC only a minority of those genes that bound REST nearby are functionally repressed by it (Johnson et al. 2008). Thus, the genomic recruitment of a transcriptional repressor is not sufficient for gene repression. To address this issue, we sought to determine whether all REST binding sites recruit transcriptional cofactors similarly in ESC. Our results demonstrate that recruitment of REST and its cofactors is not uniform—half the sites strongly recruit REST together with its cofactors, while little or no cofactor recruitment could be detected at the remaining sites within the sensitivity of our method. Most importantly, we found by integrating gene expression data from REST knock-down cells that the sites that robustly recruit REST cofactors are frequently associated with genes that are strongly repressed. Alternatively, those genes that recruit REST alone, without cofactors, are typically not

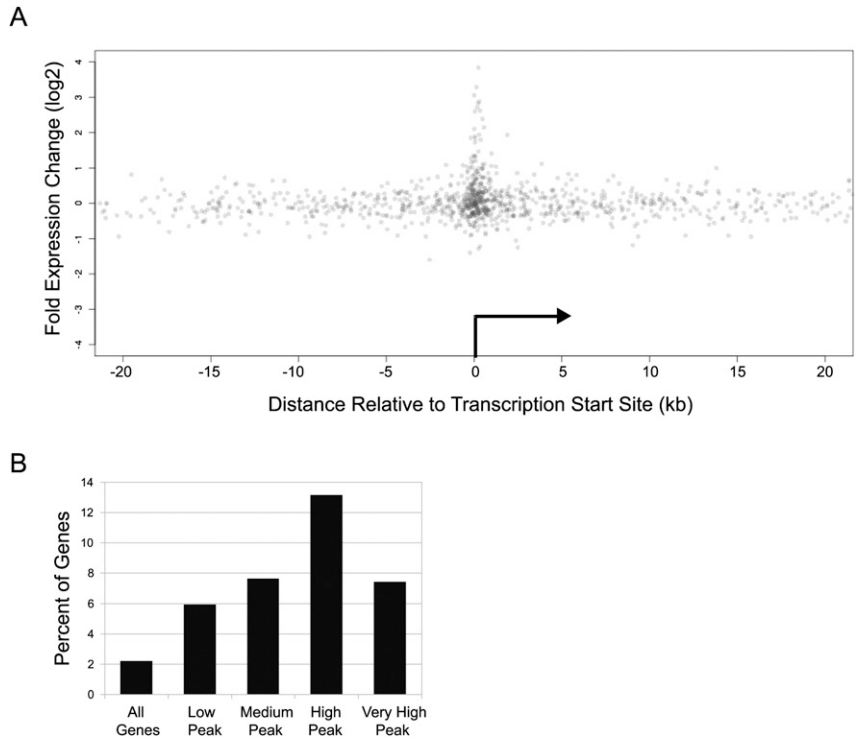

Figure 6. Factors influencing the probability of gene repression by REST. (*A*) For every gene, the unique closest REST peak was identified and plotted, such that the *x*-axis indicates the relative position of the peak with respect to the TSS, and the *y*-axis indicates the log2 fold change in gene expression in response to shREST. Height ranges, as reported by the SISSRS program, were defined as the following: Low (peak height 0–20), Medium (20–40), High (40–60), Very High (60+). (*B*) For all genes having a REST peak within 10 kb of their TSS, the proportion significantly increasing ($P < 0.01$, corrected) in response to shREST are shown.

We have broadly grouped the REST binding sites as "Complex" and "Alone" based upon our ability to detect the presence of the REST cofactors with our ChIP methodology. However, it could be argued that there are not two types of sites, but rather a continuum of sites that recruit REST with varying degrees of affinities. Our results are consistent with the notion that the "Alone" sites are simply lower affinity REST binding sites that do not efficiently recruit REST and thereby renders detection of weak cofactor recruitment below the sensitivity of our method. Indeed, this fits with our model that strong REST binding to the high affinity RE1 sites is the driver for coassembly of the corepressor complex. Our data do not indicate, nor are we arguing, that it is the presence of cofactors that lead to binding-site selection by REST. Thus, one could alternatively label these REST sites as "Strong" and "Weak" instead of "Complex" and "Alone." Importantly, our data show that it is the "Strong" (aka "Complex") sites that assemble the REST repressosome and functionally repress transcription.

It has been suggested that REST plays an essential role in maintaining ESC pluripotency (Singh et al. 2008). However, we and others have found that knockdown or knockout of REST does not lead to a loss of pluripotency (Buckley et al. 2009; Jorgensen et al. 2009b; Yamada et al. 2010). Thus, although REST is expressed at high levels and is a direct target of the POU5F1 enhanceosome in ESC (Boyer et al. 2005; Loh et al. 2006) the role of REST in ESC remains unclear. In spite of the large number of genes targeted directly by REST in ESC, REST selectively represses only a subset of these target genes (Johnson et al. 2008; Jorgensen et al. 2009b). The data presented here show that many genes, although occupied by under direct repression by REST. We also found that the degree of repression scales with the number of REST cofactors present. These data support a model where gene repression by REST requires the assembly of a mulitmeric "repressosome" complex, and that binding by REST alone is not sufficient to repress gene expression (Fig. 7).

What might be the molecular basis for differential recruitment of REST cofactors to REST binding sites? Perhaps not surprisingly, the most likely explanation from our data is that the apparent cofactor recruitment is a function of REST recruitment. Thus, the stronger REST peaks, as determined by the total number of ChIP-seq reads, tend to also recruit detectable amounts of one or more cofactors. It should be noted that cofactor binding was observed at the REST Alone sites, albeit much more weakly, indicating that it is the strength of the REST association that is the critical parameter and not the presence/absence of the cofactors, per se. We examined the sequence motif for both REST Complex sites and REST Alone binding sites and found that the Complex sites were, on average, more likely to conform to the strong RE1 consensus motif, whereas the REST Alone sites match less well with the consensus binding motif. Thus, it is possible that REST binds more weakly to certain sites, which renders them less conducive to stable recruitment of cofactors. We also note that this was not an exhaustive assessment and that there are other known corepressors of REST, which may play more critical roles in selective gene repression than the factors we evaluated in this study.

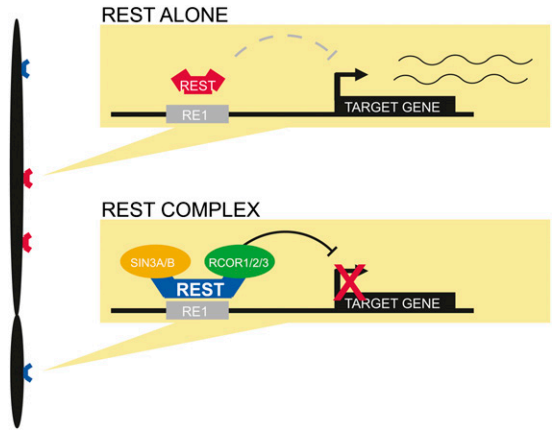

Figure 7. A general model of cofactor-dependent gene repression by REST in ESC.

REST, are not functionally repressed due to a lack of cofactor recruitment to these sites in ESC. It remains to be determined whether there is a similar distribution of REST Alone and Complex sites in neural stem cells (NSC). In NSC, knockdown of REST is sufficient to broadly derepress neuronal gene expression, which in turn leads to neural differentiation (Chen et al. 1998). It will be interesting to determine the occupancy of the cofactors at the REST binding sites in NSC. We hypothesize that many more RE1 sites will be REST Complex sites in NSC and, thus, REST will actively repress these target genes.

Our findings that target gene regulation requires co-binding of REST and its cofactor may not be a feature unique to this repressor. With the recent mapping of target sites for many TFs, it has been a conundrum that so few of the genes directly targeted by these factors are transcriptionally responsive to the binding of that factor. The data presented here highlight that TF occupancy is not sufficient to predict target gene regulation. Thus, to elucidate gene regulatory networks, it is critical that genome-wide mapping by ChIP be coupled with assessments of differential gene expression in response to knockdown of transcriptional regulators, and the recruitment of transcriptional cofactor complexes.

Methods

Cell lines

The mouse ESC cell line, E14 (American Type Culture Collection), was cultured under feeder-free conditions and shRNA knockdowns performed as described (Lim et al. 2007). Pooled, stable cell lines were generated that express V5-tagged REST cofactors RCOR1/2/3 and SIN3B. E14 cells were infected with lentiviral vectors and cells with a low level of GFP (expressed from IRES) were sorted and expanded for ChIP experiments.

Binding site identification by ChIP-seq

ChIP for REST and cofactors were carried out using protocols described (Birney et al. 2007; Yu et al. 2009). IP were carried out using the following antibodies: anti-REST (Upstate), anti-SIN3A (Santa Cruz, AK-11), anti-V5 (Invitrogen, R960-25). Approximately 20 ng of DNA was processed for ChIP-seq using the Illumina protocol and reagents. Sequencing was carried out using Illumina Genome Analyzer, and reads were mapped to mouse genome (Version mm8). The SISSRS program (Version 1.4) was used to detect peaks of ChIP enrichment (Jothi et al. 2008), by comparison to a control data set from ChIP using an anti-GFP antibody in the same cell line. SISSRS was run with the following settings: genome length 2,700,000,000, mappable fraction of genome 0.80, E value cutoff 10, *P*-value cutoff 0.001.

Co-binding analysis

We performed similar analysis as described (Chen et al. 2008) to analyze the co-binding of REST and its cofactors with various TF. The location of each binding site was defined to be a 100-bp region centered on its ChIP-seq peak. We calculated the number of intersects of each TF's peaks with the other TF, within 100 bp. We then used the R program to generate a Pearson correlation matrix to visualize the result. Locations of TF binding and histone modifications in mouse ESC were obtained from previous studies (Mikkelsen et al. 2007; Chen et al. 2008; Ku et al. 2008; Meissner et al. 2008). We also separated the REST binding sites into two groups: REST "Complex" sites that have at least a cofactor binding within 100 bp, and REST "Alone" sites that do not have any detected cofactor peaks within 100 bp. To check for possible artifacts in clustering arising from the recruitment of cofactors to factors other than REST, all clusterings were repeated with REST, but in the absence of cofactors: No substantial differences were observed in resultant clusters (Supplemental Fig. S8).

Motif analysis

To find the motifs we used repeat-masked sequences from 100 bp surrounding the top 200 peaks of each factor as input for the MDmodule program. When the program was instructed to return a 21-bp motif, all of the libraries returned the known REST motif, except for SIN3B library, which did not produce any result. When instructed to return a 11-bp result, again, all of the libraries except for SIN3B returned the REST right- and left-half sites. The SIN3B library returned a novel motif. To quantify the numbers of peaks containing RE1 motifs, we used the program Seqscan (Johnson et al. 2006) to score the best RE1 motif in a 100-bp window centered on each peak, using a relaxed motif quality score cutoff of 0.88. For analysis of noncanonical RE1 motifs, we generated a range of RE1 PWM matrices with internal spacers of $n = 0–12$ bp. All REST peaks were scanned with each PSSM, and the highest scoring motif was retained. Only peaks containing a motif scoring greater than 0.91 were considered.

Microarray expression analysis

Gene expression arrays (Mouse Ref-8 Expression Bead Chip, Illumina) were used to profile gene expression in cells where REST mRNA was knocked down by shRNA. The background-adjusted results were normalized using MeV by performing log2 transformation, followed by median centering on samples and median centering of genes across the samples (Saeed et al. 2003). *T*-test with *P*-value <0.01 and fold cut-off >1.5 was applied to find genes with statistically significant changes in expression.

REST target gene analysis

Direct target genes were defined to be RefSeq genes that (1) had a REST ChIP-seq peak lying within a 20-kb window centered on the annotated transcriptional start site, and (2) up-regulated at least twofold in response to REST shRNA.

Data access

The data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) under accession no. GSE28289.

Acknowledgments

This work was funded by the Singapore Agency for Science, Technology, and Research (A*STAR). We thank Noel Buckley and Shyam Phabhakar for their insightful intellectual contributions, and Kee-Yew Wong, Zhei-Hwee Yap, Chin-Thing Ong, Gireesh Bogu, and Felicia Hong for expert technical support.

References

Andres ME, Burger C, Peral-Rubio MJ, Battaglioli E, Anderson ME, Grimes J, Dallman J, Ballas N, Mandel G. 1999. CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc Natl Acad Sci* **96:** 9873–9878.

Ballas N, Grunseich C, Lu DD, Speh JC, Mandel G. 2005. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* **121:** 645–657.

Battaglioli E, Andres ME, Rose DW, Chenoweth JG, Rosenfeld MG, Anderson ME, Mandel G. 2002. REST repression of neuronal genes requires components of the hSWI.SNF complex. *J Biol Chem* **277:** 41038–41045.
Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, 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.
Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816.
Boyer L, Lee T, Cole M, Johnstone S, Levine S, Zucker J, Guenther M, Kumar R, Murray H, Jenner R, et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122:** 947–956.
Boyer LA, Mathur D, Jaenisch R. 2006. Molecular control of pluripotency. *Curr Opin Genet Dev* **16:** 455–462.
Bruce AW, Lopez-Contreras AJ, Flicek P, Down TA, Dhami P, Dillon SC, Koch CM, Langford CF, Dunham I, Andrews RM, et al. 2009. Functional diversity for REST (NRSF) is defined by in vivo binding affinity hierarchies at the DNA sequence level. *Genome Res* **19:** 994–1005.
Buckley NJ, Johnson R, Sun YM, Stanton LW. 2009. Is REST a regulator of pluripotency? *Nature* **457:** E5–E6.
Chen ZF, Paquette AJ, Anderson DJ. 1998. NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nat Genet* **20:** 136–142.
Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, 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.
Greenway DJ, Street M, Jeffries A, Buckley NJ. 2007. RE1 Silencing transcription factor maintains a repressive chromatin environment in embryonic hippocampal neural stem cells. *Stem Cells* **25:** 354–363.
Johnson R, Gamblin RJ, Ooi L, Bruce AW, Donaldson IJ, Westhead DR, Wood IC, Jackson RM, Buckley NJ. 2006. Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication. *Nucleic Acids Res* **34:** 3862–3877.
Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316:** 1497–1502.
Johnson R, Teh CH, Kunarso G, Wong KY, Srinivasan G, Cooper ML, Volta M, Chan SS-l, Lipovich L, Pollard SM, et al. 2008. REST regulates distinct transcriptional networks in embryonic and neural stem cells. *PLoS Biol* **6:** e256. doi: 10.1371/journal.pbio.0060256.
Johnson R, Samuel J, Ng CK, Jauch R, Stanton LW, Wood IC. 2009. Evolution of the vertebrate gene regulatory network controlled by the transcriptional repressor REST. *Mol Biol Evol* **26:** 1491–1507.
Jorgensen HF, Chen ZF, Merkenschlager M, Fisher AG. 2009a. Is REST required for ESC pluripotency? *Nature* **457:** E4–E5.
Jorgensen HF, Terry A, Beretta C, Pereira CF, Leleu M, Chen ZF, Kelly C, Merkenschlager M, Fisher AG. 2009b. REST selectively represses a subset of RE1-containing neuronal genes in mouse embryonic stem cells. *Development* **136:** 715–721.
Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. 2008. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res* **36:** 5221–5231.
Kim J, Chu J, Shen X, Wang J, Orkin SH. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132:** 1049–1061.
Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, et al. 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4:** e1000242. doi: 10.1371/journal.pgen.1000242.
Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, et al. 2008. Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* **6:** e27. doi: 10.1371/journal.pbio.0060027.
Liang J, Wan M, Zhang Y, Gu P, Xin H, Jung SY, Qin J, Wong J, Cooney AJ, Liu D, et al. 2008. Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol* **10:** 731–739.
Lim LS, Loh YH, Zhang W, Li Y, Chen X, Wang Y, Bakre M, Ng HH, Stanton LW. 2007. Zic3 is required for maintenance of pluripotency in embryonic stem cells. *Mol Biol Cell* **18:** 1348–1358.
Loh Y, Wu Q, Chew J, Vega V, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, et al. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38:** 431–440.
Lunyak VV, Burgess R, Prefontaine GG, Nelson C, Sze SH, Chenoweth J, Schwartz P, Pevzner PA, Glass C, Mandel G, et al. 2002. Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* **298:** 1747–1752.
Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454:** 766–770.
Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553–560.
Ooi L, Wood IC. 2007. Chromatin crosstalk in development and disease: lessons from REST. *Nat Rev Genet* **8:** 544–554.
Ooi L, Belyaev ND, Miyake K, Wood IC, Buckley NJ. 2006. BRG1 chromatin remodeling activity is required for efficient chromatin binding by repressor element 1-silencing transcription factor (REST) and facilitates REST-mediated repression. *J Biol Chem* **281:** 38974–38980.
Roopra A, Sharling L, Wood IC, Briggs T, Bachfischer U, Paquette AJ, Buckley NJ. 2000. Transcriptional repression by neuron-restrictive silencer factor is mediated via the Sin3-histone deacetylase complex. *Mol Cell Biol* **20:** 2147–2157.
Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, et al. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34:** 374–378.
Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119:** 941–953.
Singh SK, Kagalwala MN, Parker-Thornburg J, Adams H, Majumder S. 2008. REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature* **453:** 223–227.
Sun YM, Greenway DJ, Johnson R, Street M, Belyaev ND, Deuchars J, Bee T, Wilde S, Buckley NJ. 2005. Distinct profiles of REST interactions with its target genes at different stages of neuronal development. *Mol Biol Cell* **16:** 5630–5638.
Tahiliani M, Mei P, Fang R, Leonor T, Rutenberg M, Shimizu F, Li J, Rao A, Shi Y. 2007. The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature* **447:** 601–605.
Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457:** 854–858.
Yamada Y, Aoki H, Kunisada T, Hara A. 2010. Rest promotes the early differentiation of mouse ESCs but is not required for their maintenance. *Cell Stem Cell* **6:** 10–15.
Yeo M, Lee SK, Lee B, Ruiz EC, Pfaff SL, Gill GN. 2005. Small CTD phosphatases function in silencing neuronal gene expression. *Science* **307:** 596–600.
Yu HB, Kunarso G, Hong FH, Stanton LW. 2009. Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. *J Biol Chem* **284:** 31327–31335.
Zheng D, Zhao K, Mehler MF. 2009. Profiling RE1/REST-mediated histone modifications in the human genome. *Genome Biol* **10:** R9. doi: 10.1186/gb-2009-10-1-r9.

Received August 30, 2010; accepted in revised form May 24, 2011.