



## Functional diversity for REST (NRSF) is defined by in vivo binding affinity hierarchies at the DNA sequence level

Alexander W. Bruce, Andrés J. López-Contreras, Paul Flicek, et al.

*Genome Res.* 2009 19: 994-1005 originally published online April 28, 2009

Access the most recent version at doi:[10.1101/gr.089086.108](https://doi.org/10.1101/gr.089086.108)

---

**References** This article cites 53 articles, 24 of which can be accessed free at:  
<http://genome.cshlp.org/content/19/6/994.full.html#ref-list-1>

**Open Access** Freely available online through the *Genome Research* Open Access option.

**License** Freely available online through the Genome Research Open Access option.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Copyright © 2009 by Cold Spring Harbor Laboratory Press

# Functional diversity for REST (NRSF) is defined by in vivo binding affinity hierarchies at the DNA sequence level

Alexander W. Bruce,<sup>1,5,10</sup> Andrés J. López-Contreras,<sup>2,4</sup> Paul Flicek,<sup>3,4</sup> Thomas A. Down,<sup>1,4,5</sup> Pawandeep Dhani,<sup>1,6</sup> Shane C. Dillon,<sup>1,7</sup> Christoph M. Koch,<sup>1</sup> Cordelia F. Langford,<sup>1</sup> Ian Dunham,<sup>1,8</sup> Robert M. Andrews,<sup>1</sup> and David Vetrie<sup>1,9,10</sup>

<sup>1</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, United Kingdom;

<sup>2</sup>Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Murcia, Campus de Espinardo, 30100 Murcia, Spain; <sup>3</sup>EMBL—European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom

The molecular events that contribute to, and result from, the in vivo binding of transcription factors to their cognate DNA sequence motifs in mammalian genomes are poorly understood. We demonstrate that variations within the DNA sequence motifs that bind the transcriptional repressor REST (NRSF) encode in vivo DNA binding affinity hierarchies that contribute to regulatory function during lineage-specific and developmental programs in fundamental ways. First, canonical sequence motifs for REST facilitate strong REST binding and control functional classes of REST targets that are common to all cell types, whilst atypical motifs participate in weak interactions and control those targets, which are cell- or tissue-specific. Second, variations in REST binding relate directly to variations in expression and chromatin configurations of REST's target genes. Third, REST clearance from its binding sites is also associated with variations in the REI motif. Finally, and most surprisingly, weak REST binding sites reside in DNA sequences that show the highest levels of constraint through evolution, thus facilitating their roles in maintaining tissue-specific functions. These relationships have never been reported in mammalian systems for any transcription factor.

[Supplemental material is available online at [www.genome.org](http://www.genome.org). The REST ChIP-chip data from this study have been submitted to ArrayExpress (<http://www.ebi.ac.uk/microarray-as/ae/>) under accession no. E-TABM-660.]

In eukaryotes, the in vivo binding of transcription factors (TFs) to genomic DNA is thought to be a complex process where TFs are directed to some, but not all, of their respective cognate DNA sequence motifs, thus allowing cell- or lineage-specific coordination of transcriptional programs. Factors thought to be involved in this directed activity include TF protein expression levels and TF cooperativity (Das et al. 2004), chromatin accessibility or histone modification signatures (Orphanides and Reinberg 2002; Guccione et al. 2006), nuclear localization (Nunez et al. 2008), and the local sequence environment of the genomic DNA itself (The ENCODE Project Consortium 2007). Gel-shift and similar assays have been the conventional methods of empirically testing transcription factor–DNA interactions. However, they provide only

a low-throughput in vitro read-out of TF binding at non-physiological conditions (without taking into account the considerations mentioned above). Large-scale experimental analysis of TF binding sites in multiple cell types has been greatly facilitated using chromatin immunoprecipitation (ChIP) in combination with microarrays (ChIP-chip) (Ren et al. 2000, 2002; Horak et al. 2002; Cawley et al. 2004) or high-throughput sequencing (ChIP-seq) (Impey et al. 2004; Wei et al. 2006; Johnson et al. 2007). The power of these methods, based on identifying large numbers of in vivo TF–genome interactions, has provided far and away the best opportunity to determine the role that the DNA sequence itself contributes to TF binding in eukaryotic genomes in a stand-alone fashion. Yet to date, high-affinity TF binding sites are easily validated in cross-platform comparisons, whilst low-affinity sites have been far more difficult to identify reproducibly (Euskirchen et al. 2007). Thus, the relative contributions that a full repertoire of variations of DNA sequence motifs have in directing TF binding in vivo and the downstream biological consequences of site selection have not yet been reported. This analysis has also been further compounded by issues with computational methods used to detect differences in TF binding motifs, as there are no completely “natural” distance metrics for comparing motif variants in silico.

Repressor element 1 silencing transcription factor (REST; also referred to as NRSF) is an essential vertebrate zinc finger transcriptional repressor protein (Chen et al. 1998). Its role in neuronal development is widely studied (Lunyak et al. 2002; Roopra et al. 2004; Ballas et al. 2005): REST silences neural target genes in neural progenitor cells, and its loss at terminal differentiation allows their expression in mature neurons (Ballas et al. 2005).

<sup>4</sup>These authors contributed equally to this work.

**Present addresses:** <sup>5</sup>Wellcome Trust/Cancer Research UK Gurdon Institute, The Henry Wellcome Building of Cancer and Developmental Biology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK; <sup>6</sup>University College London Cancer Institute, Paul O’Gorman Building, University College London, 72 Huntley Street, London WC1E 6BT, UK; <sup>7</sup>School of Genetics and Microbiology, Moyné Institute of Preventative Medicine, Trinity College, Dublin 2, Ireland; <sup>8</sup>EMBL—European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK; <sup>9</sup>Section of Pathology and Gene Regulation, Division of Cancer Sciences and Molecular Pathology, University of Glasgow, Western Infirmary, Dumbarton Road, Glasgow G11 6NT, UK.

<sup>10</sup>Corresponding authors.

E-mail [dlv1q@clinmed.gla.ac.uk](mailto:dlv1q@clinmed.gla.ac.uk); fax 44-0141-337-2494.

E-mail [awb41@cam.ac.uk](mailto:awb41@cam.ac.uk); fax 44-01223-334089.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.089086.108>. Freely available online through the *Genome Research* Open Access option.

However, its role in various other developmental programs and pathophysiological conditions is less well understood. These include cardiac (Kuwahara et al. 2003), hematopoietic (Scholl et al. 1996), pancreatic (Atouf et al. 1997; Martin et al. 2003; Martin et al. 2008), and smooth muscle development (Cheong et al. 2005), Huntington's disease (Zuccato et al. 2003), epilepsy (Palm et al. 1998; Garriga-Canut et al. 2006), cardiac hypertrophy (Kuwahara et al. 2003), global ischemia (Calderone et al. 2003), and cancer (Lawinger et al. 2000; Coulson 2005; Fuller et al. 2005; Westbrook et al. 2005).

REST represses target gene expression by binding to a DNA sequence motif known as the repressor element 1 (RE1; also known as NRSE) (Chong et al. 1995; Schoenherr and Anderson 1995). The RE1 is rare amongst transcription factor binding sites, being 21 bp in length (Mori et al. 1992). Based on length alone, RE1 is relatively uncommon in mammalian genomes. The complexity of the RE1 motif also provides additional scope to examine how variations of the motif contribute to REST binding affinities and function. For these reasons, REST represents a simple experimental paradigm to comprehensively study putative TF binding site interactions.

Based on the known RE1 canonical sequence, it is estimated that there are ~1000–2000 RE1s in the human genome (Bruce et al. 2004; Johnson et al. 2006, 2007), although estimates as high as nearly 10,000 have been predicted in the mouse (Otto et al. 2007). It has been speculated that the majority of RE1 sites may be occupied by REST in mouse, although only 13% of predicted RE1s have been shown to be occupied (Otto et al. 2007), a figure not that dissimilar to that reported in humans (Johnson et al. 2007). These analyses demonstrate that a majority of the predicted RE1s are associated with putative target genes that are expressed mainly in neural cell types, a finding that is not consistent with evidence linking REST to a wide variety of developmental programs. In addition, sequence variants of the RE1 motif (Johnson et al. 2007; Otto et al. 2007) may constitute a small proportion of REST/RE1 interactions in both species, although the exact role and extent of such interactions in REST function are not known.

It is against this background that we sought to address whether features of the RE1 sequence itself were key determinants in REST binding affinity and its functional diversity. To this end, we identified a broad spectrum of REST binding sites across a panel of diverse human cell lines, deduced the target genes they regulate, and determined how these interactions relate to the DNA sequence of RE1. Our results demonstrate, for the first time in a mammal, that an *in vivo* binding affinity hierarchy defines the functional diversity of a transcription factor and that this hierarchy is evident at the DNA sequence level. The findings of this study in the context of our current understanding of the molecular and functional requirements for transcription factor binding and developmental transcriptional programs are discussed.

## Results

### Approach

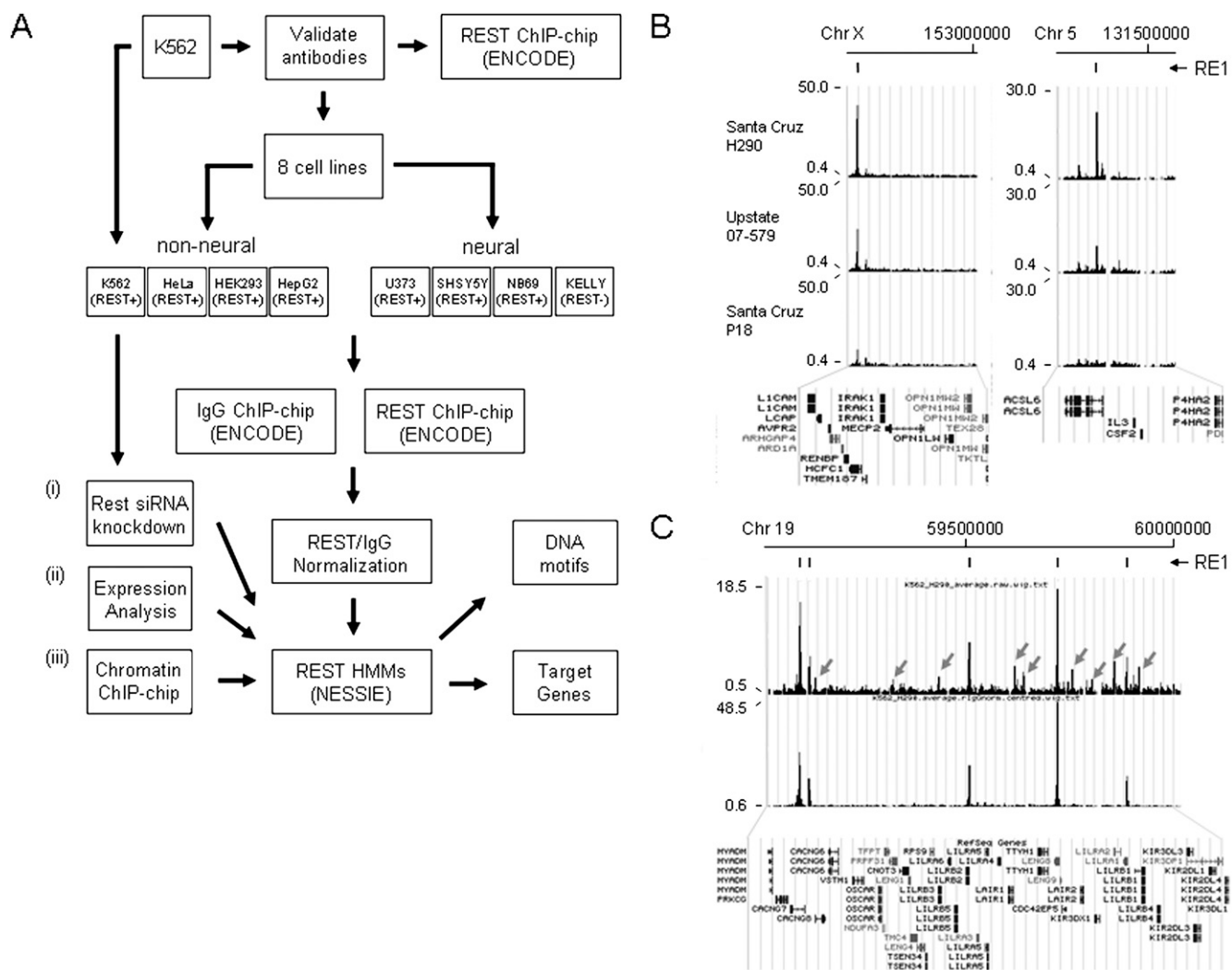
We employed a combined experimental and computational strategy that would allow us to study a variety of functional aspects related to REST binding and clearance from its sites of interaction within the human genome. An overview of the approach is summarized in Figure 1A. To ensure the highest possible confidence in our datasets, we hybridized DNA generated from chromatin immunoprecipitation (ChIP) against REST to a genomic DNA microarray comprising the 44 tiled regions of the human

genome selected for the pilot phase of the ENCODE project (The ENCODE Project Consortium 2004, 2007; Koch et al. 2007). This array utilized a PCR-product-based array technology proven to be highly sensitive and quantitative (Dhami et al. 2005). To maximize the detection of cell-type-specific interactions, we performed REST ChIP-chip across a diverse set of eight cell lines (seven REST-expressing, one REST nonexpressing), representing both neural and non-neural lineages (Supplemental Table S.1; Supplemental Fig. S.1). We optimized the REST ChIP-chip assay (i.e., tested three antibodies) and selected the best-performing antibody to profile all eight cell lines (Fig. 1B). To minimize the discovery of false-positive REST binding sites, we normalized the REST ChIP-chip datasets with corresponding datasets using a nonspecific antisera derived from the same chromatin source. This had the effect of reducing the degree of nonspecific enrichment in the assay that could be incorrectly assigned as REST binding events (Fig. 1C). Using similar approaches, we investigated gene expression, histone modification patterns, and chromatin accessibility across the ENCODE regions and related these data to REST binding patterns and target genes in the K562 cell line. siRNA-mediated knockdown of REST in combination with ChIP-chip was used to investigate clearance of REST binding sites in K562 cells. Computationally, we used a two-state hidden Markov model ("NESSIE"; Koch et al. 2007) to define REST binding sites in our datasets and "Nested-MICA" (Down and Hubbard 2005) to identify features of the underlying RE1 motifs therein.

### REST occupancy is predominantly cell- or tissue-specific

We detected a total of 591 NESSIE HMM hits for REST binding across the ENCODE regions in seven REST-expressing cell lines. We did not detect any hits from the KELLY cell line (REST-nonexpressing), providing confidence that our false-discovery rate was low (<1.2%, based on the average number of hits obtained in the other seven cell lines). The number of hits identified in each REST-expressing cell line ranged from 56 to 163 and broadly reflected levels of the REST protein (Supplemental Fig. S.1). By clustering the HMM hits, we resolved them into 250 independent sites of REST binding and determined the occupancy patterns across the seven cell lines (Fig. 2). The majority of these sites (64.8%) were closely associated with genes, either within transcribed regions (46.8%) or within 2.5 kb of promoters (18%), whilst the remainder were intergenic (Fig. 2A). Two-thirds of the sites (166) were within 1 kb of CpG islands.

REST occupied sites in both neural and non-neural cell types, with the proportion of sites being occupied uniquely in either the neural (47.6%) or non-neural (17.6%) compartments (Fig. 2B), reflecting the degree of lineage-specificity in REST binding in the cell lines we studied. The REST occupancy patterns detected in each cell type showed three discrete classes of binding sites (Fig. 2C). Sites were found to be common to all cell lines (common sites), restricted to between two and six cell lines (restricted sites), and unique to a single cell line (unique sites). Only 27 of the 250 sites were common, revealing that the vast majority of REST binding sites (89.2%) showed a degree of cell line specificity, with two-thirds of these sites showing unique occupancy in a single cell line (Fig. 2D). These data indicate that REST occupancy *in vivo* is predominantly cell- or tissue-specific, a view currently not well supported in the literature (Mortazavi et al. 2006; Johnson et al. 2007; Otto et al. 2007). There were few obvious differences between the common, restricted, and unique classes in their association with gene features and CpG islands (Supplemental Figs.



**Figure 1.** Approach to identify REST binding sites across the ENCODE regions. (A) Initially, three anti-REST antibodies were validated for enrichment levels in K562 cell lines (see also panel B). One of these, as well as rabbit IgG antisera, was used for subsequent ChIP-chip across the ENCODE regions in eight diverse human cell lines representing both neural and non-neural lineages. REST datasets were normalized with rabbit IgG datasets and resultant peaks of enrichments were identified computationally using a hidden Markov model (NESSIE). RE1 DNA motifs and target genes found at or near various classes of REST binding sites were subsequently identified. In parallel, K562 cells were also used to generate expression data and ChIP-chip data for chromatin states (histone modifications, FAIRE, and histone density) and to perform siRNA experiments against REST. (B) UCSC genome browser visualization of enrichments levels obtained in K562 cells using three anti-REST antibodies across two ENCODE regions (ENm006 and ENm002) upstream of the *L1CAM* (left) and *ACSL6* (right) genes, respectively. (Top) Genome coordinates. Enrichment scales (y-axis) in both regions are the same for all three antibodies. Performance of antibody H290 was in the range of ~2–10 times higher than the other two antibodies tested. (C) (Top panel) UCSC genome browser view of ChIP-chip enrichments (y-axis) detected across the 1-Mb ENCODE region ENm007 (genome coordinates along x-axis shown at top of panel) for raw anti-REST antibody-derived data. (Lower panel) Same data that have been normalized using data derived from the corresponding rabbit IgG dataset. (Black bars, top) The five resultant “normalized” peaks all are associated with canonical RE1 motifs (Johnson et al. 2006); (gray arrowheads) peaks of REST enrichment that were removed when correcting the data in this way.

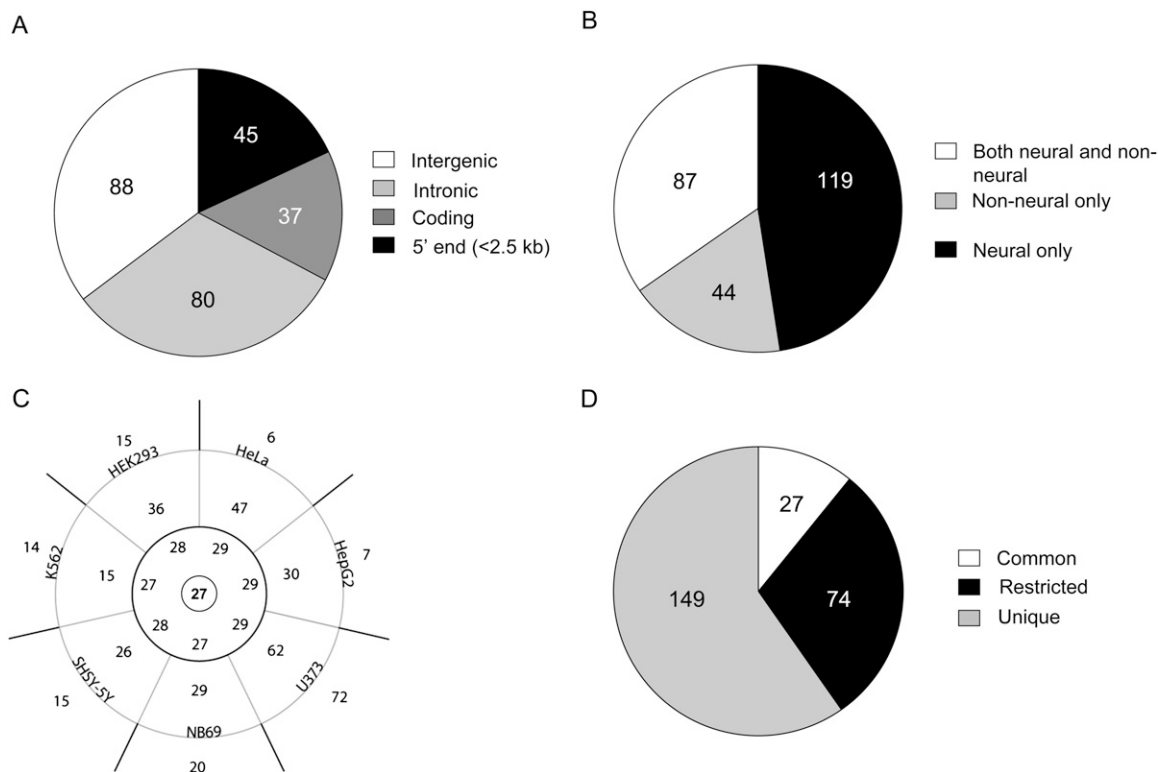
S.2, S.3), although the common class showed a twofold increase in its association with gene promoters.

Our 250 REST binding sites provide an almost 10-fold increase in the number of previously identified REST binding sites (26 sites) that had been discovered across the ENCODE regions using ChIP-sequencing in the Jurkat T-cell line (Johnson et al. 2007). We identified all 26 of these sites and could classify the vast majority of them (19) as common sites, and the remainder as restricted or unique. The vast majority (94.5%) of sites found only in our study were those that showed restricted or unique occupancy (Supplemental Figs. S.4, S.5). Thus, our view of REST occupancy

can be attributed to the identification of a large number of cell- or tissue-specific binding sites that had previously not been identified.

#### REST occupancy is reflected in its binding affinities, variations in the RE1 motif, and sequence conservation

We next examined the relationships amongst REST binding affinity, occupancy, variations in the RE1 motif (Fig. 3), and evolutionary constraints on DNA sequences where REST binds. The binding affinity of REST, as measured by enrichment levels in our ChIP-chip experiments, was directly related to occupancy patterns



**Figure 2.** REST occupancy patterns across ENCODE regions. (A) Proportions of sites in each category as determined by capturing VEGA gene features within a 2.5-kb window around each REST binding site. Numbers of sites in each category are shown. (B) Proportions and numbers of REST binding sites in neural and non-neural lineages. Sites common to both are also shown. (C) REST occupancy for each of seven cell lines using the two-state HMM algorithm NESSIE. Numbers of sites in each cell line are shown as those (1) present in only one cell line (“unique,” *outside* the circle), (2) two to six of the cell lines (“restricted,” *middle* section), or (3) common to all seven cell lines (“common,” *inner* spoke), some of which were seen at distinct sites in a given cell line, but clustered into 27 common sites when compared with sites from across all seven cell lines (central hub). (D) Relative proportions and numbers of unique, restricted, and common REST binding sites found in this study.

across cell lines. Common sites showed, on average, the highest levels of enrichment, followed by restricted and unique sites, in that order (Fig. 3A).

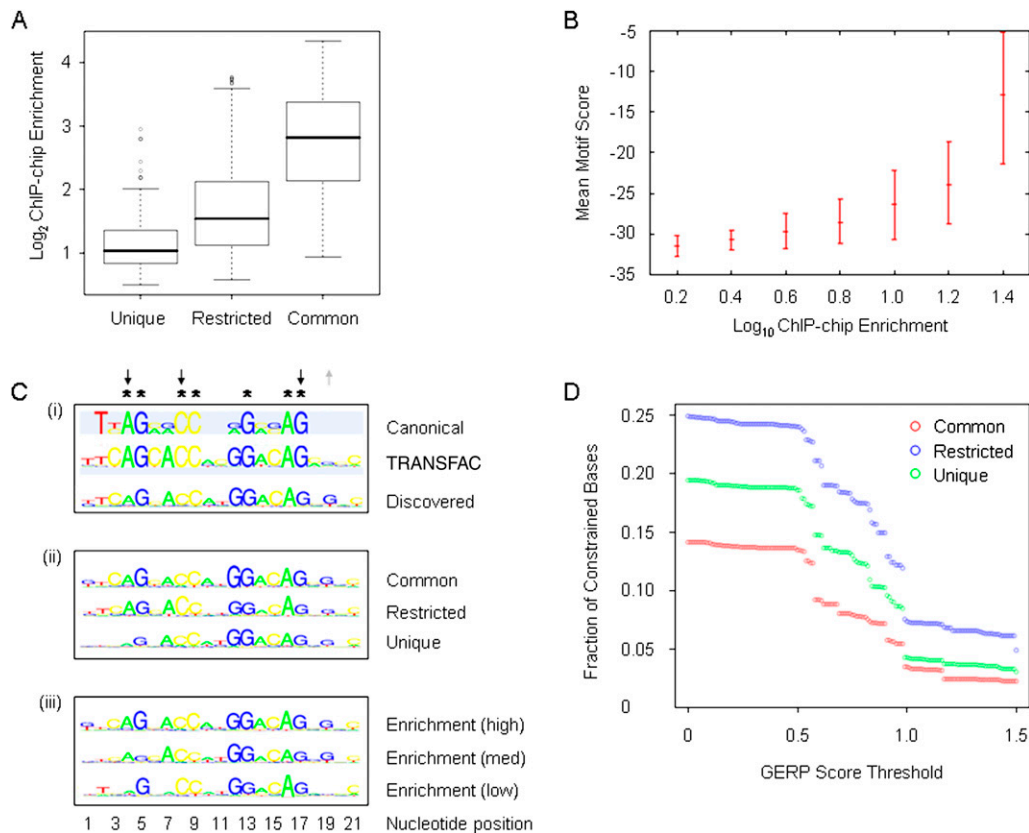
We determined the location of the highest-scoring RE1 motifs for every REST binding site; the relative strength of these RE1 motifs also showed a direct relationship with enrichment in ChIP-chip (Fig. 3B). Furthermore, whilst the consensus RE1 motif we identified from this analysis (the “discovered” motif) showed similar nucleotide constraints as the canonical or TRANSFAC RE1 motif (Fig. 3Ci), consensus motifs defined by occupancy levels showed distinct differences at a variety of nucleotide positions. Common sites most resembled the discovered and TRANSFAC 21 nucleotide RE1 motif (Fig. 3Cii). Restricted sites showed reduced or, in some cases, increased constraints at a number of base positions. Unique sites, however, showed substantial degeneracy of the left-hand portion of the RE1, and resembled an RE1 half-site (Johnson et al. 2007). Similarly, by partitioning REST binding sites into three groups based on ChIP-chip enrichments alone (high, medium, low), we also saw differences in the consensus RE1 motifs derived for each group (Fig. 3Ciii). These, by and large, showed similarities to those consensus motifs defined by common, restricted, and unique binding sites, demonstrating that binding affinity was a major determinant in distinguishing subtypes of RE1s.

Sequence conservation levels of the DNA environments in which bound RE1s were located revealed a somewhat surprising and intriguing relationship with REST occupancy. While all

regions containing bound RE1s showed higher levels of sequence conservation than would be expected at random, restricted and unique sites were found in DNA sequences that showed higher levels of sequence conservation than the common sites, which had generally higher binding affinity for REST (Fig. 3D). Thus, evolutionary pressures on RE1s may not be driven solely to attain high binding affinity and common occupancy, and additional factors, such as other TF sites, may play roles in maintaining particular types of RE1 motif structures, occupancy patterns, and binding affinities in various functional contexts.

#### REST occupancy reflects its functional diversity

We further explored these functional contexts of REST occupancy by examining the types of genes controlled by the 250 REST binding events across all seven cell lines. REST binding sites were associated with the nearest annotated genes in the ENCODE regions, and significant Gene Ontology (GO) terms were determined for REST sites classified according to occupancy in each cell line. Repression of distinct patterns of common and tissue-specific cellular functions also showed a clear relationship to patterns of REST occupancy (Table 1; Supplemental Table S.5). Common REST sites were associated with GO terms reflecting intrinsic cellular processes and neuronal function. Restricted sites were associated with lineage/tissue-specific GO terms including a wider set specifying neuronal function in the non-neural lineage



**Figure 3.** Relationship between REST occupancy, binding affinities, RE1 motifs, and sequence conservation. (A) Relationship between REST ChIP-chip enrichment (y-axis) and REST sites categorized by common, restricted, or unique cell type distribution. (Boxes) Interquartile ranges; error bars are 95% confidence intervals calculated by bootstrapping. (Open circles) Individual enrichment values that fall outside of this range. (B) The mean motif score (red dashes within error bars) for each REST binding site, defined by genome coordinates of relevant microarray tiles, was calculated by the nmscan tool of NestedMICA (y-axis) and is shown as a function of REST ChIP-chip enrichments. REST ChIP-chip enrichments were divided into 0.2- $\log_{10}$  bins. Error bars (red) indicate 95% confidence intervals on the mean motif score, obtained by bootstrapping. (C) (i) The identified motif ("Discovered") was compared with the sequences of the RE1 consensus motif ("Canonical," Bruce et al. 2004) and the RE1 motif described in the TRANSFAC database ("TRANSFAC"). The position of each nucleotide is numbered with respect to the TRANSFAC-derived sequence. (Asterisks) Nucleotides that are described as invariant in the RE1 reference motifs ("Consensus" and "TRANSFAC"). (Upward gray arrows) Some of the increased nucleotide constraints in the discovered motif, (downward black arrows) decreased nucleotide constraints. (ii) RE1 motifs identified in REST binding sites categorized by occupancy. (iii) RE1 motifs identified in REST binding sites binned by ChIP enrichment levels. (D) Evolutionary DNA sequence conservation constraints on classes of REST binding sites. The graph shows fraction of constrained bases across the entire sequence of each microarray tile containing the REST binding site (y-axis) with respect to constraint thresholds (x-axis) using the GERP conservation metric.

and terms related to interferon-mediated responses in the neural lineage. Unique REST sites were associated with GO terms found only in two of the neural cell lines (either in NB69 or U373). This, taken in the context of the relationships established above, provides a molecular basis (in terms of binding affinity, occupancy, and variations in the DNA sequence of the consensus motifs) for defining the functional diversity associated with REST within the cell.

### REST occupancy reflects chromatin structure and expression of target genes






The association between REST occupancy, binding affinity, and functional diversity also extended to REST's regulatory role, resulting in characteristic chromatin signatures across sets of REST target genes. REST enrichment levels and occupancy patterns in the K562 cell line typified those we observed in other cell lines (Supplemental Fig. S.6). Therefore, we extended our analysis of chromatin features in K562 (Koch et al. 2007) to examine these in

the context of REST binding sites and expression of its target genes. Overall, the consensus chromatin configurations of REST target genes associated with the three classes of REST occupancy patterns showed observable differences between classes (Supplemental Fig. S.7). REST sites with common occupancy were associated with genes with the lowest levels of mRNA expression (Supplemental Fig. S.8), reflected in chromatin configurations across target genes with predominantly repressive H3 K9me2 and K27me3 histone marks. This is consistent with REST's known role in gene repression. Surprisingly, for the restricted and unique classes of REST sites, target genes showed increasing levels of mRNA expression (highest for the unique class) and marked increases in levels of active H3 K4me1, K36me1 and K9ac, K18ac and K27ac histone marks, at the expense of the repressive marks. However, both active marks and the repressive H3 K27me3 mark coexisted on target genes associated with both of these classes of REST sites, mirroring signatures found in bivalent chromatin structures (Azua et al. 2006; Bernstein et al. 2006; Roh et al. 2006). Paradoxically, all classes of REST sites showed similar

**Table 1.** GO terms of REST target genes

		plasma membrane	multicellular organismal process	system process	transmission of nerve impulse	generation of a signal involved in cell-cell signaling	neurotransmitter secretion	regulation of neurotransmitter levels	synaptic transmission	plasma membrane part	response to virus	interferon binding	interferon receptor activity	interferon-alpha/beta binding	interferon-alpha/beta receptor activity	cell-cell signaling	sensory perception of light stimulus	visual perception	nitric-oxide synthase binding	basal part of cell	basal plasma membrane	caveola	
HEK293	COMMON HMMs																						
HeLa	COMMON HMMs																						
HepG2	COMMON HMMs																						
K562	COMMON HMMs																						
NB69	COMMON HMMs																						
SHSY5Y	COMMON HMMs																						
U373	COMMON HMMs																						
HEK293	RESTRICTED HMMs																						
HeLa	RESTRICTED HMMs																						
HepG2	RESTRICTED HMMs																						
K562	RESTRICTED HMMs																						
NB69	RESTRICTED HMMs																						
SHSY5Y	RESTRICTED HMMs																						
U373	RESTRICTED HMMs																						
HEK293	UNIQUE HMMs																						
HeLa	UNIQUE HMMs																						
HepG2	UNIQUE HMMs																						
K562	UNIQUE HMMs																						
NB69	UNIQUE HMMs																						
SHSY5Y	UNIQUE HMMs																						
U373	UNIQUE HMMs																						

not significant	
<0.01-0.001	
<0.001-0.0001	
<0.0001-0.00001	
0.00001-0.000001	

Statistically significant terms across REST binding sites are categorized by common, restricted, or unique cell type distribution. Data for each cell line are shown separately. Key shows level of significance for filled boxes. All terms are shown that were statistically significant for REST targets, although some remain unfilled in the data set shown here (see also Supplemental Table S.5).

repressive H3 K27me3 chromatin configurations reinforced by a lack of obvious open chromatin signatures such as depletions in nucleosome density or FAIRE peaks (Supplemental Fig. S.6).

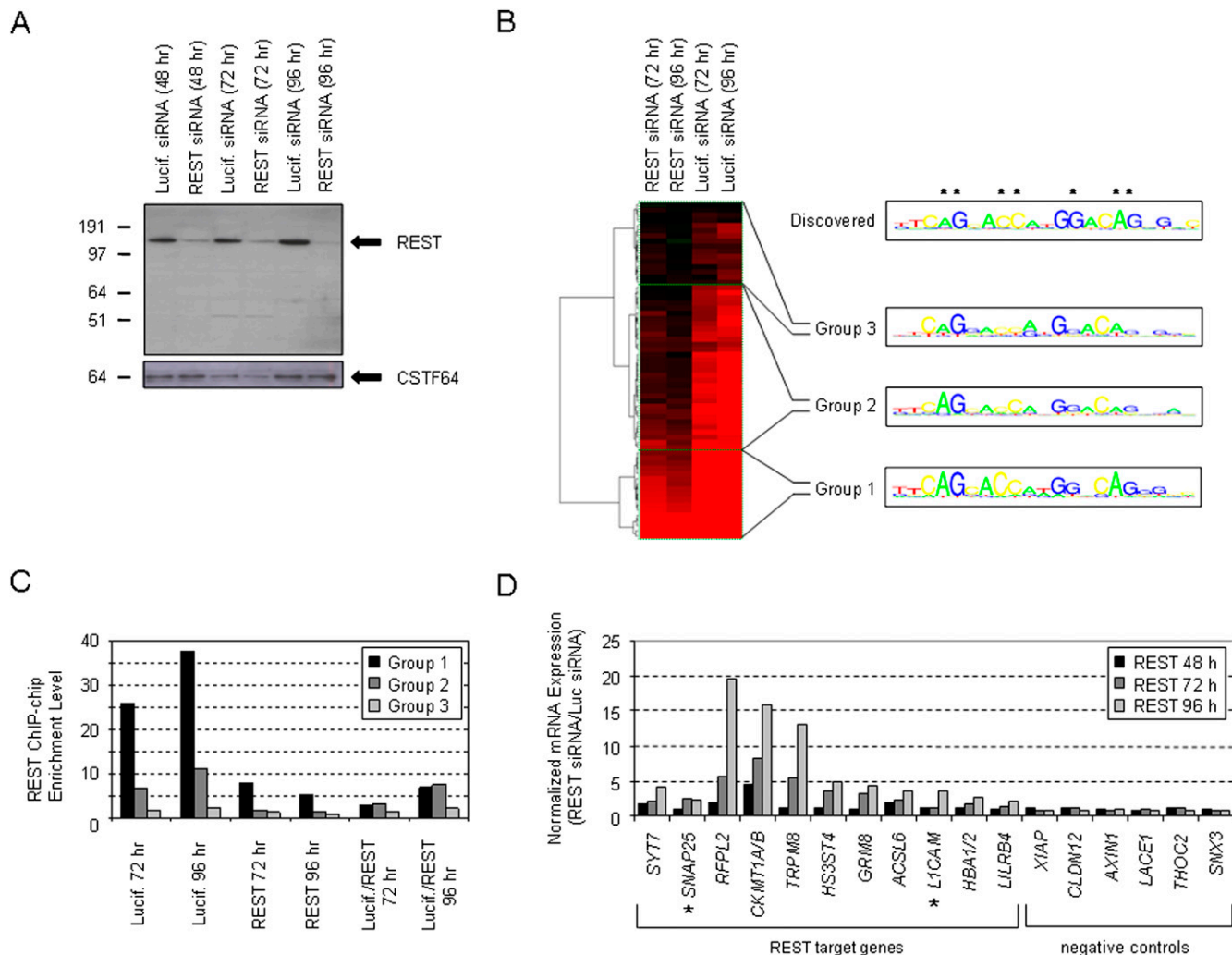
### REST clearance is defined by variations in the sequence of RE1s

It is known that within the neuronal development program, REST silences its target genes in neural progenitors, and reduction in REST expression at terminal differentiation coincides with their expression into mature neurons (Ballas et al. 2005). We used siRNA-mediated knockdown of REST to mirror this process in K562 cells, and to address what roles occupancy, binding affinity, and the RE1 motif itself play in clearance of REST from its *in vivo* binding sites and reactivation of its target genes (Fig. 4).

We performed unsupervised hierarchical clustering analysis of all microarray tiles that are normally associated with REST binding sites in K562 and examined their REST ChIP-chip enrichment levels during a time course experiment where our REST siRNA-mediated knockdown resulted in substantial removal of the REST protein (Fig. 4A). Using this two-state experimental paradigm (before and after REST knockdown), we could reliably use a clustering approach to observe the effects of both REST binding and release—and determine whether both processes are likely to be controlled by different factors, and whether these processes are

reflected in the RE1 motif. Tiles for the 56 REST enrichments in K562 that had been defined by occupancy resolved into 63 individual sites and into three distinct clusters (group 1, group 2, and group 3), each defined by distinctive versions of the RE1 motif (Fig. 4B). These groups, and the three motifs associated with them, showed features that were strikingly similar to those defined by REST occupancy (Figs. 2, 3). Group 1, 2, and 3 sites showed similar patterns of overlap with sites previously identified by ChIP sequencing, with group 1 sites sharing the highest level of overlap between the two studies (Supplemental Fig. S.9). Group 2 and 3 sites accounted for the majority of binding events (74.6% of sites). The group 1 motif most resembled the discovered motif while those of groups 2 and 3 showed the greatest deviations, particularly at a number of base positions. We were also able to detect the group 1 motif in sites with common occupancy at a higher level of significance than in the restricted and unique occupancy classes (Supplemental Fig. S.10). Group 2 and 3 sites were found with higher significance than group 1 sites in all three occupancy classes (due to the reduced levels of constraint on their motifs), but this was much more obvious for the restricted and unique occupancy classes.

Indeed, the REST groups defined by knockdown were also defined by their binding affinities for REST. Group 1 sites showed higher REST binding affinity than group 2 and group 3 sites both



**Figure 4.** Relationship between REST binding affinities, RE1 motifs, and clearance of REST from binding sites. (A) REST protein levels during siRNA knockdown. (Top panel) Nuclear protein extracts were prepared from K562 cells and subject to Western blotting using an anti-REST antibody, to confirm REST protein knockdown at 48, 72, and 96 h post-transfection with *REST* and luciferase (Lucif) siRNAs. (Lower panel) Membrane was re probed using an anti-CSTF64 antibody as a loading control. (B) Treeview diagram of hierarchical clustering of REST-enriched genomic tiles that were identified using an HMM analysis in *REST* and luciferase siRNA knockdowns in the K562 cell line. (Left) Tree diagram. Each tile is represented by a horizontal bar and is shaded in red scale or green scale to denote the level of enrichment or depletion, respectively. Knockdown experimental states are labeled at the top of each column of tiles. (Right) Recurring RE1 sequence motifs found in tiles of the three principle clusters of the tree (groups 1, 2, and 3). The overall discovered motif shown in Fig. 3 is also shown. (Asterisks) Nucleotides that are highly conserved in the discovered motif, or that were found to be invariant in previous studies. (C) Mean REST ChIP-chip enrichments (y-axis) for the three groups of REST binding sites (groups 1, 2, and 3 as defined in Results) in both luciferase and *REST* siRNA knockdown conditions (x-axis) at 72 and 96 h post-transfection. The mean fold reductions in enrichment levels at both time points (based on luciferase/*REST* ChIP-chip enrichments) are also shown at the right of the panel. (D) Induction of gene expression of REST target genes during siRNA knockdown. Total RNA was extracted from cells at 48, 72, and 96 h time points, from either *REST* or luciferase (luc) siRNA transfected K562 cells. Quantitative RT-PCR was performed using oligonucleotide primers specific to 11 known (\*) or putative REST target gene transcripts and six control genes not associated with REST binding sites. Fold changes in mRNA levels (*REST* siRNA/luciferase siRNA) are shown on the y-axis; genes are shown along the x-axis.

before and after knockdown (Fig. 4C). Group 3 sites showed the lowest affinity before knockdown and were virtually cleared of REST during knockdown, approaching baseline levels. Yet, the rate at which REST clearance occurred (measured by the ratio of ChIP-chip enrichments before and after knockdown) was not a distinguishing feature of the three groups, with group 1 and 2 sites clearing at similar rates, and complete clearance being the rate limiting factor for group 3 sites (Fig. 4C). Furthermore, target genes associated with REST sites with a range of binding affinities and RE1 subtypes showed no clear relationship in the degree to which

they were transcriptionally activated during REST knockdown (Fig. 4D). Thus, while REST binding and clearance show a clear relationship to the RE1 sequence, the kinetics of how clearance is manifested at the site of binding and on derepression of REST target genes are not obviously related to features of only the RE1.

## Discussion

There is a paucity of experimental evidence in the literature that describes *in vivo* DNA binding affinities of transcription factors as

defining functional contexts, and what has been reported is limited to specific loci (Hentsch et al. 1992). However, it is a widely held view that the DNA sequence motifs themselves are one of the major determinants of TF binding specificity, and that degeneracy at some nucleotide positions of binding sites selected through evolution may play roles in determining function (Stormo 2000). The findings of this study have begun to systematically explore these questions and issues in the human genome using a combined experimental and computational approach.

The data presented here support the existence of an *in vivo* regulatory hierarchy for one transcription factor, REST (NR5F1). This hierarchy confers functional information through DNA binding affinities encoded by variations in its RE1 binding motif as follows. Binding affinities of REST are linked to patterns of occupancy across cell lines. REST binding events that are shared in common with all cell lines studied have the highest levels of ChIP-chip enrichments, whilst weaker enrichments show restricted or unique occupancy amongst cell lines. In turn, these binding patterns and affinities relate to distinct functional groups of REST target genes across cell lines and between cell lines derived from neural and non-neural lineages. The levels of transcriptional repression exerted by REST and reflected by chromatin configurations on its target genes also show similar relationships to binding affinity. The ability to relieve targets from REST-mediated transcriptional repression is also hierarchical—sites of weak interaction are capable of completely clearing REST from their binding sites (given we could not detect binding at these sites after knockdown) whilst strong binding sites retain REST to some degree. All of these features are linked to subclasses of the RE1 site, where differences in constraints at individual DNA nucleotide positions are evident in the various subclasses. The existence of such a hierarchy for REST has direct relevance not only for our current understanding of TF binding, but for how the REST regulome is capable of orchestrating diverse and complex transcriptional programs using contextual cues encoded by genomic sequence information.

We identified 250 independent REST binding sites across the ENCODE pilot regions, only 26 of which had previously been mapped (Johnson et al. 2007). Strikingly, while we screened only 1% of the human genome, our total sample size of identified REST sites is only six- to sevenfold smaller than the total number of sites found in the whole genome (Johnson et al. 2007). Analysis across the ENCODE regions has shown to be a strong indicator of trends across the whole genome (The ENCODE Project Consortium 2007). Thus, by extrapolation, we would estimate ~25,000 REST binding sites in the human genome—a greater than 13-fold increase in the number of sites experimentally verified in either the human or mouse genomes (Johnson et al. 2007; Otto et al. 2007). These findings resonate well with the idea that the majority of the estimated 10,000 RE1 sites in the mouse genome may be occupied by REST (Otto et al. 2007), and our figures would suggest that perhaps even this may be an underestimate, if binding events at degenerate RE1 are considered in these figures. Furthermore, 89.2% of the REST–genome interactions we identified across the ENCODE regions show occupancy that is restricted to one or several, but not all, of the cell lines studied here, suggesting that the vast majority of REST binding sites are cell- or lineage-specific. This challenges the current view that REST is primarily a global repressor of neuronal gene expression (Chong et al. 1995; Schoenherr and Anderson 1995; Palm et al. 1998) in both neural and non-neural cell types. In support of the current view, we identified significant GO terms for neuronal function in the target

gene sets found in common for all of the cell lines we studied. These represented common high-affinity REST binding events with RE1 motifs most closely resembling the canonical motif. Yet, the majority of cell- or lineage-specific REST binding sites are ones we considered to be weak affinity sites with noncanonical sequence variations in RE1, and only a very small proportion of them (1.7%) have been identified by other studies (Johnson et al. 2007). Thus, it is clear why REST has not previously been seen to have much more specialized functional repertoires among cell types.

Our data and existing knowledge about the role of REST suggest that this hierarchy is governed not only by binding preferences of target sites in the genome, but also by the level of REST protein expression (Bruce et al. 2004), and assume even greater biological importance when one considers the situations in which REST levels have been shown to be dynamically regulated. As neurons differentiate from the proliferative zone of the neural tube and begin their migration to the marginal zone, REST levels decrease to virtually nondetectable (Schoenherr et al. 1996). In epileptic seizure rat models, REST levels increase rapidly in the hippocampus after the onset of seizure, after which they return to basal level (Palm et al. 1998). In rat models of ischemia, levels increase in the hippocampus (Calderone et al. 2003), and in vascular smooth muscle, fluctuating REST levels have been shown to affect their proliferative properties that in turn are thought to contribute to certain types of vascular pathology (Cheong et al. 2005). The aberrant expression of REST in pancreatic islet cells has also been implicated in diabetes (Atouf et al. 1997; Martin et al. 2003, 2008). Thus, with an improved understanding of the functional diversity of REST through the hierarchical model, we can begin to better appreciate how such cell- and tissue-specific (mis)regulation involving REST can be achieved.

We admit that whilst our knockdown data support a view of a REST hierarchy, neither the rate at which REST was lost from RE1 sites nor the rate target genes were relieved from transcriptional repression showed this same bias. We interpret this to suggest that the binding of REST to RE1 and the rate at which REST is lost are likely to be controlled in different ways and that there may be additional requirements for REST clearance or for target gene transcriptional activation (including the presence of cell-type specific regulators). Indeed, given that the variants of the RE1 motif we identified from the siRNA knockdown experiments are not entirely the same as those deduced from basal occupancy or enrichment patterns, the role that the sequence of RE1 plays in loss of REST may also be different from binding or reflect the limited sample size of RE1 motifs we analyzed in K562. Notwithstanding this, the subclasses of RE1 sites identified in the K562 cell line during knockdown allowed us to study their cognate subclasses of target genes and provide evidence of how induction of expression may be achieved for targets that bind REST with different affinities. Target genes that bind REST with low affinity, when compared with high-affinity binders, are more likely to be already expressed at low levels and have chromatin configurations with histone modifications of both active and repressed genes. These characteristics resemble genes that have “bivalent” histone modification structures (Azuara et al. 2006; Bernstein et al. 2006; Roh et al. 2006). Whilst the exact role of bivalency is unclear, it may allow some classes of developmentally regulated target genes to be primed for expression where rapid dynamic transcriptional responses are required (Bernstein et al. 2006; Roh et al. 2006). As weak REST binders also tend to be cell- or lineage-specific target genes, such “opposing” chromatin configurations may also

provide a means of control in cell-specific developmental programming. The role of weak affinity binding, in this context, has been proposed in other organisms and fits well with a hierarchical structure of transcriptional control (Gaudet and Mango 2002).

The contribution that the DNA sequence of the RE1 has in defining the binding affinity hierarchy cannot be overstated. Our data support a redefined view of what degeneracy the consensus RE1 motif can tolerate, and how this degeneracy facilitates the range of binding affinities that REST requires to fulfill its functions in the context of common or cell- and lineage-specific roles. Our clustering and motif analyses provide evidence for three types of RE1 motifs (canonical, degenerate, and half-sites), are consistent with the three-state model that we propose for occupancy (universal, cell-specific, unique), and are consistent with previous studies that define three classes of RE1 sites (Johnson et al. 2007; Otto et al. 2007). However, it was previously thought that degenerate or half RE1 sites were much less prevalent at sites of REST binding than the canonical RE1. Our motif analysis for sites showing restricted or unique REST occupancy suggests that degenerate RE1s and RE1 half-sites may account for the overwhelming majority (89.2%) of binding sites. Furthermore, the REST binding hierarchy points to these sites as those that regulate key cell- and tissue-specific target genes in developmental programs.

Based on the binding affinity framework we have described here, an important question we asked was how does the affinity of binding relate to the evolutionary constraint on RE1s and the genomic DNA sequence environment in which they reside? A logical a priori assumption we made was that the canonical RE1, represented in the high-affinity “common” subclass, would reside in sequences that are more evolutionarily constrained than RE1s that confer weak REST binding. This assumption was proven to be wrong, with the data supporting increased evolutionary constraint on DNA sequences in which weak-binding RE1s reside. Given, however, that two-thirds of our REST sites are associated with CpG islands, evolutionary constraints may be difficult to measure and may skew the results to reflect CpG content rather than occupancy. This difficulty is somewhat lessened here, as our three REST occupancy classes are represented by similar proportions of CpG island-associated REST binding sites. Nonetheless, this result is intriguing and would lend weight to the idea that RE1s that bind REST weakly are constrained to remain weak binders. Given that we see that cell- or tissue-specific function of REST is related to weak REST binding at atypical RE1s, including half-sites, this level of constraint would allow cell- or tissue-specific functions to be maintained during evolution via selective pressures on RE1s that vary according to their positions in the hierarchy and in what contextual functional roles those RE1s are involved. A previous model of the role and evolution of degenerate RE1s (Zhang et al. 2006) does not accurately account for such functional constraints.

We believe that our data and interpretations, supported by rigorous statistics, more accurately reflect the hierarchical nature of the REST regulome than has been reported previously. For a number of reasons, we believe our ChIP-chip strategy using a sensitive array platform and applied across seven different cell lines was instrumental in uncovering aspects of this hierarchy, which detected a wide spectrum of both strongly and weakly enriching REST–genome interactions, the majority of which have not been identified by ChIP-sequencing (Johnson et al. 2007). With further improvements to ChIP-sequencing platforms and

deeper sequencing across multiple cell types, we believe these limitations will be overcome for further genome-wide studies. It is only once complete spectra of binding affinity profiles for other transcription factors are uncovered across the human genome that we believe that hierarchies of this nature will begin to emerge for other transcriptional regulators.

## Methods

### Cell line culture

Eight human cell lines (K562, HEK293, HeLa, HepG2, U373, SHSY5Y, KELLY, and NB69; Supplemental Table S.1) were obtained from the European Collection of Cell Cultures ([www.hpacultures.org.uk/](http://www.hpacultures.org.uk/)). The following cell lines were all grown in RPMI 1640 media with the specified supplements: K562 (10% fetal calf serum [FCS] and 2 mM L-glutamine [LQ]), SHSY-5Y (15% FCS, 2 mM LQ, and 1% nonessential amino acids [NEAA]), NB69 (15% FCS), and KELLY (10% FCS). HEK293, HeLa, and U373 cells were grown in DMEM supplemented with 10% FCS, 2 mM LQ, and 5 units/mL penicillin/streptomycin. HepG2 cells were cultured in EMEM containing 10% FCS, 2 mM LQ, and 1% NEAA.

### siRNA transfections

Per condition,  $5 \times 10^6$  K562 cells were resuspended in 100  $\mu$ L of room-temperature serum-free RPMI 1640 media and double-stranded siRNA oligonucleotides (annealed at 100  $\mu$ M) targeted to either REST (sense; 5'-GGAGCAAGUCCUUAUUGAdTdT-3') or luciferase (5'-CUUACGCUGAGUACUUCGAdTdT-3') were added to a final concentration of 20 nM. The cells were transfected using the T16 program on an Amaxa Nucleofector and cultured for up to 96 h. For cells used for ChIP (see below), 10 individual transfections were combined to yield  $5 \times 10^7$  cells for chromatin preparation. For the 96-h time-point described in the Results section,  $5 \times 10^7$  cells were incubated for 48 h after transfection with anti-REST or luciferase siRNA, counted, and retransfected as described above in order to prepare chromatin after a further 48-h incubation.

### RNA extraction and PCR-based gene expression analysis

Total cellular RNA from the eight human cell lines was purified using TRIzol reagent (Invitrogen) and is described in full at [www.sanger.ac.uk/Projects/Microarrays/arraylab/methods.shtml](http://www.sanger.ac.uk/Projects/Microarrays/arraylab/methods.shtml). Reverse transcription was performed from 1  $\mu$ g of total RNA using the Superscript II kit (Invitrogen) according to manufacturer's protocols. The resultant cDNAs were used to quantify the levels of REST or beta-actin transcripts using gene-specific oligonucleotide primers (REST: forward-CATTCGCTGTGACCGCTG, reverse-CAATTGAACTGCCGTGGG; beta-actin; forward-CCCTGGAGAAGGATGACTG, reverse-ACATCTGCTGGAAGGTGGAC) and AmpliTaq DNA polymerase according to manufacturer's protocols (Applied Biosystems). PCR products were resolved by electrophoresis and visualized by ethidium bromide staining and UV transillumination. cDNA was prepared from siRNA-transfected K562 cells in an identical manner to assay REST target gene transcripts by quantitative real-time PCR (Supplemental Table S.2). 2 $\times$  SYBR-green PCR mastermix (Eurogentec) was used according to supplied protocols on an ABI Prism-7000 real-time PCR machine (Applied Biosystems). The  $\Delta\Delta$ Ct method was used to derive normalized fold changes (with respect to beta-actin) in transcript levels from cells transfected with anti-REST siRNA and compared with K562 cells transfected with anti-luciferase (Livak and Schmittgen 2001).

## Protein analysis

The protocol is identical to that used previously (Bruce et al. 2006). Ten million ( $10^7$ ) cells were used for nuclear protein extraction to detect REST protein expression in all eight individual cell lines, and  $5 \times 10^6$  cells were used for siRNA transfection in K562 cells. REST protein was detected using the anti-REST antibody from Upstate Biotechnology (07-579), and anti-CSTF64 antibody (sc-16473; Santa-Cruz Biotechnology) was used to detect the loading control protein CSTF64. The immunoblotted membrane was stained in diluted Bradfords reagent (Bio-Rad; one part reagent to four parts water) for 1 h at room temperature with agitation. The blot was then destained in water, dried, and photographed.

## Chromatin immunoprecipitation (ChIP)

ChIP was essentially the same as previously reported (Koch et al. 2007) with cells being cross-linked with formaldehyde (to a final concentration of 0.75% or 1%) for 10 min. When assaying REST binding in the eight different cell lines, aliquots equating to  $1 \times 10^7$  cell equivalents of chromatin were used per immunoprecipitation using 10  $\mu$ g of the relevant antibody. ChIP assays from K562 cells transfected with siRNA were performed using  $5 \times 10^6$  cell equivalents and 5  $\mu$ g of antibody. Supplemental Table S.3 details all the antibodies used. All ChIP assays were performed as three independent biological replicates. Formaldehyde-based chromatin fractionation (FAIRE) (Giresi et al. 2007) assays were performed using cross-linked chromatin prepared as for ChIP.

## ChIP-chip labeling and microarray hybridization

The labeling protocols were the same as previously reported (Koch et al. 2007). Labeled ChIP DNA samples and input samples were hybridized onto the Sanger ENCODE PCR product tiling array (Koch et al. 2007). Hybridized microarrays were scanned using a ScanArray 4000 confocal laser-based scanner (Perkin Elmer).

## Gene expression analysis on arrays

We prepared total RNA from three bioreplicates of K562 cells using TRIzol reagent as described above. Each purified RNA sample was assessed for quality and integrity using the 2100 Bioanalyzer (Agilent) following the manufacturer's instructions. Transcriptional activity in K562 was determined by hybridizing samples to the Sanger ENCODE array. Labeled samples were prepared by oligo-dT priming or random priming as described at <http://www.sanger.ac.uk/Projects/Microarrays/arraylab/protocol5.pdf>. RNA samples (labeled with Cy3) and K562 genomic DNA (labeled with Cy5) were hybridized, scanned, and quantitated as for ChIP-chip analysis. Overall expression levels were computed across all three bioreplicates as the mean of the polyA+ and oligo(dT) Z-scored data mapping to the 2000 bp at the 3' end of each gene. REST target genes were grouped according to variations of the RE1 motif (see Results; Fig. 4), and the mean expression levels were calculated for all genes within each group.

## Computational analyses of ChIP-chip data

### Pre-processing and normalization

Pre-processed data were created from raw enrichment data from three bioreplicate hybridizations as previously described (Koch et al. 2007). In addition, histone modification ChIP-chip datasets were normalized by dividing the final composite median data for each ChIP-chip assay with histone density (based on averaged H3 and H2B levels) and normal rabbit IgG antisera datasets. Similarly,

REST ChIP-chip experiments were normalized with the rabbit IgG antisera control datasets (in order to account for nonspecific binding not attributable to REST). Final normalized datasets were used to create three datasets used in the analyses: (1) median-centered enrichment data, (2)  $\log_2$ -centered enrichment data, and (3)  $\log_2$ -centered Z-scored enrichment data ( $\log_2$ -centered data minus the mean were divided by the standard deviation of the entire dataset). All data were stored and analyzed on NCBI genome build 35 (hg17).

### Determination of REST binding sites

Peaks of REST enrichment within the data were identified using a two-state hidden Markov model ("NESSIE"; Koch et al. 2007). An Excel workbook describing the exact genomic coordinates for each identified REST/genome interaction and a file compatible for UCSC genome browser upload showing the location of each hit as genome annotation are available (Supplemental Table S.4). Overlaps between the HMM sites identified in each cell line were determined using an in-house script (available on request) in order to derive nonredundant clusters of REST binding sites.

### REST target genes and Gene Ontology (GO)

Genes whose transcriptional start sites lay within 100 kb of HMM-identified peaks (NCBI build 35, [hg17]) were identified, and all GO terms for each of these genes were retrieved from the Gene Ontology database ([www.geneontology.org](http://www.geneontology.org)). A complete list of all the HMMs and their associated target genes and GO terms are available as supplemental data. The genes nearest to REST binding sites within these 100-kb intervals were used to determine statistically significant GO terms (<http://go.princeton.edu/cgi-bin/GOTermFinder>) for various subgroupings of REST sites as described in Results.

### RE1 motif and DNA sequence conservation analysis

Genomic DNA sequences corresponding to the HMM hits in each cell line were probed for recurring DNA sequence motifs using 95% confidence levels as determined by bootstrap analysis with the "NestedMICA" tool (Down and Hubbard 2005). In all cases, the DNA sequence coordinates of the microarray tiles that gave the highest enrichment for REST binding within each HMM were used to determine window sizes for computing the highest-scoring RE1 motifs. Hierarchical clustering of REST binding data in K562 cells prior to and after siRNA knockdown of REST was performed using Cluster 3.0 (Euclidean distance similarity metric with average linkage) and visualized with Java Treeview. Evolutionary constraints on DNA sequences contained within a  $\sim$ 1-kb window around highest-scoring RE1 sites were derived using the GERP conservation metric found at <http://www.ensembl.org/index.html>.

### Analysis of REST binding sites, histone enrichments, and genomic features

Normalized ChIP-chip datasets described above were viewed within the UCSC genome browser as formatted wiggle tracks (<http://genome.ucsc.edu/goldenPath/help/wiggle.html>) permitting the visualization of continuous-valued data in the context of annotated genome features. Histone modification behavior in the context of genes and REST binding sites was plotted using the R programming language (<http://www.r-project.org/>). Consensus histone modification profiles for subsets of REST target genes were visualized based on RE1 motif groupings as described in the Results. Gene coordinates for VEGA genes were downloaded from the Ensembl database (<http://www.ensembl.org/>). Genomic coordinates for REST binding sites were determined from HMM "NESSIE" analysis described above. Perl scripts were used compute

the average environment around genes or REST binding, using a sliding window.

## Acknowledgments

We thank Peter Ellis and Oliver Dovey from the Sanger Institute Microarray Facility and Gayle Cleland, Sarah Wilcox, and Joanne Fowler from the Sanger Institute ENCODE consortium team for generating the human ENCODE genomic DNA microarray. This work was funded by the Wellcome Trust. A.W.B. was a Wellcome Trust Sanger Institute post-doctoral fellow.

## References

- Atouf, F., Czernichow, P., and Scharfmann, R. 1997. Expression of neuronal traits in pancreatic beta cells. Implication of neuron-restrictive silencing factor/repressor element silencing transcription factor, a neuron-restrictive silencer. *J. Biol. Chem.* **272**: 1929–1934.
- Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merckenschlager, M., et al. 2006. Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8**: 532–538.
- Ballas, N., Grunseich, C., Lu, D.D., Speh, J.C., and Mandel, G. 2005. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* **121**: 645–657.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315–326.
- Bruce, A.W., Donaldson, I.J., Wood, I.C., Yerbury, S.A., Sadowski, M.I., Chapman, M., Gottgens, B., and Buckley, N.J. 2004. Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc. Natl. Acad. Sci.* **101**: 10458–10463.
- Bruce, A.W., Krejci, A., Ooi, L., Deuchars, J., Wood, I.C., Dolezal, V., and Buckley, N.J. 2006. The transcriptional repressor REST is a critical regulator of the neurosecretory phenotype. *J. Neurochem.* **98**: 1828–1840.
- Calderone, A., Jover, T., Noh, K.M., Tanaka, H., Yokota, H., Lin, Y., Grooms, S.Y., Regis, R., Bennett, M.V., and Zukin, R.S. 2003. Ischemic insults derepress the gene silencer REST in neurons destined to die. *J. Neurosci.* **23**: 2112–2121.
- Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**: 499–509.
- Chen, Z.F., Paquette, A.J., and Anderson, D.J. 1998. NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nat. Genet.* **20**: 136–142.
- Cheong, A., Bingham, A.J., Li, J., Kumar, B., Sukumar, P., Munsch, C., Buckley, N.J., Neylon, C.B., Porter, K.E., Beech, D.J., et al. 2005. Downregulated REST transcription factor is a switch enabling critical potassium channel expression and cell proliferation. *Mol. Cell* **20**: 45–52.
- Chong, J.A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altshuler, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. 1995. REST: A mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**: 949–957.
- Coulson, J.M. 2005. Transcriptional regulation: Cancer, neurons and the REST. *Curr. Biol.* **15**: R665–R668.
- Das, D., Banerjee, N., and Zhang, M.Q. 2004. Interacting models of cooperative gene regulation. *Proc. Natl. Acad. Sci.* **101**: 16234–16239.
- Dhami, P., Coffey, A.J., Abbs, S., Vermeesch, J.R., Dumanski, J.P., Woodward, K.J., Andrews, R.M., Langford, C., and Vetrie, D. 2005. Exon array CGH: Detection of copy-number changes at the resolution of individual exons in the human genome. *Am. J. Hum. Genet.* **76**: 750–762.
- Down, T.A. and Hubbard, T.J. 2005. NestedMICA: Sensitive inference of over-represented motifs in nucleic acid sequence. *Nucleic Acids Res.* **33**: 1445–1453.
- The ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**: 636–640.
- The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799–816.
- Euskirchen, G.M., Rozowsky, J.S., Wei, C.L., Lee, W.H., Zhang, Z.D., Hartman, S., Emanuelsson, O., Stolc, V., Weissman, S., Gerstein, M.B., et al. 2007. Mapping of transcription factor binding regions in mammalian cells by ChIP: Comparison of array- and sequencing-based technologies. *Genome Res.* **17**: 898–909.
- Fuller, G.N., Su, X., Price, R.E., Cohen, Z.R., Lang, F.F., Sawaya, R., and Majumder, S. 2005. Many human medulloblastoma tumors overexpress repressor element-1 silencing transcription (REST)/neuron-restrictive silencer factor, which can be functionally countered by REST-VP16. *Mol. Cancer Ther.* **4**: 343–349.
- Garriga-Canut, M., Schoenike, B., Qazi, R., Bergendahl, K., Daley, T.J., Pfender, R.M., Morrison, J.F., Ockuly, J., Stafstrom, C., Sutula, T., et al. 2006. 2-Deoxy-D-glucose reduces epilepsy progression by NRSF-CtBP-dependent metabolic regulation of chromatin structure. *Nat. Neurosci.* **9**: 1382–1387.
- Gaudet, J. and Mango, S.E. 2002. Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science* **295**: 821–825.
- Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., and Lieb, J.D. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**: 877–885.
- Guccione, E., Martinato, F., Finocchiaro, G., Luzi, L., Tizzoni, L., Dall’Olio, V., Zardo, G., Nervi, C., Bernard, L., and Amati, B. 2006. Myc-binding-site recognition in the human genome is determined by chromatin context. *Nat. Cell Biol.* **8**: 764–770.
- Hentsch, B., Mouzaki, A., Pfeuffer, I., Rungger, D., and Serfling, E. 1992. The weak, fine-tuned binding of ubiquitous transcription factors to the Il-2 enhancer contributes to its T cell-restricted activity. *Nucleic Acids Res.* **20**: 2657–2665.
- Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Snyder, M. 2002. GATA-1 binding sites mapped in the beta-globin locus by using mammalian ChIP-chip analysis. *Proc. Natl. Acad. Sci.* **99**: 2924–2929.
- Impey, S., McCorkle, S.R., Cha-Molstad, H., Dwyer, J.M., Yochum, G.S., Boss, J.M., McWeeney, S., Dunn, J.J., Mandel, G., and Goodman, R.H. 2004. Defining the CREB regulon: A genome-wide analysis of transcription factor regulatory regions. *Cell* **119**: 1041–1054.
- Johnson, R., Gamblin, R.J., Ooi, L., Bruce, A.W., Donaldson, I.J., Westhead, D.R., Wood, I.C., Jackson, R.M., and Buckley, N.J. 2006. Identification of the REST regulon reveals extensive transcription element-mediated binding site duplication. *Nucleic Acids Res.* **34**: 3862–3877.
- Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. 2007. Genome-wide mapping of in vivo protein–DNA interactions. *Science* **316**: 1497–1502.
- Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P., et al. 2007. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res.* **17**: 691–707.
- Kuwahara, K., Saito, Y., Takano, M., Arai, Y., Yasuno, S., Nakagawa, Y., Takahashi, N., Adachi, Y., Takemura, G., Horie, M., et al. 2003. NRSF regulates the fetal cardiac gene program and maintains normal cardiac structure and function. *EMBO J.* **22**: 6310–6321.
- Lawinger, P., Venugopal, R., Guo, Z.S., Immaneni, A., Sengupta, D., Lu, W., Rastelli, L., Marin Dias Carneiro, A., Levin, V., Fuller, G.N., et al. 2000. The neuronal repressor REST/NRSF is an essential regulator in medulloblastoma cells. *Nat. Med.* **6**: 826–831.
- Livak, K.J. and Schmittgen, T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* **25**: 402–408.
- Lunyak, V.V., Burgess, R., Prefontaine, G.G., Nelson, C., Sze, S.H., Chenoweth, J., Schwartz, P., Pevzner, P.A., Glass, C., Mandel, G., et al. 2002. Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* **298**: 1747–1752.
- Martin, D., Tawadros, T., Meylan, L., Abderrahmani, A., Condorelli, D.F., Waeber, G., and Haefliger, J.A. 2003. Critical role of the transcriptional repressor neuron-restrictive silencer factor in the specific control of connexin36 in insulin-producing cell lines. *J. Biol. Chem.* **278**: 53082–53089.
- Martin, D., Allagnat, F., Chaffard, G., Caille, D., Fukuda, M., Regazzi, R., Abderrahmani, A., Waeber, G., Meda, P., Maechler, P., et al. 2008. Functional significance of repressor element 1 silencing transcription factor (REST) target genes in pancreatic beta cells. *Diabetologia* **51**: 1429–1439.
- Mori, N., Schoenherr, C., Vandenbergh, D.J., and Anderson, D.J. 1992. A common silencer element in the SCG10 and type II Na<sup>+</sup> channel genes binds a factor present in nonneuronal cells but not in neuronal cells. *Neuron* **9**: 45–54.
- Mortazavi, A., Leeper Thompson, E.C., Garcia, S.T., Myers, R.M., and Wold, B. 2006. Comparative genomics modeling of the NRSF/REST repressor network: From single conserved sites to genome-wide repertoire. *Genome Res.* **16**: 1208–1221.
- Nunez, E., Kwon, Y.S., Hutt, K.R., Hu, Q., Cardamone, M.D., Ohgi, K.A., Garcia-Bassets, I., Rose, D.W., Glass, C.K., Rosenfeld, M.G., et al. 2008.

- Nuclear receptor-enhanced transcription requires motor- and LSD1-dependent gene networking in interchromatin granules. *Cell* **132**: 996–1010.
- Orphanides, G. and Reinberg, D. 2002. A unified theory of gene expression. *Cell* **108**: 439–451.
- Otto, S.J., McCorkle, S.R., Hover, J., Conaco, C., Han, J.J., Impey, S., Yochum, G.S., Dunn, J.J., Goodman, R.H., and Mandel, G. 2007. A new binding motif for the transcriptional repressor REST uncovers large gene networks devoted to neuronal functions. *J. Neurosci.* **27**: 6729–6739.
- Palm, K., Belluardo, N., Metsis, M., and Timmusk, T.O. 1998. Neuronal expression of zinc finger transcription factor REST/NRSF/XBR gene. *J. Neurosci.* **18**: 1280–1296.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* **290**: 2306–2309.
- Ren, B., Cam, H., Takahashi, Y., Volkert, T., Terragni, J., Young, R.A., and Dynlacht, B.D. 2002. E2F integrates cell cycle progression with DNA repair, replication, and G<sub>2</sub>/M checkpoints. *Genes & Dev.* **16**: 245–256.
- Roh, T.Y., Cuddapah, S., Cui, K., and Zhao, K. 2006. The genomic landscape of histone modifications in human T cells. *Proc. Natl. Acad. Sci.* **103**: 15782–15787.
- Roopra, A., Qazi, R., Schoenike, B., Daley, T.J., and Morrison, J.F. 2004. Localized domains of G9a-mediated histone methylation are required for silencing of neuronal genes. *Mol. Cell* **14**: 727–738.
- Schoenherr, C.J. and Anderson, D.J. 1995. The neuron-restrictive silencer factor (NRSF): A coordinate repressor of multiple neuron-specific genes. *Science* **267**: 1360–1363.
- Schoenherr, C.J., Paquette, A.J., and Anderson, D.J. 1996. Identification of potential target genes for the neuron-restrictive silencer factor. *Proc. Natl. Acad. Sci.* **93**: 9881–9886.
- Scholl, T., Stevens, M.B., Mahanta, S., and Strominger, J.L. 1996. A zinc finger protein that represses transcription of the human MHC class II gene, DPA. *J. Immunol.* **156**: 1448–1457.
- Stormo, G.D. 2000. DNA binding sites: Representation and discovery. *Bioinformatics* **16**: 16–23.
- Wei, C.L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., et al. 2006. A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**: 207–219.
- Westbrook, T.F., Martin, E.S., Schlabach, M.R., Leng, Y., Liang, A.C., Feng, B., Zhao, J.J., Roberts, T.M., Mandel, G., Hannon, G.J., et al. 2005. A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121**: 837–848.
- Zhang, C., Xuan, Z., Otto, S., Hover, J.R., McCorkle, S.R., Mandel, G., and Zhang, M.Q. 2006. A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome. *Nucleic Acids Res.* **34**: 2238–2246.
- Zuccato, C., Tartari, M., Crotti, A., Goffredo, D., Valenza, M., Conti, L., Cataudella, T., Leavitt, B.R., Hayden, M.R., Timmusk, T., et al. 2003. Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nat. Genet.* **35**: 76–83.

Received November 10, 2008; accepted in revised form February 23, 2009.