Supplementary Figure Legends

Supplementary Figure S1: Expression levels of tagged and endogenous proteins

Western blot analyses were performed on ESC expressing V5-tagged REST cofactors; RCOR1, RCOR2, RCOR3, and SIN3B (as indicated) and cells transfected with empty vector (Vec). Antibodies specific for V5, RCOR1, RCOR2, RCOR3, SIN3B, and ACTINB were used to detect levels of exogenous and endogenous proteins. Results indicate that V5-tagged protein levels were similar to endogenous protein levels.

Supplementary Figure S2 : Peak finding.

(A) Using motif enrichment to compare peak-finding methods. To compare the efficacy of peak finding algorithms, we performed the following analysis. We reasoned that peaks containing REST recognition motifs, RE1s, are likely to be bona fide binding sites, whereas those without RE1s are more likely to be spurious peaks. The REST ChIP-seq library was subjected to peak-finding using three independent algorithms: SISSRs (Jothi et al. 2008), CCAT (Xu et al.), MACS (Zheng et al. 2009). The discovered peaks were ranked by P-value. The genomic DNA surrounding each binding peak was extracted and searched for RE1 motifs using Seqscan (Johnson et al. 2006). A sliding window was moved across the ranked peaks, and the mean RE1 motif score across those peaks was calculated and plotted. Using this analysis, we decided to employ SISSRs, as it represented the best trade-off between (a) the number of peaks it reported, and (b) the proportion of true sites amongst those peaks.

(B) Different behavior between SISSRs-discovered peaks. We applied SISSRs with default cutoff threshold to discover binding peaks in each of the six ChIP-Seq libraries in this study. For each peak, SISSRs provides an estimated significance value, P. Peaks were ranked by P, and their values plotted as shown. The number of peaks in each library is also displayed.

Supplementary Figure S3 : Motif discovery.

(A) RE1 motif rates in ChIP-Seq libraries. The program Seqscan was used to discover RE1 motifs in 100bp windows centered on binding peaks from ChIP-Seq libraries (Johnson et al. 2006). A relaxed cutoff score of 0.88 was employed, and both canonical (internal spacer length n=2) and non-canonical (spacer length $n \leq 12$, $n \neq 2$) RE1 motifs were considered. The percentages of peaks containing at least one such RE1 are displayed. We also searched five independent sets of randomly selected genomic fragments, of equal number to the REST ChIP-Seq peak set ("Random n=5").

(B) Comparing Canonical and Non-Canonical RE1 frequencies between REST Alone and Complex sites. We separately searched 100bp windows centered on either REST Alone or Complex sites, using Seqscan at identical settings as above. Shown is the distribution of spacer widths (n), amongst the highest scoring RE1 under each peak.

(C) No evidence for signals determining cofactor recruitment within RE1 spacer region. Previous evidence suggests that transcriptional cofactor recruitment can be controlled by

apparently nonconserved nucleotides within transcription factor recognition elements (Leung et al., 2004). We extracted all high-quality (RE1 score >0.91) RE1s from REST Alone and Complex peaks, and visualized their nucleotide frequencies using Weblogo tool (Crooks et al. 2004). Among both canonical and noncanonical (n=9) RE1s, we could find no candidate nucleotides in the spacer region that might specify cofactor recruitment.

Supplementary Figure S4 : Effects of cofactor recruitment on gene expression for REST and other stem cell factors. Similar to REST Alone and Complex, we divided other known ESC regulatory factors and complexes based on the presence or absence of regulatory cofactors. We defined the target genes of these to be the nearest RefSeq gene having a particular binding peak within a 40kb window centered on its transcriptional start site. We extracted gene expression profiles for pluripotent and differentiating ESC from Ivanova et al (Ivanova et al. 2006). We plotted the mean probe log2 expression level for various gene sets, using the normalized data provided by Ivanova et al. (A) Oct4 target genes were divided according to whether Oct4 was also recruiting the P300, SIN3A or SIN3B ("SIN3A/B"), or none. (B) CTCF target genes were divided amongst those where CTCF also recruits RCOR3, or not. (C) REST Alone target genes were compared to genes where REST recruits 3, 4 or 5 cofactors. (D) MYC MTLs (Chen et al. 2008) – defined as the intersection of CMYC and E2F1 binding sites – were divided amongst those recruiting SIN3A or SIN3B, or neither.

Supplementary Figure S5: Cofactor recruitment to REST Alone sites

To test whether there is any evidence of cofactor recruitment at REST Alone sites, we extracted ChIPSeq reads mapping in a 700bp window centered on the REST peak location. The figures show the density of reads falling on each nucleotide in this region, summed for either REST Alone (left) or REST Complex (right) sites. Top panels show this analysis for REST ChIPSeq reads, lower panels show analysis for the cofactors. The grey line denotes the corresponding density for non-specific Input reads.

Supplementary Figure S6: Correlation between cofactor recruitment and REST recruitment for both REST Alone and REST Complex sites.

We investigated whether cofactor recruitment depends directly on REST recruitment. Lower panels: For every REST Alone (left) or REST Complex (right) peak, we summed the total number of cofactor sequence reads within a 700bp window centered on the peak, and plotted this value as a function of REST reads in the same window. Top panels: A control analysis was also carried out, substituting non-specific Input reads. Here no relationship was observed between Input reads and REST reads, indicating that the cofactor read enrichment is specific for recruitment by REST. All charts show log10 transformed counts of ChIPseq reads.

Supplementary Figure S7: Evolutionary conservation of REST binding sites.

To determine evolutionary conservation, we used the following method. A REST binding region of 100bp (centered on the reported binding peak) from the mouse genome was searched for RE1 motifs using Seqscan (Johnson et al. 2006). Only sites containing a canonical or noncanonical RE1 motif of moderate score (>0.88) were retained. These mouse loci were mapped to the orthologous region of the human genome (hg18), using the LiftOver tool at the Galaxy web suite

(Giardine et al. 2005) on default settings. Sites that could not be lifted over – due to absence of orthologous sequence conservation - were considered "not conserved". Human genomic regions were searched for RE1 motifs using the same procedure as above. Those sites containing at least one canonical or noncanonical RE1 (score >0.88) were considered "conserved". Those where no RE1 scoring >0.88 was discovered, were also considered "not conserved". Plotted is the percent of conserved RE1s compared to the total number of RE1-containing mouse REST peaks.

Supplementary Figure S8: Clustering analysis of REST, omitting cofactors.

In order to rule out the possibility that the recruitment of cofactors by factors other than REST could affect the clustering shown in Figure 3, we repeated the analysis in the absence of RCOR and SIN3 ChIPSeq data.

References

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res* **14:** 1188–1190.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski, L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, et al. 2005. Galaxy: A platform for interactive large-scale genome analysis. *Genome Res* **15:** 1451–1455.

Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka IR. 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442:** 533–538.

Xu H, Handoko L, Wei X, Ye C, Sheng J, Wei CL, Lin F, Sung WK. 2010. A signal-noise model for significance analysis of ChIP-seq with negative control. *Bioinformatics* **26:** 1199–1204.