Cell, Vol. 119, 1041–1054, December 29, 2004, Copyright ©2004 by Cell Press

Defining the CREB Regulon: A Genome-Wide Analysis of Transcription Factor Regulatory Regions

Resource

Soren Impey,[1,*] Sean R. McCorkle,[2] Hyunjoo Cha-Molstad,[1] Jami M. Dwyer,[1] Gregory S. Yochum,[1] Jeremy M. Boss,[4] Shannon McWeeney,[3] John J. Dunn,[2] Gail Mandel,[5] and Richard H. Goodman[1]

[1]Vollum Institute
Oregon Health and Science University
Portland, Oregon 97239

[2]Biology Department
Brookhaven National Laboratory
Upton, New York 11973

[3]Department of Public Health and Preventative Medicine
Oregon Health and Science University
Portland, Oregon 97239

[4]Department of Microbiology and Immunology
Emory School of Medicine
Atlanta, Georgia 30322

5Howard Hughes Medical Institute
Department of Neurobiology and Behavior
The State University of New York at Stony Brook
Stony Brook, New York 11794

Summary

The CREB transcription factor regulates differentiation, survival, and synaptic plasticity. The complement of CREB targets responsible for these responses has not been identified, however. We developed a novel approach to identify CREB targets, termed serial analysis of chromatin occupancy (SACO), by combining chromatin immunoprecipitation (ChIP) with a modification of SAGE. Using a SACO library derived from rat PC12 cells, we identified ~41,000 genomic signature tags (GSTs) that mapped to unique genomic loci. CREB binding was confirmed for all loci supported by multiple GSTs. Of the 6302 loci identified by multiple GSTs, 40% were within 2 kb of the transcriptional start of an annotated gene, 49% were within 1 kb of a CpG island, and 72% were within 1 kb of a putative cAMP-response element (CRE). A large fraction of the SACO loci delineated bidirectional promoters and novel antisense transcripts. This study represents the most comprehensive definition of transcription factor binding sites in a metazoan species.

Introduction

The transcription factor CREB was originally identified by virtue of its role in cAMP signaling (for review, see Mayr and Montminy [2001]). CREB has since been found to mediate calcium, neurotrophin, and cytokine signals as well as those of a variety of cellular stresses. CREB binds to the cAMP-response element (CRE), a sequence identified in the promoters of many inducible genes (Montminy et al., 1986). The presence of CRE sequences in multiple target genes is believed to contribute to coordinate regulation. Although a multitude of kinases phosphorylate CREB in vitro, thus far only PKA, CaMKIV, MSK, and RSK have been shown to activate CREB-dependent transcription (Lonze and Ginty, 2002). Phosphorylation of CREB at Ser133 triggers the recruitment of the coactivator CBP, which induces transcription via its intrinsic and associated acetylase activities and/or by interacting with the core transcriptional machinery (Vo and Goodman, 2001).

The ability of CREB to be activated by multiple signaling pathways has led investigators to examine its role in long-term adaptative responses to various extracellular stimuli. CREB regulates the differentiation of T cells, hepatocytes, and spermatocytes, and CREB-dependent gene expression plays a particularly important role in the central nervous system, where it regulates neuronal survival, memory consolidation, addiction, entrainment of the biological clock, and synaptic refinement. Surprisingly, despite strong evidence for its involvement in gene expression, CREB binding to native promoters has been documented for only a few target genes. Moreover, many CREB targets (e.g., C/EBPβ, Egr1, and Nurr1) are themselves transcription factors that regulate other genes. Thus, it is often uncertain which genes are activated by CREB directly and which are activated indirectly.

One approach for identifying CREB targets is to measure the changes in gene expression that occur in response to an activating or interfering CREB mutant (Fass et al., 2003; McClung and Nestler, 2003). A caveat to this approach is that the overexpressed protein may bind inappropriately or compete for coactivators. Additionally, effects can be indirect. For example, one of the genes most highly induced by VP16-CREB (a constitutively active CREB mutant) is ICER, a potent inhibitor of CREB function, which itself alters the expression of CREB-regulated genes (Fass et al., 2003).

Chromatin immunoprecipitation (ChIP) measures binding of endogenous transcription factors to native promoters. Theoretically, immunoprecipitated DNA fragments can be cloned and sequenced to determine their identities (Hug et al., 2004). This method is not suitable for identifying transcription factor binding sites across the genome, however. For this goal, a more comprehensive approach is needed, such as the "ChIP-on-chip" method. In this procedure, chromatin immunoprecipitated DNA is amplified using PCR and hybridized to promoter sequences displayed on a microarray (the second chip). The major limitation of this method lies in defining what constitutes a promoter. Initial microarrays utilized CpG islands, which tend to be enriched in promoter regions (Weinmann et al., 2001). Second generation ChIP-on-chip assays used specific oligonucleotides or DNA fragments derived from known promoter sequences. This approach has been highly effective in analysis of yeast transcriptional programs, where the transcriptional units are well characterized (Lee et al., 2002).

More recent studies have extended the ChIP-on-chip approach to mammalian systems. Odom et al., for exam-

*Correspondence: impeys@ohsu.edu

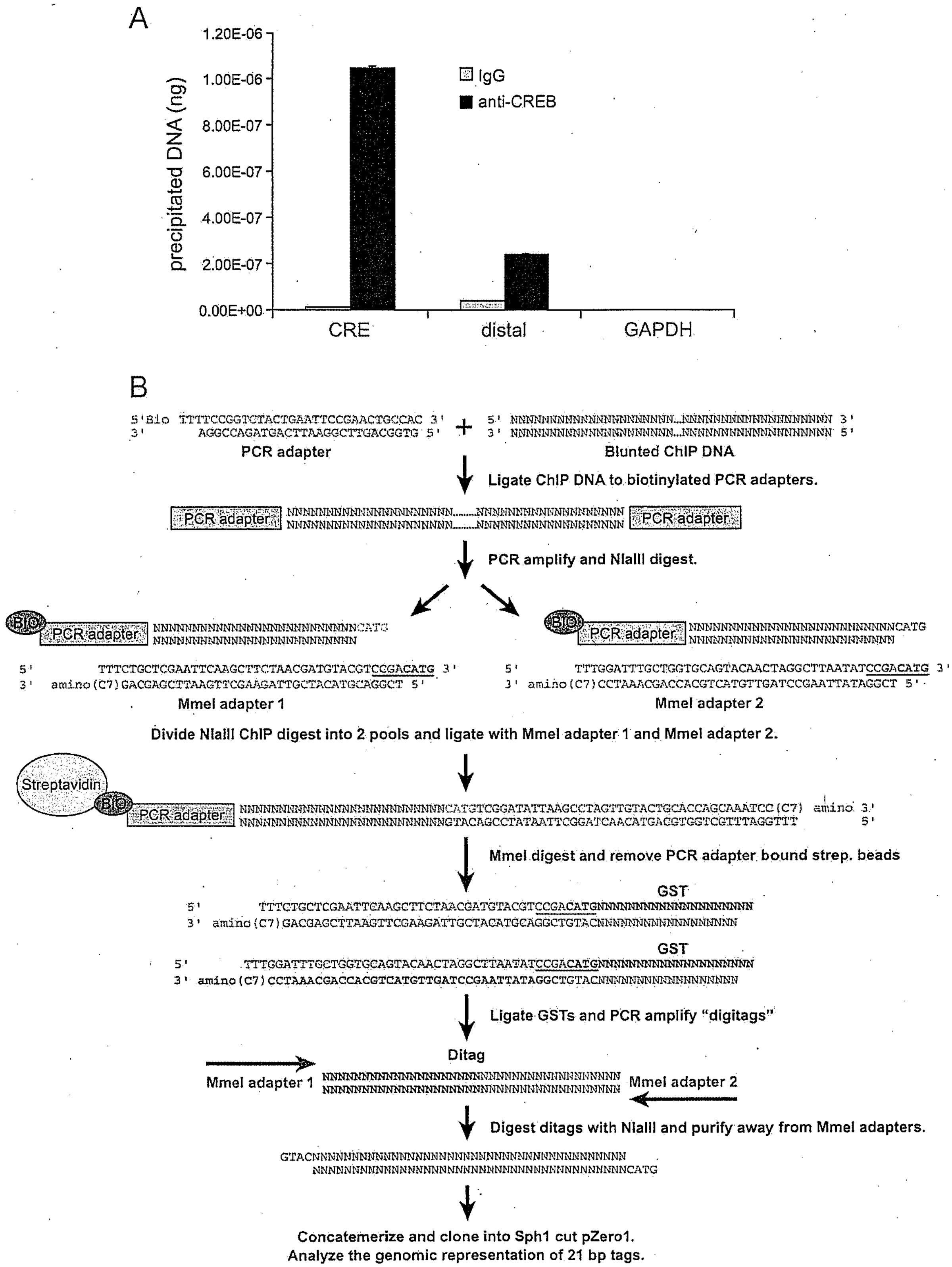

Figure 1. Construction of the SACO Library

(A) CREB ChIP. Forskolin-treated PC12 cells were subjected to a ChIP assay using an anti-CREB antibody or nonspecific IgG as described in the methods. The levels of immunoprecipitated DNA encompassing the *c-fos* CRE, a distal *c-fos* region (−439 to −678), and a GAPDH control were assessed by real-time quantitative PCR using an amplicon standard curve. The data from triplicate determinations were normalized

ple, recently identified binding sites for Hnf1, Hnf4, and Hnf6 in genomic DNA from pancreatic and liver cells (Odom et al., 2004). The microarray in this study contained a collection of sequences believed to lie within 750 base pairs of the initiation sites of annotated genes. While most of these sequences are likely to have been characterized correctly, annotation of mammalian genomic databases remains problematic, and many mammalian promoters have not been well characterized. Another problem is that important transcription factor binding sites may lie outside this 750 base pair region.

A recent report by Cawley et al. suggested that these concerns are more than theoretical (Cawley et al., 2004). These workers examined the binding of Sp1, c-*myc*, and p53 to a microarray representing the nonrepetitive sequences of chromosomes 21 and 22. (The small sizes of these two chromosomes, representing only 2% of the genome, made this study feasible; such an analysis of the whole genome would not be possible.) Remarkably, when this unbiased approach was used, only 22% of the transcription factor binding sites were located near known promoters. Even if part of the discrepancy between Cawley et al. (2004) and Odom et al. (2004) is due to different standards of defining promoters, it still appears that promoter microarrays may miss important aspects of biological regulation. Other problems of ChIP-on-chip assays, the low sensitivity inherent in hybridization assays and the limited potential for quantification, detract from both methods.

CREB binding sites have been identified on chromosome 22 using a ChIP-on-chip assay, but, because only ~1% of the genome was interrogated, relatively few responsive genes were detected (Euskirchen et al., 2004). For analysis of a factor like CREB, promoter microarrays are also of limited value because internal or intronic promoters specify a significant percentage of known CREB-regulated genes (e.g., ICER, BDNF, nNos, TrkB, PDE4B, PDE4D). Thus, an unbiased methodology is critical for characterizing CREB binding sites in vivo. We developed such an approach, which we term serial analysis of chromatin occupancy (SACO). This method can quantitatively interrogate an entire metazoan genome by combining ChIP with a modification of long serial analysis of gene expression (Long-SAGE), a method normally used for mRNA analysis. A similar method has been used to map acetylation of histone H3 and H4 in yeast (Roh et al., 2004). SACO relies on the observation that 21 base pair DNA fragments can be used to identify specific chromosomal locations from a genomic database. By sequencing large numbers of concatemerized 21 base pair genomic signature tags (GSTs) generated from CREB ChIPs, we can identify and quantify the frequency of CREB binding sites. In this study, we describe the first results of an unbiased analysis of transcription factor binding across an entire mammalian genome. Most of the confirmed CREB binding loci isolated were located in or near transcriptionally active regions, and the majority of these were upstream from genes not previously known to be regulated by CREB. A significant percentage of CREB GSTs were found in introns or at the 3′ ends of genes, however. Most of the GSTs were tightly associated with CRE elements and CpG islands. This study defines the CREB regulon in forskolin-treated PC12 cells and provides the most comprehensive representation of transcription factor binding sites across the entire genome in a metazoan species.

Results

Optimizing the CREB ChIP

To generate a specific SACO library, we first optimized our ChIP methodology. We utilized PC12 cells because they have been used extensively for studies of CREB function. Cells were stimulated with forskolin because of the possibility that cAMP may stimulate CREB binding to certain promoters (Hiroi et al., 2004; Nichols et al., 1992; Wolfl et al., 1999). In general, however, CREB association with CRE sequences is not believed to require cAMP signaling. Binding was analyzed using a ChIP protocol that includes an extended period of formaldehyde fixation followed by a pH 9.4 wash. The CREB antibody provided an ~100-fold enrichment for c-*fos* (and other CREB targets) as compared to IgG (Figure 1A). We also analyzed the GAPDH promoter and an upstream region of c-*fos* (−439 to −678), neither of which are believed to contain CREB binding sites. No enrichment of the GAPDH promoter fragment was detected. Binding to the c-*fos* distal fragment was about 20% that detected for the CRE. Because the chromatin fragments in this assay averaged around 700 bp in length, we expected that some fragments would contain both the CRE and the upstream site, which could account for some of the signal obtained with the upstream primers. Alternatively, there could be a cryptic CREB binding site in or near the distal fragment. Because the standard ChIP assay does not assess the level of nonspecific chromatin contamination, we additionally measured the level of contaminating LINE L1 DNA. (The degenerate L1 retrotransposons comprise ~25% of the rat genome). When compared to the original genomic input, our CREB ChIP showed an ~9×10^5-fold depletion of L1 DNA. There was no significant enrichment for L1 DNA in the IgG control.

Generation and Verification of CREB SACO Library

Sonicated CREB ChIP DNA was polished (protruding 3′ and 5′ ends were made flush) and ligated to adapters for limited amplification. The resulting DNA was digested

for amplicon size. Error bars denote SEM.

(B) SACO methodology. ChIP material (blue) was polished and ligated to PCR adapters (red). Following limited amplification, the ChIP material was digested with NlaIII and divided into two pools that were ligated to distinct MmeI adapters (green and red). The two ligations were bound to streptavidin beads and digested with MmeI. This allowed purification of GST fragments without contaminating adaptor or nondigested ChIP DNA. The GST fragments were ligated to form GST "ditags," which were purified and concatemerized. Concatemers (500–1100 bp) were ligated into pZero1 and sequenced.

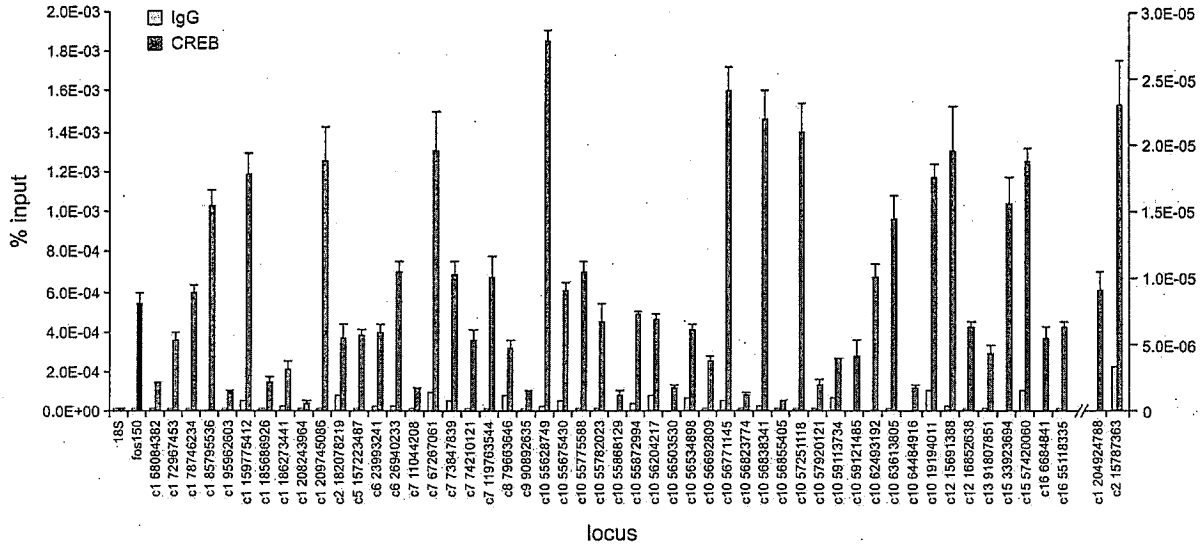

Figure 2. Confirmation of CREB Binding Sites

Forskolin-treated PC12 cells were subjected to repeat ChIP assays using primers designed from regions surrounding duplicated GSTs. Real-time quantitative PCR was used to assess DNA levels. Data are expressed as the percentage input DNA normalized to LINE L1 levels. All samples showed at least a 2-fold enrichment using CREB antiserum as compared to IgG. The black bar shows the enrichment for c-*fos*. Error bars denote SEM of quadruplicate determinations.

with NlaIII, which cleaves genomic DNA approximately every 120 base pairs, and a modified SAGE procedure was used to create concatemerized 21 bp genomic signature tags (GSTs). An outline of the method is presented in Figure 1B. The SACO library contained in excess of 3×10^6 GSTs. To date, we have sequenced approximately 76,000 GSTs and have obtained approximately 41,000 that identify a single locus in the rat genome. Because sonication generates a population of genomic fragments in which NlaIII sites are localized randomly, a particular CREB binding site may be represented by clusters of distinct GSTs. Therefore, we defined a SACO locus as any collection of GSTs that are within 2 kb of each other. A computational analysis of the rat genome predicted that 75% of all theoretical 21 bp GSTs should be assignable to unique loci. In practice, unambiguous correlation with a unique site is somewhat less than 75% because GSTs containing repetitive sequences cannot be localized. Overall, we have found that we can assign approximately 70% of the GSTs to a single genomic locus.

One can increase specificity in SAGE analyses by considering only duplicate tags. Therefore, we focused our SACO analysis on the 6302 loci represented by more than one GST (see Supplemental Table S1 at http://www.cell.com/cgi/content/full/119/7/1041/DC1/). To determine the validity of these GSTs, we utilized ChIP to test all multiply identified tags within a gene-dense span of chromosome 10. Quantitative PCR confirmed CREB binding to 100% (32 of 32) of chromosome 10 SACO loci identified by duplicate tags (Figure 2). Because limiting our analysis to the chromosome 10 region could introduce bias, we selected another set of multiply identified loci from other chromosomes. One hundred percent (21 of 21) of these were also confirmed by repeat ChIP assay (Figure 2). A separate ChIP assay performed using a different CREB antibody also confirmed the interaction of CREB to all (24 of 24) SACO loci tested (data not shown).

Because many loci were identified by only a single GST, we also analyzed CREB binding to this subgroup. We found that CREB binding to 69% of these loci (43 of 62) could be confirmed in a second ChIP assay (data not shown). The frequency of any GST in our library should be proportional to its abundance in the original ChIP, which in part measures the strength of the CREB interaction. Nevertheless, GSTs are anchored to their genomic NlaIII sites, raising the possibility that some loci might not be confirmed because of the distance between the NlaIII site and the actual CREB binding element. Therefore, we performed additional PCR analyses to look for distal CRE sequences in several of the SACO loci that did not confirm. Fifty-six percent (five of nine) of these GSTs could ultimately be confirmed by using different sets of PCR primers. Based on this result and the likelihood that some nonconfirmed GSTs represent PCR artifacts, we believe that almost 90% of the GSTs overall define in vivo CREB binding sites.

Loci Identified in the SACO Screen Are Enriched for CREB Binding Motifs and CpG Islands

CREB was originally identified through its ability to bind to a palindromic element, TGACGTCA (Montminy et al., 1986). Nevertheless, many CREB binding elements consist of half-palindromic sites (TGACG) or contain multiple substitutions (Mayr and Montminy, 2001). Indeed, 72% of the loci represented by more than one GST were within 1 kb of a half-palindromic CRE. This association was highly significant (see Supplemental Data on the *Cell* web site). The palindromic CRE sequence was also significantly enriched in regions associated with SACO

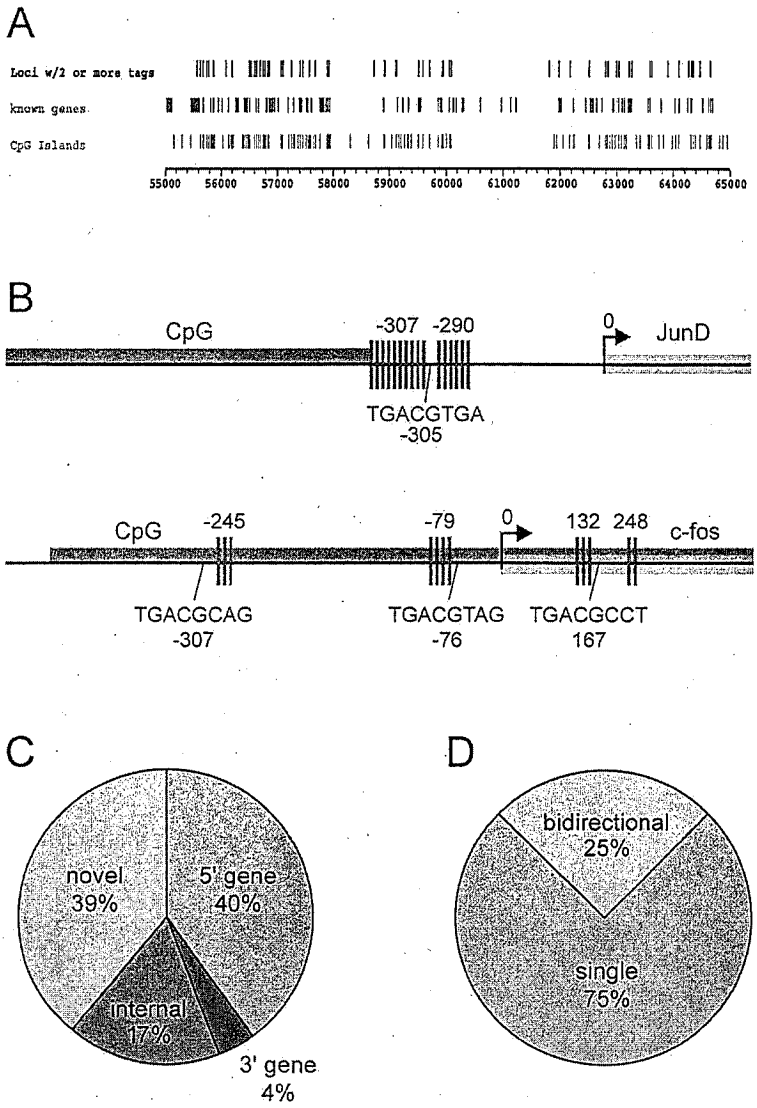

Figure 3. Analysis of Duplicated SACO Loci

(A) Chromosomal map of a gene-rich region of chromosome 10 (bp 55,000,000–65,000,000). The "Loci" track shows the first base of duplicated loci on the assembled chromosome. The "known gene" track depicts the 5′ ends of known genes from the UCSC genome center and ENSEMBL. The "CpG islands" track indicates the first base of UCSC genome center CpG island annotation. Known genes and CpG islands were highly correlated with duplicated SACO loci across the entire genome.

(B) Clustering of GSTs in two representative promoter regions. Transcriptional start sites are based on ENSEMBL annotation and are indicated by arrows. Vertical bars represent the number of individual GSTs at a given site.

(C) Classification of SACO loci within 2 kb of the start sites of annotated genes (5′), based on known (UCSC genome center) or ENSEMBL full-length mRNA gene categories. Another 23% of duplicated loci localize 5′ to ECgene (EST- and mRNA-based) transcript predictions. SACO loci within 2 kb of the end of an annotated gene (3′) and within a gene (internal) are also depicted. The subset of 3′ loci that were present in the 5′ category and the subset of internal loci that were present in either the 5′ or 3′ category were not considered. "Novel" denotes loci that are not near or within annotated genes.

(D) "Bidirectional" denotes loci flanked by two transcripts (annotated genes or ECgene transcripts) in a head-to-head arrangement. Only bidirectional transcripts separated by less than 2 kb were considered. "Single" denotes loci within 2 kb of a 5′ start site.

loci as compared to random genomic DNA (see Supplemental Data).

The distribution of CpG islands and transcribed genes is highly correlated in the rat genome; likewise, we found that the distribution of CpG islands and GSTs is significantly correlated (Figure 3A and Supplemental Data). Approximately half of the duplicated GSTs were within 1 kb of a CpG island (UCSC genome annotation). Because CpG islands occur on average only once every ~175,000 bp in the rat genome, this level of enrichment is highly significant (Supplemental Data). Nonetheless, many CREB binding sites were not associated with CpG islands. Consequently, a ChIP-on-chip assay based on a CpG-enriched microarray would miss most of these CREB binding sites.

GST Clusters Define CREB Binding Sites

Because loci identified in our SACO library typically contain multiple NlaIII sites, the clustering of GSTs can delineate distinct CREB binding regions within a promoter. Interestingly, some well-characterized CREB targets, such as *c-fos*, show multiple clusters of tags (Figure 3B). This result suggests that the ChIP signal detected by the distal primers in Figure 1A results from CREB binding to sequences in or near that region. Whether the reiterated CREs in a promoter interact with multiple CREB molecules simultaneously is unclear. It is possible that only one CRE in a single promoter is functional at any given time.

Analysis of Transcripts Associated with SACO Loci

Although dozens of potential CREB target genes have been described (see Mayr and Montminy [2001] for review), much of the evidence for these targets derives from artificial reporter assays and in vitro binding assays. In this study, we identified 32,723 potential CREB-regulatory regions. Therefore, we sought to determine the relationship of these potential regulatory regions with genes and transcripts by computational

analysis of the rat genomic database. We focused on the 6302 loci supported by multiple tags because, when we tested a sample of these loci, CREB binding could be validated in all cases by quantitative PCR.

The density of duplicated SACO loci significantly paralleled the density of well-annotated genes genome-wide, as shown along a 10 million base pair stretch of chromosome 10 (Figure 3A and Supplemental Data). Forty percent of these duplicated loci fell within two kilobases of the 5′ region of an annotated gene (defined as the most 5′ region of UCSC "known gene" or ENSEMBL gene annotation, the association was highly significant; Figure 3C and Supplemental Data). Another 23% were localized within 2 kb of the start sites of transcripts predicted by ECgene EST and mRNA clustering annotation (UCSC genome browser). Interestingly, many ECgene predictions had limited protein coding potential and, barring sequencing or assembly errors, may represent noncoding RNAs. We analyzed a subset of these ECgene predictions by quantitative PCR and found that nine of 12 were expressed in PC12 cells (data not shown). This suggests that many of the duplicated SACO loci that are not upstream from the promoters of annotated genes may, nonetheless, drive expression of novel RNA transcripts.

Approximately 17% of SACO loci (1070 of 6302) were located downstream of a start site, suggesting that they may denote alternative promoters (Figure 3C). In support of this possibility, several known intronic promoters that drive CREB target genes (TrkB, CREM) were identified by GST evidence. Interestingly, 4% of the duplicated SACO loci were located near the 3′ ends of annotated genes, and ECgene or EST evidence of transcription was often associated with these loci. Although this percentage was lower than that reported by Cawley et al. (2004), the enrichment for 3′ ends was nonetheless highly significant (Supplemental Data). This suggests that at least some of these 3′ loci represent functional promoters. Consistent with the observations of Cawley et al. (2004), we also found many examples of SACO loci at both the 5′ and 3′ ends of some genes.

Functional Annotation of SACO Loci

Having established that most duplicated SACO loci were tightly associated with genes and gene regulatory regions, we categorized a selection of these potential CREB target genes by function (Tables 1 and 2). The vast majority of genes associated with SACO loci had not previously been proposed to be CREB targets (Table 2). These included transcriptional regulators, including chromatin modifying enzymes, coactivators, and corepressors. Consistent with the neuroendocrine origin of PC12 cells, many genes associated with neurotransmitter synthesis, transport, and release were also identified. In addition, multiple genes involved in vesicle docking and release were observed. Although many kinds of signaling intermediates were identified, it is striking that several are involved with cAMP and Ca^{2+} signaling in particular. These pathways are important for the vesicle loading and secretion of the catecholamine and peptide neurotransmitters produced in PC12 cells.

Another interesting category included genes involved in mitochondrial homeostasis and protein import. Mitochondrial function provides energy for the release and regeneration of neurotransmitters and plays a critical role in regulating Ca^{2+} signals that trigger neurotransmitter release. It is also not surprising that many of the targets regulate cell survival, because CREB has been implicated in somatic and mitochondrial survival pathways. Other major categories of CREB target genes were associated with proliferation and cell cycle entry, metabolism, proteases, transporters, and chaperones.

Is CREB the Major Transcriptional Mediator of cAMP Signaling?

Although many cAMP-responsive genes contain CREs (Mayr and Montminy, 2001), the general importance of CREB as a mediator in this pathway is not known. We utilized Affymetrix oligonucleotide arrays containing probes to 28,800 potential genes to define what proportion of cAMP-regulated genes were CREB targets. The levels of 1621 transcripts were significantly increased by forskolin relative to control cells treated with an inactive analog, and 1333 were mapped to genomic regions (Supplemental Table S2). Some of these genes were known to be regulated by cAMP, but most had not previously been linked to this pathway (Table 3). Approximately half of the cAMP-regulated transcripts were also identified in the SACO library, suggesting that these represent direct CREB targets.

We next asked whether the CREB-regulatory regions defined by SACO loci are functionally important for cAMP induction. We used quantitative PCR to assess mRNA levels of a selection of transcripts identified in the Affymetrix screen. As expected, all (15 of 15) were significantly stimulated by forskolin (Figure 4 and data not shown). Expression of ACREB, a selective CREB inhibitor, attenuated transcriptional activation of 93% of this gene subset (14 of 15).

CREB Loci Define Genes Induced by Diverse Extracellular Stimuli

Only a fraction of the genes identified in the SACO analysis were found to be upregulated by cAMP signaling in the Affymetrix study. Because we assessed mRNA levels at only one time point, a more detailed kinetic analysis could reveal additional cAMP-regulated genes. Alternatively, the Affymetrix screen may not be sufficiently sensitive to detect changes in nuclear transcription at the level of cytosolic mRNA. Although the Affymetrix screen likely underrepresents the number of cAMP-stimulated genes, analysis of mRNA levels suggests another possibility. Many cAMP-regulated genes responded more effectively to NGF or UV irradiation than they did to forskolin (Figure 4 and data not shown). Moreover, expression of ACREB significantly attenuated stimulation by NGF and UV for 12 of 15 genes studied. Because CREB can be activated by multiple hormone-, neurotransmitter-, and stress-regulated kinases, these results suggest that many genes linked to a SACO locus will be activated by signaling pathways other than cAMP. Indeed, a large fraction of the targets detected in the SACO library were previously identified as primary response genes to cytokines, neurotrophins, TNFα, UV, serum, TGFβ, radiation, cold shock, retinoids, etc. Moreover, several CREB target genes responded differen-

Table 1. Categorization of Novel SACO-Associated Transcripts

Transcription/Nuclear	Signaling	Synaptic/Endocrine	Metabolic/Structural	Cell Cycle/Proliferation	Survival/Stress
Fra2	NT4	Synaptotagmin V	Hydroxyacyl glutathione hydrolase	Cyclin B1	Siva
FosB	IGF2 receptor	Synaptotagmin 7	Serine palmitoyltransferase 1	DNA topoisomerase II	Defender against cell death
Btg2	GDNF receptor 2	Synaptotagmin 13	Pancreasin	Cyclin E1	Bcl2l12
TOB1	TrkA	VAMP-1 synatobrevin 1	Cytochrome c oxidase, subunit 4a	Cyclin G AK	Survival motor neuron
Hnf4a	p75 NGF R	VAMP-associated A	Cytochrome P450 20a1	CHK1	Survival motor neuron 1
Sp4	IL6 receptor	GABARAP	Glutathione S-transferase, mito.	Elac2 prostate cancer susceptibility	Programmed cell death 8
SREBP1	Bradykinin receptor 1	SNAP-α	Neutral sphingomyelinase	Tumor differentially expressed 1	Death-associated-like kinase
ATF4	RAGE	NSF	Mitochondrial VDAC2	Tumor-rejection antigen SART3	Apoptosis inhibitor 2
ATF5	IL11 Receptor	VAMP-2	Glyoxalase ii	Translationally controlled tumor protein	BAX inhibitor-1
Sp1	P2X4	Semaphorin 6B	Peroxiredoxin 5	myeloid leukemia factor 2	DEDD
Stat5a	PAK-2	Reticulon 4/NOGO	Thyroxine deiodinase	Putative tumor supressor ST13	BCL-W
Klf5	Grk2	CRMP1	APEX nuclease	Wd-repeat tumor rejection antigen	Caspase-activated DNAase
Onecut	CKII beta	CRMP3	o-sialoglycoprotease	Testosterone regulated tumor suppressor	TRADD
MHC II	ERK5	Nudel	Sialidase 1	TRADD	Neuronal cell death inducible
transactivator	Mapkapk5	Shank2	Thioredoxin peroxidase 1	Elavl2/HUB	Bim
Id1	GSK3α	GKAP	NA(+)/H(+) exchanger	Wisp2	RAD50
Id3	S6 kinase 1	MAGUK p55	Transaminase A	Tumor susceptability gene 101	DDIT3
Hes3	AKAP8	SV2	TIM44	Werner helicase interacting protein	CiRBP
HesB	cGMP PDE	Shank-interacting	HSP10	v-erba-related	GADD45γ
Prmt1	PDE4c	Ephrin A5	Mitochondrial aconitase	c-crk	HSP70
Prmt3	cAMP-GEF 2	DAP-4	TIM13	c-mer	HSP90
DNMT1	PP5C	Netrin3	TIM22	Rb-like 2	Catalase
TIP160	PP2B Sub 1	Syntaxin 5	TIM9	NF-I	TRAF4
NRBF-1	PP1αC	Cdk5	α actinin	PIN3	
p105 coactivator	14-3-3 tau	NR1ζ	MAP1A/MAP1B LC3 MAP 1B	DHFR	
PPARγ coactivator 1	14-3-3 sigma	Chloride channel 3	Dynein		
RUVB-like 1	14-3-3 gamma	SK1	Dynein light chain		
NCOA-2	IRS1	TRP 5 4	MAP1A LC2		
KAP-1	R-Ras GAP	TRPC1			
NAC1	IKK-β	Vanilloid receptor-like			
TIEG1	SOCS	ASIC1			
Nuclear cap binding protein 2	SOCS1	atrophin-related			
Nuclear RNA export factor	PAR3	GABA-B rec. 2			
Nucleoporin p58	N-Ras	Neuronal acetycholine rec.			
NUP98	R-Ras	Neuronal pentraxin rec.			
NUP88	Rab3A	P2X4			
NUP155	Rab28				
NOPP140	Ran				
poly A binding	ARF4				
HMG-2	ARF5				
	p85 PI3K				

Table 2. Categorization of Known SACO-Associated Transcripts

Transcription/Nuclear	Signaling	Neuronal	Metabolic	Cell Cycle/Proliferation
c-*fos*	SSTR2	Synaptotagmin IV	tPA	Cyclin D1
Nurr1	Cardiotrophin-1	Kv3.1	Carnitine palmitoyltransferase	DNA polymerase β
NOR1	αA crystallin	Presenilin-2	Aquaporin 2	PCNA
c-jun	Heme oxygenasel		Ornithine decarboxylase	DNA polymerase γ
c-maf	Galanin receptor		Aryl hydrocarbon receptor	NF-I
JunD	TGFbeta2		HMG CoA reductase	BRCA1
ICER	14-3-3 eta		Cytochrome C	BRCA2
Hes1	Inhibin α		Lactate dehydrogenase A	
C/EBPβ	MKP1		SOD2	
Nur77	GIP receptor		Hexokinase 2	
Egr1	VGF		Pyruvate carboxylase	
Per1	Glycoprotein α		VDAC2	
Per2	IGF-I		Spermine synthase	
Fra1	Class II MHC β chain		Fibronectin	
JunB			UCP-3	
			VMAT-1	

A partial list of genes not previously shown to be CREB targets (novel) and known CREB targets is presented.

tially to cAMP, NGF, and UV, suggesting that these genes are able to discriminate between activation signals. Whether this discrimination is due to differential regulation of CREB or to the existence of distinct transcriptional complexes is not known.

CREB Regulates Bidirectional Transcription

Although mammalian promoters are typically thought to regulate single genes, recent analyses of human and mouse genomes has revealed that 5%–10% of genes are organized in a head-to-head arrangement (Adachi and Lieber, 2002; Kiyosawa et al., 2003; Trinklein et al., 2004). Because the transcriptional start sites and promoters of these bidirectional genes are poorly characterized, it is not known whether they share regulatory elements. Of the duplicated SACO loci that define the 5′ flanking regions of genes or ECgene transcripts, approximately 25% predict bidirectional transcripts separated by less than 2 kb (Figure 3D). The two divergent genes comprising a bidirectional promoter are typically within 2 kb of a CpG island (94%). Interestingly, GSTs located in bidirectional promoters often contain multiple CREs.

Because the incidence of SACO loci in bidirectional promoters was so high, CREB may play a particularly important role in their coordinate regulation. To test this idea, we used quantitative PCR to determine whether some of these bidirectional transcripts were induced by cAMP or other CREB activators. As shown in Figure 5A, mRNA levels of two pairs were significantly stimulated by forskolin, and these increases were attenuated by ACREB (Figure 5A and data not shown). Two additional bidirectional pairs showed a reduction in their overall mRNA levels in response to ACREB (data not shown). Of interest, there is EST evidence of bidirectional transcription for several known immediate-early CREB targets. One such example is a spliced EST that represents a bidirectional partner of the c-*fos* gene (Figure 5B). Basal mRNA levels for this novel gene and c-*fos* are far lower than for many other bidirectional genes. Forskolin markedly stimulated transcription of both genes, and these increases were attenuated by expression of ACREB. The ability of a single transcription factor to regulate pairs of genes coordinately likely has important biological implications.

Antisense Transcription

Antisense transcripts have been shown to have important functions in gene regulation in both eukaryotes and prokaryotes. Recent studies suggest that there are between 1600 and 2500 sense-antisense transcript pairs in the human and mouse genomes (Kiyosawa et al., 2003; Shendure and Church, 2002; Yelin et al., 2003). Interestingly, 4% of our duplicated SACO loci were located near the 3′ ends of genes, and many additional genes had loci at both the 5′ and 3′ ends. In some cases, these loci were adjacent to ESTs or mRNAs that were clearly antisense, based on splicing and/or polyadenylation signals. We used strand-specific quantitative PCR to confirm whether some of these novel 3′ RNAs were transcribed in the antisense orientation. These studies showed that four of four were transcribed in an antisense orientation (Figure 6). Thus, some SACO loci delineate novel antisense transcripts that could be missed by hybridization to conventional oligonucleotide arrays. Whether these transcripts are actually regulated by CREB has not been determined.

Discussion

A variety of in vitro methods has been used to identify and characterize transcription factor binding sites. The presence of a binding site in genomic DNA does not necessarily indicate whether that site is occupied in vivo, however. Multiple epigenetic mechanisms, including DNA methylation and nucleosome position, can influence whether a particular regulatory element is functional. In addition, DNA sequences flanking a potential regulatory element can affect transcription factor binding by inducing conformational changes or allowing interaction with competing transcription factors. Moreover, multiple transcription factors can compete for the same site. Several of these mechanisms have been shown to influence CREB binding to the CRE (Cox et al., 1995; Fink et al., 1991; Iguchi-Ariga and Schaffner, 1989; Schild-Poulter et al., 1996). Consequently, assessing whether a regulatory element is occupied in vivo

Table 3. Affymetrix Analysis of cAMP-Regulated Genes

Known Targets	SACO	Novel Targets	SACO
c-fos	+	BAMBI	+
Nurr1	+	Cited2	+
Nur77	+	Rgs2	−
NOR1	+	Id1	+
EGR1	+	Claudin 7	+
C/EBPβ	+	Id3	+
Per1	+	ARC	+
c-jun	+	DAP-4	+
N-NOS	+	Nudel	+
JunD	+	TIEG1	+
JunB	+	Bhlhb2	+
PDE4D	+	Nup98	+
inhibin α	+	Onecut	+
tyrosine hydroxylase	+	βIII tubulin	+
presenilin-1	+	MKP5	+
glycoprotein α	+	Nfil3	−
SSTR2	+	GDF15	+
Btg2	+	Id2	−
DβH	+	GSK3β	+
secretogranin II	+	IRS1	+
tPA	+	DAP-1	+
c-maf	+	Metallothienen	+
Hes1	+	Synaptotagmin 13	+
synaptotagmin IV	+	Neuregulin	+
HMG Co-A synthase	+	Caveolin	−
MAT IIA	+	Hes3	+
		Nueropilin-2	+
		GABA B rec. 1	+
		β arrestin 1	+
		Ninjurin	+
		Myd116	+
		BMP2	+
		CIRBP	+
		Oazi	+
		VEGF	+
		Klf4	−
		Dyrk1a	+
		Mef2D	−
		RhoB	+
		PP5	+
		Synaptotagmin 7	+
		Pim1	+
		Bhlhb2	+
		Egl 9 homolog	+
		Synapsin 3	+

See Supplemental Data for experimental details. (Left) Selection of known cAMP-regulated genes whose mRNA levels were significantly increased by forskolin. "SACO" denotes the presence of a SACO locus within 2 kb (Affymetrix RAE230). (Right) Selection of genes showing significant forskolin-stimulated increases in mRNA but not previously known to be regulated by cAMP.

requires experimental confirmation, as provided by ChIP assays.

In general, ChIP assays utilize specific PCR primers flanking a suspected binding site to determine whether the site is occupied. Recent advances in this methodology, as exemplified by the ChIP-on-chip approach, can determine the identity of the regulatory element itself. The ChIP-on-chip assay depends upon an a priori knowledge of gene regulatory regions, however, and recent studies by Cawley et al. (2004) suggest that these regions, at least for mammalian genomes, are more complex than previously imagined. Their study additionally suggests that the ChIP-on-chip assay may miss most regulatory regions.

SACO was adapted from a procedure used previously to analyze the *Yersinia pestis* genome (Dunn et al., 2002). This method depends upon having a complete (or nearly complete) genome sequence but is not limited to analyzing 5′ flanking regions. SACO is based on the isolation and analysis of genomic signature tags, 21 base pair sequences that can be localized to unique sites in the mammalian genome. Unlike the tiled microarrays used by Cawley et al. (2004), which considered only genes on chromosomes 21 and 22, SACO can interrogate the entire genome, providing a significant advantage for identifying transcriptional regulatory elements. Although the requirement for extensive DNA sequencing is an apparent disadvantage of SACO (and SAGE), this is more than compensated for by the comprehensive information obtained. The amount of potential background binding in ChIP assays would seem to preclude this type of analysis, and we devoted considerable effort to maximizing specificity of the ChIP starting material in constructing our SACO library. It may, in fact, be possible to utilize less purified material if one is willing to do more DNA sequencing. Requiring duplication of each GST increased the confirmation rate in repeat ChIP assays to 100%. Like SAGE, SACO should also be quantitative if sufficient GSTs are analyzed.

The CRE was originally identified as an eight nucleotide palindrome in the somatostatin promoter (Montminy et al., 1986), and similar sequences have been detected in hundreds of other genes. The palindromic sequence can be identified in random oligonucleotide selection assays when CREB is used as a probe (Paca-Uccaralertkun et al., 1994), but this type of assay has not been particularly informative for characterizing transcriptional targets in native genes. Many genes contain CREs that include only a portion of the palindrome, and abundant evidence supports the idea that these variant CREs nonetheless interact with CREB and mediate gene activation through CREB-dependent pathways. The affinity of CREB for the eight base pair palindrome is considerably higher than that for some of the variant sequences (Craig et al., 2001), but there is no reason, a priori, to believe that the highest-affinity sites would be represented preferentially throughout the genome. Indeed, deviations from the palindromic sequence, along with correspondingly weaker affinities for CREB, tend to be conserved evolutionarily, suggesting a biological value for the weaker binding. A central tenet in the CREB model is that CREB binds to the CRE constitutively. This idea is not consistent with observations that cAMP appears to activate distinct sets of genes in different cell types, however. Thus, we hypothesize that CREB binding sites might only be occupied under certain conditions and that the population of binding sites, the CREB regulon, could differ from one cell type to another. The SACO method provides an ideal approach for testing this hypothesis.

Recently, a hidden Markov model was trained using ten well-characterized CREs to interrogate the human genome for novel CREB-responsive motifs (Conkright et al., 2003). Although this approach yielded 75 sites (of which ten were experimentally confirmed), virtually all the sites conformed to the palindromic sequences used to train the model. The large number of experimentally confirmed CREB binding sites identified in our study should allow us to redefine what sequences are required for CREB interaction in vivo. Of note, some of the con-

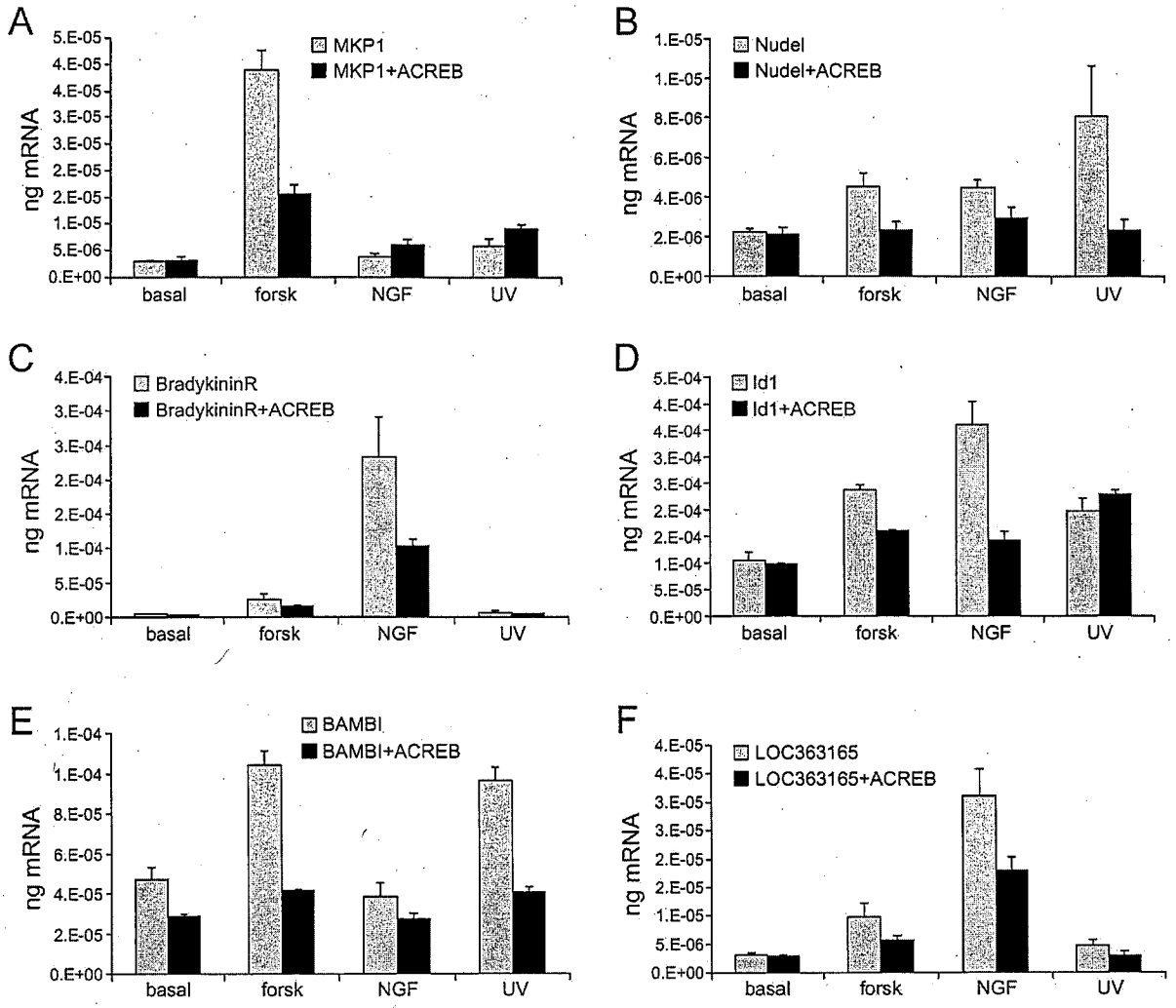

Figure 4. CREB Regulation of cAMP-Inducible Genes

Genes examined were shown to be induced by forskolin by Affymetrix. PC12 cells were electroporated with vector or the dominant-negative CREB mutant ACREB. Two days after transfection, cells were stimulated with the indicated agonists or UV-B. One hour later, RNA was isolated, reverse transcribed, and analyzed by real-time quantitative PCR; mRNA levels were normalized to 18S RNA. Data represents 6–10 replicates; error is SEM. (MKP1, MAP kinase phosphatase 1, BAMBI, BMP, and activin membrane bound inhibitor; LOC363165 is an unknown gene.)

sensus binding sites proposed by Euskirchen et al. (2004) on the basis of a ChIP-on-chip analysis of chromosome 21 were not confirmed in our study (data not shown). This difference probably relates to the paucity of CREB binding sites on chromosome 21 and the depth of sequencing of our SACO library.

A difference between our genome-wide study and that of Cawley et al. (2004) was the localization of transcription factor binding sites. Cawley et al. (2004) determined that only 22% of the binding sites were at expected promoter locations. When we considered the population of CREB binding loci identified by duplicate GSTs, we found that 40% were near the 5′ flanking regions of annotated genes. This set of GSTs was markedly enriched for CRE motifs and CpG islands. Another 23% of loci were upstream from ECgene transcript evidence. Thus, at least 63% of the CREB binding sequences were 5′ to expected start sites of transcription. While not as prevalent as described by Cawley et al. (2004), 4% of CREB binding sites were detected in 3′ locations. It is likely that some of these elements drive expression of antisense transcripts and are thus upstream from transcription start sites as well. The biological roles of these antisense transcripts and their mechanisms of activation have yet to be determined. Nonetheless, we suggest that the vast majority of CREB binding sites are upstream from promoters. In many instances, the ESTs adjacent to CREB binding sites had little coding potential, suggesting that a subset of CREB-regulated genes may not encode proteins. Although prediction of noncoding mRNAs is difficult, eight duplicated SACO loci were closely associated with previously described miRNAs. The presence of CREB binding sites in genomic regions outside annotated promoters supports the idea that the transcriptome is larger than previously envisioned.

Interestingly, although a few duplicated GST loci were

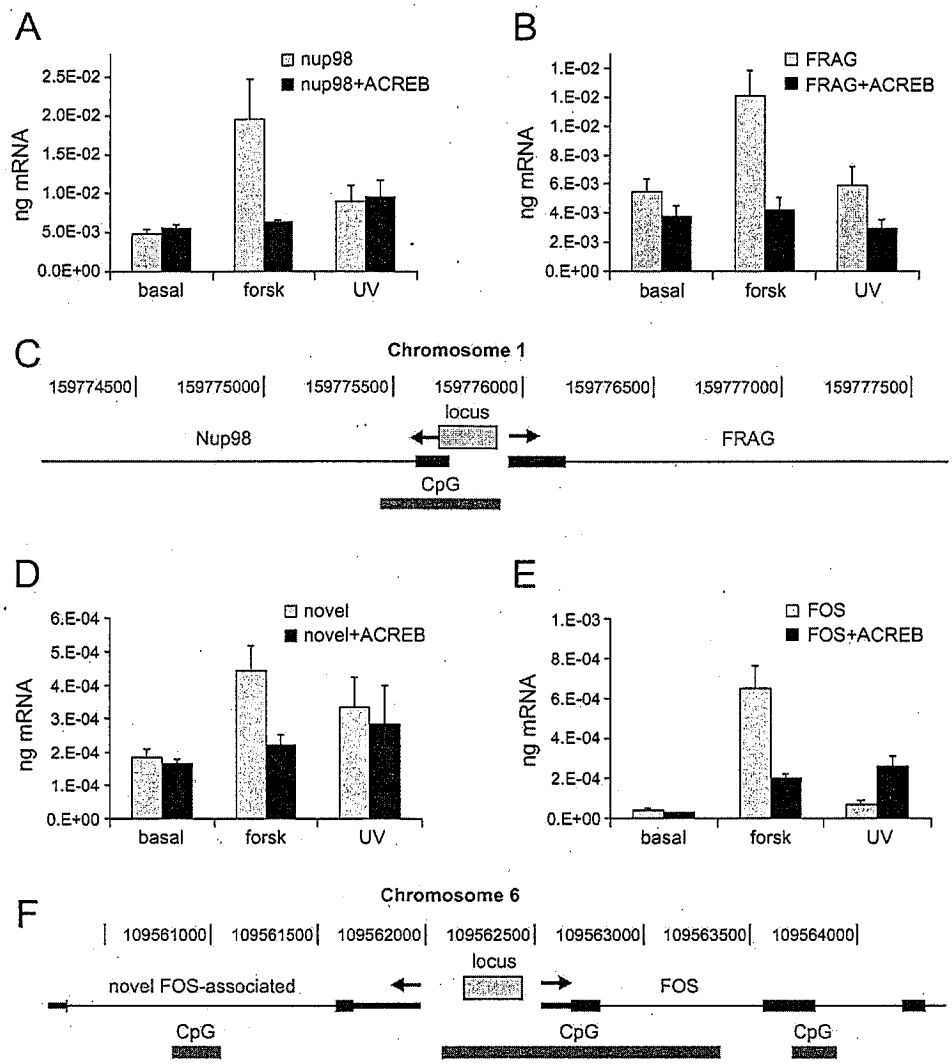

Figure 5. Quantitative PCR Analysis of Bidirectional Genes

(A, B, D, and E) (A) and (B) designate a gene pair on chromosome 1. (D) and (E) designate a gene pair on chromosome 6. PC12 cells were electroporated with vector or ACREB. Two days after transfection, cells were stimulated with the indicated agonists or UV-B. One hour later, RNA was isolated, reverse transcribed, and subjected to real-time quantitative PCR analysis using primers specific for the indicated genes; mRNA levels were normalized to 18S RNA. Data represents 6–10 replicates; error is SEM. (C) and (F) depict the positions of the SACO loci in the two gene pairs. Nup98 is a nucleoporin; FRAG is an interferon-stimulated gene. The novel c-*fos*-associated transcript is the syntenic ortholog of RIKEN cDNA 1810048J11.

not adjacent to discernable CRE-like sequences, we were able to confirm that several of these loci interact with CREB in vivo. Although these regions could define a novel CREB recognition sequence, they may also represent indirect interactions (Giebler et al., 2000; Van Orden et al., 1999).

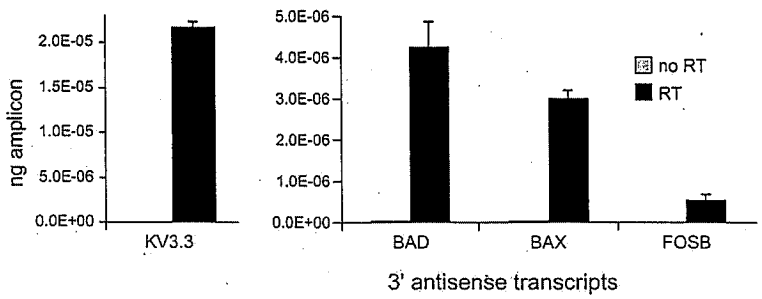

Figure 6. Quantitative PCR Analysis of Antisense Transcription

Cytoplasmic RNA was isolated and reverse transcribed with strand-specific primers. Real-time quantitative PCR analysis was performed using primers to antisense transcripts at the 3′ ends of the indicated genes. Data represents nine replicates; error is SEM.

Surprisingly, the set of 32,723 loci supported by a single GST was less likely (46%) than the duplicated loci to be associated with the 5′ ends of annotated genes or ECgene transcripts. These loci were also less frequently localized near CRE motifs or CpG islands, but 69% could be confirmed by a separate ChIP assay. Thirty percent of these singly detected loci were located within an annotated gene, suggesting that they may tend to identify internal regulatory regions. The basis for the discrepancy between the localization of singly and multiply identified loci is unclear. One possibility is that exchange of nucleosomes during transcriptional elongation could decrease the frequency of CREB interactions at internal sites. Whether these internal regions represent alternative promoters, distal enhancers, or function in some novel manner awaits a detailed analysis of transcriptional activity surrounding these regions. Because several known internal promoters are associated with single GST loci (TrkB, nNos, and PDE4), it is likely that at least some of these loci function as conventional promoters.

Another surprising finding was the frequent occurrence of CREB binding sites in bidirectional promoters. Although bidirectional promoters are relatively common in mammalian genomes (Adachi and Lieber, 2002; Kiyosawa et al., 2003; Trinklein et al., 2004), the observation that 25% of loci associated with transcriptional start sites localized to such promoters was unexpected. In some instances, as exemplified by c-*fos*, one of the two components was a well-characterized gene. In others, both components were unknown. The particular sets of genes comprising a bidirectional pair tend to be conserved evolutionarily, suggesting that this arrangement has biological importance. In the few gene pairs that we have analyzed, regulation of the two components appears to be coordinated. Disparate regulation of transcripts from bidirectional promoters has also been observed, however (Momota et al., 1998; Schuettengruber et al., 2003). Further studies are clearly needed to elucidate the biological implications of this interesting genomic organization.

Our Affymetrix analysis indicated that hundreds of genes were induced by activation of the cAMP pathway, but even more contained CREB binding sites. Our data differed from that of Euskirchen et al. (2004) in that most of the genes upregulated by forskolin in our study contained CREB binding sites. Because the Affymetrix screen was performed at a single time point, it is possible that we detected only a fraction of the induced genes. Alternatively, microarray analysis may be too insensitive to detect the full complement of cAMP-activated genes. Three other possibilities should also be considered, however. First, the expectation that all cAMP-activated genes will behave like immediate-early genes, with an extremely low basal level of expression, short mRNA half-life, and highly stimulated induced level, may be unjustified. Many genes may be induced by cAMP in a more subdued fashion. Second, it is also likely that, for some genes, changes in nuclear transcription may not be easily quantifiable at the level of cytoplasmic mRNA. Third, cAMP may not always be the most important signal for stimulating CREB activity. We found, for example, that NGF activated some CREB binding genes but forskolin did not. If CREB is indeed the target of multiple signaling pathways, it is likely that, in particular contexts, some pathways will be more effective than others. It will be interesting to determine whether cAMP signaling leads to CREB phosphorylation in all genomic contexts.

Finally, given the comprehensive nature of our screen, it is also interesting to note that some well-studied CREs were not represented in our library. For example, the somatostatin CRE, which binds to CREB with a very high affinity in vivo, was not detected by even a single GST. This suggests that CREB binding is not constitutive.

Experimental Procedures

Reagents

NGF (Chemicon, Temecula, CA) and KCl (Sigma), were dissolved in dH_2O. Forskolin (Sigma, St. Louis, MO) and 1,9 dideoxy-forskolin were dissolved in DMSO. One hour treatments were at the following concentrations: forskolin, 10 μM; NGF, 100 ng/ml; KCl, 30 mM. UV-B treatment was 300 J/m^2.

Plasmids

GFP and ACREB (Ahn et al., 1998) were cloned into the pCAG vector (Niwa et al., 1991) and verified by automated DNA sequencing. Details on the subcloning will be provided on request.

Cell Culture

PC12 cells (passage 20–30) were grown in high glucose Dulbecco's modified Eagle's medium (Invitrogen) with 10% neonatal calf serum and 5% fetal bovine serum plus 100 units/ml penicillin and 100 units/ml streptomycin. Cells were serum starved for 12–16 hr prior to drug treatments. Cultures were maintained at 37°C and 5% CO_2.

Transfection

PC12 cells (5×10^6) were electroporated using the Nucleofector kit according to the manufacturer's protocol (Amaxa, Köln, Germany) except that 5 μg of DNA was used. Separate transfection of a green fluorescent protein-expressing vector showed that 90%–100% of the cells were transfected.

ChIP

PC12 cells (5×10^5) were serum starved for 12–16 hr in a 10 cm dish, then stimulated with 10 μM forskolin (Sigma) for 15 min. Cells were fixed in 1% formaldehyde/1× PBS for 20 min at 25°C. Formaldehyde was removed, 2 ml harvesting buffer (100 mM Tris-HCl [pH 9.4] and 10 mM DTT) was added, and the dishes were placed on ice. Cells were centrifuged at 2000 × g for 5 min at 4°C. Cell pellets were washed with 1 ml ice-cold 1× PBS, resuspended in 0.6 ml lysis buffer (0.1% SDS, 0.5% Triton X-100, 20 mM Tris-HCl [pH 8.1], 150 mM NaCl, 1× protease inhibitor cocktail; Roche Molecular Biochemicals, Indianapolis, IN), sonicated to an average size of 600 bp using a Misonex cup horn sonicator (5 × 30 s 140–150 W pulses with 30 s rest intervals in ice water). Protein A-Sepharose beads were washed three times for 10 min at 4°C with 10 mM Tris-HCl (pH 8), then blocked overnight with 10 mM Tris-HCl (pH 8), 1 mM EDTA, 0.5% BSA, and 19.2 μg/ml glycogen. The sonicated chromatin was clarified by centrifugation for 10 min at 13,000 rpm at 4°C. Supernatants were transferred to fresh tubes and precleared with 80 μl of 50% protein A-Sepharose (Repligen) slurry by rocking for 1 hr at 4°C. Immunoprecipitation was performed overnight at 4°C with 5 μg anti-CREB antibody (Moreno et al., 1999) or affinity purified polyclonal antibody, β-gal (5′ 3′ Inc.) as a control. Immune complexes were captured with 80 μl of 50% protein A-Sepharose slurry for one hour at 4°C. The beads were collected by centrifugation at 7600 rpm for 1 min and washed as follows: four times in lysis buffer for 10 min, once in LiCl buffer (0.25 M LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA [pH 8.0], 10 mM Tris-HCl [pH 8.1]), once with 1× TE (pH 8.1) for 30 min, and a final wash with 1× TE for 5 min. CREB-DNA complexes were eluted from the beads by adding 100 μl elution buffer (1% SDS and 0.1 M $NaHCO_3$ [pH 8.0]) and rotating at room temperature for 15 min. Beads were collected by centrifugation at

13,000 rpm for 1 min, the eluates were transferred to a new tube, and the elution was repeated using 50 μl elution buffer. Eluates were pooled and heated at 65°C overnight to reverse crosslinking. DNA fragments were purified using a Qiagen PCR purification kit (Qiagen, Valencia, CA) and quantified by real-time PCR. For GST confirmation, large-scale ChIP assays were performed using 6×10^7 cells.

SACO Library

ChIP DNA was blunt ended with the DNA Terminator End Repair Kit (Lucigen), phenol/chloroform extracted, and EtOH precipitated. A duplex DNA cassette (50 pmoles), cassette P (top, 5′ TTTTCCG GTCTACTGAATTCCGAACTGCCAC; bottom, GTGGCAGTTCGGAA TTCAGTAGACCGGA), was ligated to blunt-ended ChIP DNA using 1 μl (2000 U) of T4 DNA ligase (NEB). The ligations were purified using a Qiaquik PCR purification column (Qiagen), and adaptor-ligated ChIP material (1 μl) was PCR amplified in a 100 μl reaction containing 200 pmols P-oligo (5′ bio-CCGGTCTACTGAATTCCG AAC), 2 μl platinum *Taq* polymerase (Invitrogen), 2 mM $MgCl_2$, and 0.75 mM dNTPs (Roche) for one cycle at 95°C, 50 s; 20–25 cycles at 94°C, 15 s; 55°C–60°C, 30 s; 68°C, 2 min, followed by one cycle at 68°C for 7 min. Approximately 10 μg of amplified ChIP DNA was digested with 6 μl (60 U) NlaIII (NEB) in a 200 μl reaction at 37°C for 2 hr. The digests were phenol/chloroform extracted and EtOH precipitated. A modified version of the Long-SAGE protocol (Saha et al., 2002) was used to create ditags. Briefly, 12 μl of NlaIII-digested, amplified ChIP DNA was added to 36 pmol Long-SAGE adaptor A (top, 5′ TTTGGATTTGCTGGTGCAGTACAACTAGGCTTAATATCCG ACATG; bottom, 5′ PO_4-TCGGATATTAAGCCTAGTTGTACTGCAC CAGCAAATCC [C7] amino) and 2 μl 10× T4 ligase buffer. In a second tube, 12 μl NlaIII-digested, amplified ChIP DNA was added to 36 pmol Long-SAGE adaptor B (top, 5′ TTTCTGCTCGAATTCAAGCT TCTAACGATGTACGTCCGACATG; bottom, 5′ PO_4-TCGGACGTACA TCGTTAGAAGCTTGAATTCGAGCAG [C7] amino) and 2 μl 10× T4 ligase buffer. Reactions were heated to 50°C for 2 min, 1 μl high-concentration T4 ligase was added, and ligations were incubated for 12 hr at 16°C. Streptavidin-coated magnetic beads (200 μl) (Dynal M280) were washed with 800 μl 1× BW buffer (5 mM Tris [pH 8.0], 1M NaCl, 0.2 mg/ml BSA, 0.5 mM EDTA) and the beads resuspended in 200 μl 2× BW buffer. The ligations were purified over a Qiaquik PCR column, eluted in 200 μl DEPC water, and added to the beads. The tubes were mixed for 1 hr at 25°C, and beads were washed six times for 5 min with 1× BW buffer followed by three washes with 10 mM Tris [pH 8.0]. To release the ditags, the ligation products were cleaved with MmeI (NEB) as follows: the beads were washed twice with 200 μl of 1× NEB buffer 4/1× S-adenosyl methionine (SAM) (1.0 μl, 32 nM SAM in 800 μl 10× NEB4) resuspended in 70 μl of 10 mM Tris (pH 8.0). 1X NEB4 (10 μl), 10 μl 10× SAM (1 μl 32 mM SAM diluted in 79 μl DEPC), and 10 μl MmeI were added and the samples incubated for 2.5 hr at 37°C. The supernatants were pooled, and 100 μl 10 mM Tris-HCl (pH 8.0) was added to yield 300 μl. The supernatant was phenol/chloroform extracted; 200 μl was set aside in one tube, and 100 μl of DEPC water was added to the remaining volume to yield 200 μl (this tube served as the no ligase control). The sample and the no ligase control were EtOH precipitated, and the pellets were air dried. The sample pellet was ligated using high-concentration T4 DNA ligase for 12 hr at 25°C. From this point on, the protocol mirrored the Long-SAGE protocol (Saha et al., 2002) with the following exceptions: concatemers were isolated using the alternate agarose gel method; NEB T4 DNA ligase (2000 U/μl) and GlycoBlue (Ambion) were used in place of T4 DNA ligase and glycogen. A detailed protocol is in Supplemental Data.

Reverse Transcription

PC12 cells ($1–5 \times 10^4$ were treated as described, and total RNA was isolated using Trizol (Invitrogen) or an RNeasy kit (Qiagen) according to manufacturer's instructions. RNA (50 ng to 3 μg) was reverse transcribed with Moloney murine leukemia virus reverse transcriptase (Invitrogen) and 50–250 ng random primers (Invitrogen).

Quantitative PCR

Primers were designed using MIT's Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) with default parameters except the following: rodent and simple repeat library was on, product size was 50–200 bp, primer size was 18–27 bases, Tm was 66°C–72°C, maximum self-complementarity was 5, and maximum 3′ complementarity was 3. For ChIP confirmation, primers were designed to encompass TGACG CRE half-sites within 1 kb of a SACO locus. In some cases, multiple primers were designed for each locus. For analysis of cDNA levels near a tag, primers were selected within the nearest downstream UTR or exon that has extensive mRNA or spliced EST evidence. Primer sequences are available on request. PCRs (10 μl) contained 1 μl 10× PCR buffer (Invitrogen), 2.5 mM $MgCl_2$, 200 μM dNTP (Roche), 0.125–0.25 μM primer (IDT), 1× SYBR green I (Invitrogen), and 1 U platinum *Taq* (Invitrogen). PCR was run on an Opticon OP346 (MJ Research) for one cycle at 95°C, 35 s, and 30–50 cycles at 94°C, 15 s; 68°C–70°C, 40 s. CREB ChIPs were expressed as either ng of gel-purified (Qiagen) amplicon or percent of input DNA and were normalized for LINE L1 DNA. Samples showing greater than 2-fold enrichment relative to an IgG control were considered confirmed. RT-PCR experiments were normalized to 18S RNA levels (other housekeeping genes showed similar results) and were expressed as ng of gel-purified (Qiagen) amplicon. All RT-PCR generated 100-fold higher levels of product than no reverse transcriptase controls. All PCRs showing multiple products on a 3%–4% Nu-Seive gel or nonlinear amplification were excluded. A subset of products were sequenced to confirm that they represented the expected amplicon.

Affymetrix Oligonucleotide Array

PC12 cells were treated with 10 μM forskolin or 1,9 dideoxy forskolin (Sigma) for 1 hr. Two biological replicates were used to prepare ~10 μg of total RNA using both the Trizol (Invitrogen) and RNAeasy (Qiagen) kits. Each sample was analyzed by the OHSU Gene Microarray Shared Resource (GMSR) Affymetrix Microarray Core (www.ohsu.edu/gmsr/amc). Labeled samples were hybridized to both the RAE 230A and 230B arrays. Robust multiarray average (RMA) (Irizarry et al., 2003) probe (low) level analysis was implemented using the Affy package from the Bioconductor version 1.3 framework (http://www.bioconductor.org/). Further details on the hybridization and statistical analysis are in Supplemental Data.

Bioinformatics

Concatemer sequences were extracted from chromatograms with the base caller "phred" using recommended settings (Ewing and Green, 1998). A custom perl script separated ditags at all CATGs. The resulting SACO GSTs were matched to genomic CATG sites using a C program. GSTs with exact matches or matches with one substitution error that were uniquely assignable to a genomic location were considered. GSTs without a unique genomic match or with multiple unique matches were not considered. GSTs within 2 kb of each other were taken to be associated with the same CREB locus. A set of scripts that automates the analysis of SACO data is available at http://genome.bnl.gov/SACO/. The CREB SACO data is available at http://saco.ohsu.edu/public/data. Additional details on the analysis of SACO loci are in Supplemental Data.

Acknowledgments

We thank David Ginty for anti-CREB antisera used in preliminary experiments and Yasuhiro Yoshimatsu and Ryan Cleland for analysis and discussion. We also thank Qinghong Zhang and Karl Obrietan for their comments on the manuscript. This work was supported by grants from the NIH, DOE, and Howard Hughes Medical Institute.

Received: July 20, 2004
Revised: September 14, 2004
Accepted: October 13, 2004
Published: December 28, 2004

References

Adachi, N., and Lieber, M.R. (2002). Bidirectional gene organization: a common architectural feature of the human genome. Cell *109*, 807–809.

Ahn, S., Olive, M., Aggarwal, S., Krylov, D., Ginty, D.D., and Vinson, C. (1998). A dominant-negative inhibitor of CREB reveals that it is a general mediator of stimulus-dependent transcription of c-fos. Mol. Cell. Biol. *18*, 967–977.

Cawley, S., Bekiranov, S., Ng, H., Kapranov, P., Sekinger, E., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A., 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.

Conkright, M., Guzman, E., Flechner, L., Su, A., Hogenesch, J., and Montminy, M. (2003). Genome-wide analysis of CREB target genes reveals a core promoter requirement for cAMP responsiveness. Mol. Cell *11*, 1101–1108.

Cox, J., Sloan, L., and Schepartz, A. (1995). Conformation of Tax-response elements in the human T-cell leukemia virus type I promoter. Chem. Biol. *2*, 819–826.

Craig, J.C., Schumacher, M.A., Mansoor, S.E., Farrens, D.L., Brennan, R.G., and Goodman, R.H. (2001). Consensus and variant cAMP-regulated enhancers have distinct CREB-binding properties. J. Biol. Chem. *276*, 11719–11728.

Dunn, J., McCorkle, S., Praissmna, L., Hind, G., Van Der Lelie, D., Bahou, W., Gnatenko, D., and Krause, M. (2002). Genomic signature tags (GSTs): a system for profiling genomic DNA. Genome Res. *12*, 1756–1765.

Euskirchen, G., Royce, T., Bertone, P., Martone, R., Rinn, J., Nelson, F., Sayward, F., Luscombe, N.M., Miller, P., Gerstein, M., et al. (2004). CREB binds to multiple loci on human chromosome 22. Mol. Cell. Biol. *24*, 3804–3814.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. *8*, 186–194.

Fass, D., Butler, J., and Goodman, R. (2003). Deacetylase activity is required for cAMP activation of a subset of CREB target genes. J. Biol. Chem. *278*, 43014–43019.

Fink, J.S., Verhave, M., Walton, K., Mandel, G., and Goodman, R.H. (1991). Cyclic AMP- and phorbol ester-induced transcriptional activation are mediated by the same enhancer element in the human vasoactive intestinal peptide gene. J. Biol. Chem. *266*, 3882–3887.

Giebler, H.A., Lemasson, I., and Nyborg, J.K. (2000). p53 recruitment of CREB binding protein mediated through phosphorylated CREB: a novel pathway of tumor suppressor regulation. Mol. Cell. Biol. *20*, 4849–4858.

Hiroi, H., Christenson, L.K., Chang, L., Sammel, M.D., Berger, S.L., and Strauss, J.F., 3rd. (2004). Temporal and spatial changes in transcription factor binding and histone modifications at the steroidogenic acute regulatory protein (stAR) locus associated with stAR transcription. Mol. Endocrinol. *18*, 791–806.

Hug, B.A., Ahmed, N., Robbins, J.A., and Lazar, M.A. (2004). A chromatin immunoprecipitation screen reveals protein kinase Cbeta as a direct RUNX1 target gene. J. Biol. Chem. *279*, 825–830.

Iguchi-Ariga, S.M., and Schaffner, W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes Dev. *3*, 612–619.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249–264.

Kiyosawa, H., Yamanaka, I., Osato, N., Kondo, S., and Hayashizaki, Y. (2003). Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. Genome Res. *13*, 1324–1334.

Lee, T., Rinaldi, N., Robert, F., Odom, D., Bar-Joseph, Z., Gerber, G., Hannett, N., Harbison, C., Thompson, C., Simon, I., et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science *298*, 763–764.

Lonze, B., and Ginty, D. (2002). Function and regulation of CREB family transcription factors in the nervous system. Neuron *35*, 605–623.

Mayr, B., and Montminy, M. (2001). Transcriptional regulation by the phosphorylation-dependent factor CREB. Nat. Rev. Mol. Cell Biol. *2*, 599–609.

McClung, C., and Nestler, E. (2003). Regulation of gene expression and cocaine reward by CREB and DeltaFosB. Nat. Neurosci. *6*, 1208–1215.

Momota, R., Sugimoto, M., Oohashi, T., Kigasawa, K., Yoshioka, H., and Ninomiya, Y. (1998). Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and alpha4(IV) collagen chains are arranged head-to-head on chromosome 2q36. FEBS Lett. *424*, 11–16.

Montminy, M., Sevarino, K., Wagner, J., Mandel, G., and Goodman, R. (1986). Identification of a cyclic-AMP-responsive element within the rat somatostatin gene. Proc. Natl. Acad. Sci. USA *83*, 6682–6686.

Moreno, C.S., Beresford, G.W., Louis-Plence, P., Morris, A.C., and Boss, J.M. (1999). CREB regulates MHC class II expression in a CIITA-dependent manner. Immunity *10*, 143–151.

Nichols, M., Weih, F., Schmid, W., DeVack, C., Kowenz-Leutz, E., Luckow, B., Boshart, M., and Schutz, G. (1992). Phosphorylation of CREB affects its binding to high and low affinity sites: implications for cAMP induced gene transcription. EMBO J. *11*, 3337–3346.

Niwa, H., Yamamura, K., and Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene *108*, 193–199.

Odom, D., Zizlsperger, N., Gordon, D., Bell, G., Rinaldi, N., Murray, H., Volkert, T., Schreiber, J., Rolfe, P., Gifford, D., et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. Science *303*, 1311–1312.

Paca-Uccaralertkun, S., Zhao, L., Adya, N., Cross, J., Cullen, B., Boros, I.M., and Giam, C.Z. (1994). In vitro selection of DNA elements highly responsive to the human T-cell lymphotropic virus type I transcriptional activator, Tax. Mol. Cell. Biol. *14*, 456–462.

Roh, T.Y., Ngau, W.C., Cui, K., Landsman, D., and Zhao, K. (2004). High-resolution genome-wide mapping of histone modifications. Nat. Biotechnol. *22*, 1013–1016.

Saha, S., Sparks, A.B., Rago, C., Akmaev, V., Wang, C.J., Vogelstein, B., Kinzler, K.W., and Velculescu, V.E. (2002). Using the transcriptome to annotate the genome. Nat. Biotechnol. *20*, 508–512.

Schild-Poulter, C., Sassone-Corsi, P., Granger-Schnarr, M., and Schnarr, M. (1996). Nucleosome assembly on the human c-fos promoter interferes with transcription factor binding. Nucleic Acids Res. *24*, 4751–4758.

Schuettengruber, B., Doetzlhofer, A., Kroboth, K., Wintersberger, E., and Seiser, C. (2003). Alternate activation of two divergently transcribed mouse genes from a bidirectional promoter is linked to changes in histone modification. J. Biol. Chem. *278*, 1784–1793.

Shendure, J., and Church, G.M. (2002). Computational discovery of sense-antisense transcription in the human and mouse genomes. Genome Biol. *3*. Published online August 22, 2002. 10.1186/gb-2002-3-9-research0044

Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. Genome Res. *14*, 62–66.

Van Orden, K., Giebler, H.A., Lemasson, I., Gonzales, M., and Nyborg, J.K. (1999). Binding of p53 to the KIX domain of CREB binding protein. A potential link to human T-cell leukemia virus, type I-associated leukemogenesis. J. Biol. Chem. *274*, 26321–26328.

Vo, N., and Goodman, R. (2001). CREB-binding protein and p300 in transcriptional regulation. J. Biol. Chem. *276*, 13505–13508.

Weinmann, A., Bartley, S., Zhang, T., Zhang, M., and Farnham, P. (2001). Use of chromatin immunoprecipitation to clone novel E2F target promoters. Mol. Cell. Biol. *21*, 6820–6832.

Wolfl, S., Martinez, C., and Majzoub, J.A. (1999). Inducible binding of cyclic adenosine 3′,5′-monophosphate (cAMP)-responsive element binding protein (CREB) to a cAMP-responsive promoter in vivo. Mol. Endocrinol. *13*, 659–669.

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., et al. (2003). Widespread occurrence of antisense transcription in the human genome. Nat. Biotechnol. *21*, 379–386.

Accession Numbers

The GEO accession number for the Affymetrix array data reported in this paper is GSE2071.