A comparative genomics approach to identifying the plasticity transcriptome

Carnegie Mellon University School of Computer Science Senior Thesis

Andreas R. Pfenning

Department of Computer Science, Carnegie Mellon University, Pittsburgh PA 15213

Abstract

The aim of this project is to use computational methods to analyze the set of genes expressed in response to neural activity. When a neuron receives a stimulus from seizure or learning and memory, the amount of certain proteins that the cell produces increases through a mechanism called gene expression. This change is governed by the binding of transcription factors to DNA at specific sequences near the gene. Using a probabilistic model and a comparison between human and mouse, we identified a set of genes with CREB, zif268, and AP-1 transcription factor binding sequences. The presence of those sequences are known in many cases increase gene expression and protein amount. This set genes provides information about the role of the transcription factors as well as a resource for biologists looking to build specific networks of protein activation. We found that the transcription factors CREB and zif268 are likely to bind near genes involved in regulatory networks. These results are compared to studies of gene expression following seizure. This work is a crucial first step in using computational methods to study learning and memory at the level of the neuron.

Introduction

One of the fundamental problems in neurobiology is how the properties of a neuron can change in response to electrical and chemical activity. The concept of altering a neuron's properties by signaling is known as plasticity. When a neuron receives an electrical stimulus, particularly a strong one, chemical signals are initiated in the neuron. These chemical signals cause proteins called transcription factors, such as CREB, zif268, and AP-1 to be activated. An activated transcription factor can bind to

DNA near the beginning of a gene called a promoter region and cause an increase or decrease in the protein that the gene encodes. This change in protein level can then affect the properties of a neuron. This process is known as activity dependent gene expression. We call the set of genes altered by activity the plasticity transcriptome.

Elucidating the plasticity transcriptome is essential for a good understanding of any biological process where neuron connections and properties are altered. The most widely studied examples are epileptogenesis, brain injury, and learning and memory. Because of the widespread increase in activity, seizure is useful model in understanding activity-dependent changes in gene expression. Following seizure, changes in gene expression facilitate the development of epilepsy disorders (reviewed by (Elliott and Lowenstein 2004; Rakhade et al. 2005). Understanding the plasticity transcriptome is also crucial in studying aspects of learning and memory. For instance, mental retardation is linked to changes in activity-dependent gene expression, such as Rubenstein-Taybi syndrome, related to a mutation in the CREB-binding protein, and Rett syndrome, tied to a defect in a DNA-binding protein that regulates the correct timing of expression of many downstream genes (Hong et al. 2005). Knowing the targets of activity dependent transcription factors allows researchers to build models of these disorders that could lead to treatments.

Experimentally based approaches dominate the literature in identifying genes that are differentially expressed. Microarrays that measure the amount of specific DNA transcribed have been frequently employed to characterize the plasticity transcriptome by identifying candidate genes that are upregulated after activity (Befort et al. 2003; Costigan et al. 2002; Laifenfeld et al. 2002; Lee et al. 2005; Luo et al. 2001; Valerio et

al. 2004; Yao et al. 2004), typically through pharmacologically-induced seizure (Del Rio and Barlow 2002; Flood et al. 2004; Hunsberger et al. 2005; Lukasiuk and Pitkanen 2004; Tang et al. 2001; Wilson et al. 2005). There are several reasons microarray technologies are insufficient for understanding activity-dependent gene expression. The large amount of noise inherent to the method only allows the detection of large changes in gene expression. Subtle, but important differences are lost. This problem is amplified by the heterogeneity of brain tissue. Even a large change in gene expression could be overlooked if it only occurred in small fraction of the neurons analyzed. Neuron-specific pathways of gene expression could also make the results sensitive to slight differences in the location and timing of neural tissue collected for the microarray. Furthermore, studies in Drosophila melanogaster suggest that the plasticity transcriptome is dependent on the method of seizure induction(Guan et al. 2005). This can cause inconsistent results based on the methods of different research groups. Finally, microarrays yield little information on which transcription factors affect which genes.

An alternate experimentally based approach known as chromatin-immunoprecipitation (ChIP) against specific transcription factors and determination of immunoprecipitated DNA binding sites by hybridization to microarrays or PCR analysis (Hakimi et al. 2002; Impey et al. 2004; Israsena et al. 2004; Sun et al. 2005; Vanderluit et al. 2004). Unlike microarrays, ChIP is able to study the binding specific transcription factors to genes. However, much like microarrays, it sufferers from tissue-specific occupancy of binding sites and the inherent heterogeneity of cell types in brain tissue (Cha-Molstad et al. 2004).

A non-genomic based approach to identifying potential activity-regulated genes has been to isolate specific candidates and examine changes in the abundance of their protein or mRNA after manipulations of activity (see, for example, Amadio et al. 2004 and Hoffman 2003). However, this approach is obviously limited because it must proceed in a highly directed, case-by-case manner, and does not accommodate the discovery of novel or unlikely gene candidates.

Identification of activity-regulated genes could be improved by genomic screens that are unbiased by average expression level, cell type, or target gene preselection. The emergence of online databases has made large amounts of data available on transcription factor binding as well as promoter DNA sequences near the beginning genes where the transcription factors act. Methods for locating transcription factor binding sites often rely upon relatively simple comparisons of a single sequences or small set of specific "consensus binding sites" with individual promoter regions (Bulyk 2003; Qiu 2003; Vavouri and Elgar 2005). Although such searches can be productive, the use of a single consensus site is too simplistic. In many cases, these searches will over- or underestimate the number binding sites that exist. This results in subjective target identifications whose reliability is difficult to judge.

As an alternative, we use a computationally sophisticated model of three transcription factor binding sites CREB, zif268, and AP-1. The probabilistic representations of these site are shown in figure 1. These transcription factors are all linked to processes involving increased neural activity (Herdegen and Leah 1998). CREB activation is necessary for the consolidation of long-term memory (Yin et al. 1994). Zif268 and the components of AP-1, fos and jun, are often used as markers of increased activity or neural plasticity. Because activity-dependent changes in gene transcription are

linked to memory consolidation and also occur as a response to pathological conditions such as seizure (Corriveau et al. 1998; Guan et al. 2005; Nedivi et al. 1993), identification of the downstream targets of these transcription factors remains of considerable interest.

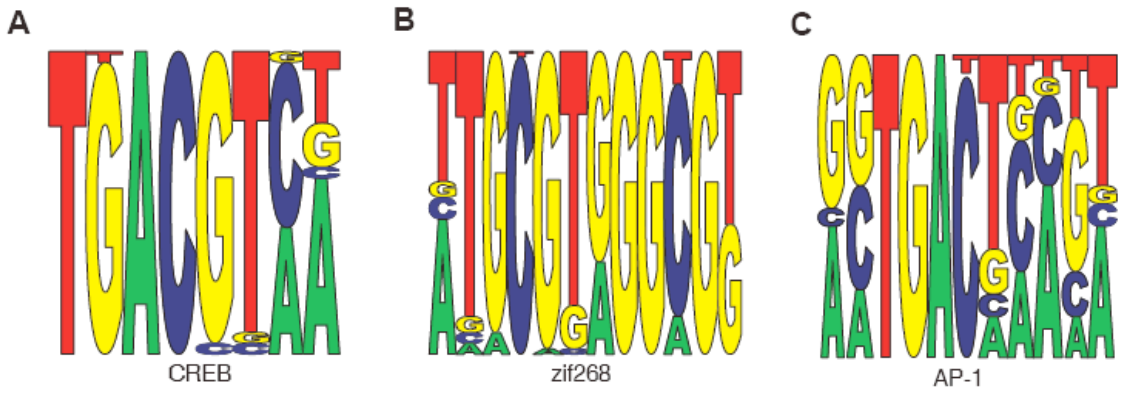

Figure 1. Activity dependent transcription factor binding sites consensus sequences. The height of the letters is proportional to their frequency in the data used to build the matrix. For instance, a large "A" in position one means that "A" belongs to the most probable consensus for that transcription factor binding site. (A) Consensus sequence given by the transfac CRE-binding matrix V$CREB_01. (B) Consensus sequence given by the transfac zif268 binding matrix V$EGR1_01. (C) Consensus sequence given by the transfac AP-1 binding matrix V$AP1_Q2.

We verified and our results using the location of transcription factor binding sites and improved target quality with comparative genomics. More general algorithms have successfully used comparative genomics between human and mouse to identify transcription factor binding sites (Elemento and Tavazoie 2005). We performed this search in parallel on all human and rat genes with an annotated transcription start sites. Dual hits from both the mouse and human genomes were considered more likely to have an be regulated by the transcription factor. Our results identify 854 candidate genes with conserved binding sites between human and mouse (see Table 1) that may be regulated by activity, 21 of which were predicted to have more than one type of transcription factor binding site. These results provide an important resource in understanding the regulatory networks that control activity-dependent programs of gene expression.

Table 1. Scores of computational search for activity-dependent transcription factor binding sites

	Mouse			Human			Homologue[a]		
Measure	Promoter[b]	Intergenic[c]	Score[d]	Promoter[b]	Intergenic[c]	Score[d]	Complete[e]	Conserved[f]	Score[g]
Total genes	18071	13475		19794	15178		13365	854	
CREB targets	1050	279	0.64	1389	388	0.64	830	356	0.83
zif268 targets	1382	200	0.81	1203	244	0.74	1136	166	0.56
AP-1 targets	4417	3612	-0.10	1623	1539	-0.24	3186	344	0.25
% CREB genes	5.81%	2.07%		7.02%	2.56%		6.21%	2.66%	
% zif268 genes	7.65%	1.48%		6.08%	1.61%		8.50%	1.24%	
% AP-1 genes	24.44%	26.81%		8.20%	10.14%		23.84%	2.57%	

[a] The homologue dataset was contructed using the Homologene resource on the NCBI website (http://www.ncbi.nih.gov).

[b] The promoter region is the area from -1,000 to 200 bp relative to transcription start.

[c] Intergenic regions refer to an area 50,000 bp away from transcription start. Fewer intergenic regions are available than genes because of gaps in sequencing.

[d] The score is determined by the positive predictive value which is a conservative estimate of the fraction of found transcription factor binding sites that are true hits. It is calculated as (%Observed - %Intergenic)/(%Observed).

[e] The complete homologue numbers refer the the number of human/mouse homologous pairs identified using the Homologene resources.

[f] The conserved dataset is a subset of the Homologene dataset that consists of only those genes for which homologous pairs contain a conserved binding site.

[g] The quality is determined by the positive predictive value which is a conservative estimate of the percentage of found transcription factor binding sites that are true hits. It is calculated as (%Conserved Observed - %Predicted Conserved)/(%Conserved Observed).

Methods

Promoter Database

We compiled a database of gene promoter regions using sequences from mouse build mm6 (Waterston et al. 2002), rat build rn3 (Gibbs et al. 2004), and human build hg17 (Lander et al. 2001; Venter et al. 2001) of the UCSC Genome Bioinformatics Resource (http://hgdownload.cse.ucsc.edu/downloads.html). Transcription start sites for these promoters organized by mRNA accession number were found in the table "knownGene.txt" for each build. Where promoter regions were reported within 50 bp of each other, only the one earlier on the chromosome was used, as the copies were presumed to be duplicates of the same promoter region (derived from otherwise identical mRNAs of different lengths). Incomplete promoters with missing sequence data were also removed from the analysis. The full promoter list was annotated with gene name, symbol, and accession number using the NCBI gene resources

(http://www.ncbi.nlm.nih.gov/) (Wheeler et al. 2005). In total, 18,071 mouse promoters, 19,794 human promoters, and 5,943 rat promoters were analyzed (Table 1).

When searching for candidate genes, we defined a putative promoter to be the genetic sequence from –1,000 bp to +200 bp of each transcription start for human, mouse, and rat genes. A set of intergenic sequences was also compiled for human and mouse to construct a "random" control dataset of 1,200 bp sequences using the regions from -51,200 bp to -50,000 bp relative to each transcription start site, where the transcription factors CREB, zif268, and AP-1 are not likely to have regulatory function. Due to a decrease in sequence quality further away from transcription start, distal sequence regions were available for only 77% of total genes, leaving 13,475 mouse intergenic regions and 15,178 human intergenic regions far analysis (Table 1). In order to confirm location-specificity trends inferred from the 1,200 bp regions, an additional search was run for each gene on an extended promoter region (-6,000 bp to +200 bp). We saw no significant difference in the region between -1,000 bp and -6,000 bp compared to the -51,200 bp to -50,000 bp region, suggesting that the initial search had identified the majority of sites with likely function.

The Homologene database provided us with human and mouse homologous pairs based on gene accession number (Wheeler et al. 2005), yielding 13,365 homologous gene pairs (Table 1). A binding site prediction was defined as conserved if the same binding site type was predicted in the promoters of both homologous genes, without regard for position in the promoter.

Transcription Factor Binding Site Inference

The goal of this search is to identify a transcription factor binding site compared to its background. This is the case when the probability that it is a binding site is greater than the probability you would observed the sequence.

$$P(Model) > P(Background)$$
$$\log(P(Model)) > \log(P(Background))$$
$$\log(P(Model)) - \log(P(Background)) > 0$$

Position specific scoring matrices (PSSMs) or position weight matrices (PWMs), are a well established method of motif finding (GuhaThakurta and Stormo 2001; Stormo 1990). We used a variant of them to find the log probability of a sequence being a part of the model. Our methods are similar to the transcription factor binding site search available through the database of transcription start sites (Suzuki et al. 2004). Binding site frequency matrices for AP-1, CREB, and zif268 were obtained from the Transfac (Wingender 2004; Wingender et al. 2000) public database (see Fig. 1). These frequency matrices give the frequency that each nucleotide is in each positive of the binding site. Scoring matrices for the present study were created from the Transfac frequency matrices with the following equation:

$$S_{n,p} = \log\left(\frac{A_{n,p} + b}{\sum_i A_{i,p} + 4b}\right)$$

where S is the scoring matrix, A is the frequency matrix, n is the nucleotide, p is the position within each binding site. The pseudocount, b, is set at the relatively small value of 0.25 to allow limited tolerance of base-pairs which have never been observed in a given position for a binding site. When comparing this scoring matrix to a sequence of

the same size, adding the scores for the nucleotide that is at that same position in the sequence gives you the log of probability that the sequence matches the model.

The probability that a sequence is part of the background, or not a binding, site is based on dinucleotide frequencies. For each species individually, we went through all promoter and calculated the probability of each dinucleotide transition. For instance, $P(AT) = \frac{Observed(AT)}{Observed(Nucleotides)}$. We also calculated the probability of observing each nucleotide individually. From those you can calculate the probability of observing any sequence by multiplying the probability of the first nucleotide by the probability of each nucleotide transition. The log probability can be found by adding the log of each probability.

$$P(ACT) = P(A) * \frac{P(AC)}{P(A)} * \frac{P(CT)}{P(C)}$$

$$\log(P(ACT)) = \log(P(A)) + \log(\frac{P(AC)}{P(A)}) + \log(\frac{P(CT)}{P(C)})$$

The goal of this study is to create a comprehensive list of possible transcription factor targets. In some studies, the log of the sequence length is subtracted to correct for the number of possible sites being searched. While subtracting by the full log of the sequence length would provide a more rigorous control, we deliberately chose to increase the sensitivity of the method at the expense of specificity at this stage in order to allow us to better take advantage of homology to enhance specificity in subsequent analysis. To analyze the quality of the data in response to decreasing the specificity, we use comparative genomics and a measure called the positive predicted value. Intuitively, it is the probability that an observed site is true positive. The positive predictive value is

defined as $\frac{TruePositives}{FalsePositives + TruePositives}$. In terms of our data, the positive predictive value is $\frac{ObservedSites - ExpectedSites}{ObservedSites}$. The expected number of sites is the number of sites expected to be conserved if there was no association between a binding site existing in mouse its human homologue. A series of possible correction terms are plotted against the positive predicted value (Figure 2). Because it is the point at which CREB and zif268 plateau, we chose to use a correction of three hundred. The final positive predictive value based on comparative genomics is found in Table 1 under homologues.

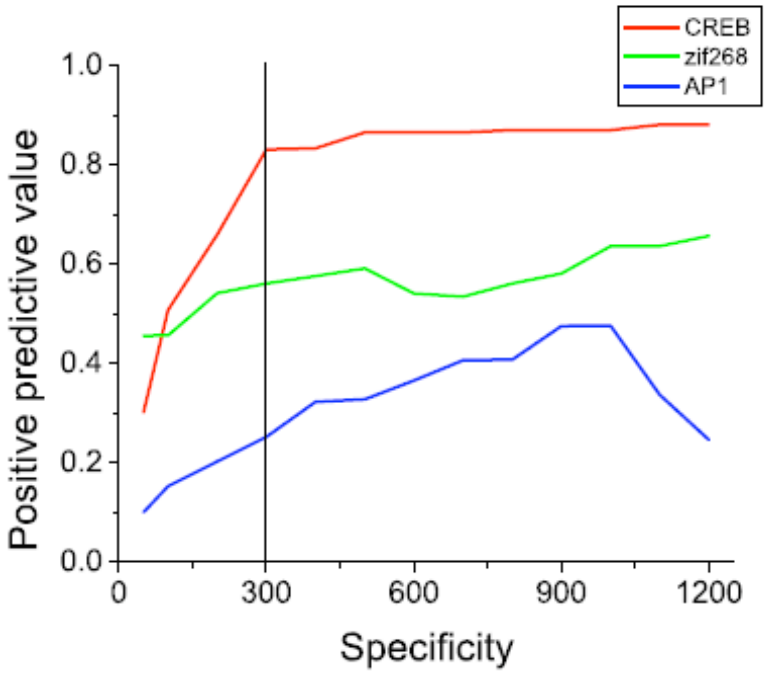

Figure 2. Fraction of binding sites expected to true positives. The specificity is a correction, the log of which is subtracted to arrive at the final score for a transcription factor binding site. The positive predicted value is calculated as shown in the methods. It represents the fraction of binding sites that are expected to be true positives based on a comparison between binding sites in human/mouse homologues. Values for CREB, zif268, and AP-1 are given.

A binding site is considered a hit if the final calculated score is above zero. The equation used to determine the final score is given below:

$$Score = \log(P(Model)) - \log(P(Background)) - \log(correction)$$

Global Data Analysis

The positive predictive values for the individual species is calculated by comparing the promoter region to the intergenic region (see Table 1). It is still calculated as $\frac{ObservedSites - ExpectedSites}{ObservedSites}$, but the observed sites is now the percentage of promoter targets with a binding site and the expected sites is the percentage of intergenic regions with that binding site.

Associations between co-occurring binding sites were analyzed by applying a 2-tailed Fisher's exact test (Agresti 1992), using a web-based calculator (http://www.matforsk.no/ola/fisher.htm), to the 2 by 2 contingency table of counts of occurrence of either, both, or neither site.

Analysis of the function of the activity-dependent transcription factor targets was done using GOstat (http://gostat.wehi.edu.au/), an online tool for finding overrepresented ontologies in a set of genes (Beissbarth and Speed 2004). The list of targets for each transcription factor binding site and species was searched against the entire list of promoters for overrepresentation of different gene ontology classes.

The rat microarray data used was obtained from the NCBI Gene Expression Omnibus (Barrett et al. 2005). The database describes how seizure was induced in rat at p15 using kainite. The mice were sacrificed at 1 hour, 6 hours, 24 hours, 72 hours, and 240 hours post-seizure and hippocampal tissue was examined. Three control and three experimental trials were done for each time point. Significance was calculated as a function of the database by comparing the intensity across trials (Barrett et al. 2005).

Results

Our research uses computationally advanced techniques to identify targets of activity-dependent transcription factors. Because these transcription factors have large roles in neural plasticity, we assume that many of them are members of the plasticity transcriptome. Our method may overestimate the transcriptome, but still has advantages over other methods that can miss important targets. We used a form of position specific scoring matrices or PSSMs to develop model for transcription factor binding sites. Using those models, we searched the promoter regions of human, mouse, and rat. Location specificity, or the tendency of valid binding sites to be close to transcription start was used to assess the accuracy of these models. Comparative genomics between human and mouse was used to find binding sites likely to be functionally conserved and important between the two species. Analysis of transcription factor targets shows that CREB and zif268, particularly if located on the same promoter, tend to regulate genes involved in regulatory network.

Developing binding site models

Several possible models exist for transcription factor binding sites. We wanted to avoid using a consensus sequence, which would be likely to miss many subtle targets. Sufficiently stringent criteria was also used to avoid the possibility of finding ten or more transcription factor binding site of one type in the same promoter, which is highly unlikely to occur naturally (James et al. 2005). Our binding site models were built from known examples of high quality, functional binding sites catalogued in the Transfac Database (Wingender et al. 2000). We also used frequency matrices that have been

experimentally developed for zif268 (Swirnoff and Milbrandt 1995; Wingender et al. 2000) and CREB (Benbrook and Jones 1994; Wingender et al. 2000). A schematic of the consensus sequences used is shown in Figure 1.

Another important aspect of this transcription factor binding site search is the correction for GC content. Mammalian promoters are known to have regions where the nucleotides G and C occur at a very high frequency. Certain pairs of nucleotides are also more likely to found within promoter regions. To distinguish the transcription factor binding sites from normal promoter sequence, it was necessary add these aspects of the promoter the background sequence model.

Computational genomic analysis: mouse, rat, and human genomes

Because binding sites are less likely to be similar in unrelated species, we analyzed the promoters of the mammals mouse, human, and rat. Although a comparative genomic analysis could have been conducted on rat, there were only 5,943 annotated promoters. The 18,071 mouse genes and 19,988 human genes allowed for a greater number of conserved targets (see Table 1). This represents all the annotated promoters from these species (approximately two-thirds of the total estimated distinct human and mouse genes; (Lander et al. 2001; Ota et al. 2004; Venter et al. 2001). Due to the amount and quality of data available, many of our analyses are restricted to mouse and human.

CREB binding sites were predicted in 6% of mouse promoters, 7% of human promoters, and 11% of rat promoters (Figure 3a). Zif268 binding sites were predicted in 8% of mouse promoters, 6% of human promoters and 5% of rat promoters (Figure 3b). AP-1 sites were the most common of the both the human and mouse datasets, with

predicted frequencies of approximately 24%, 8%, and 11% in mouse, human and rat promoters (Figure 3c). Although this number is high, it may be partly attributable to artificially high degeneracy of the AP-1 search consensus sequence, which stems in part from the diversity of AP-1 family members that can bind this sequence (Hai and Curran 1991). The large difference in AP-1 frequency could be due to differences in dinucleotide between human and mouse. An overcorrection for a high GC content in human promoters could easily have increased the frequency of the AT rich AP-1 transcription factor binding site.

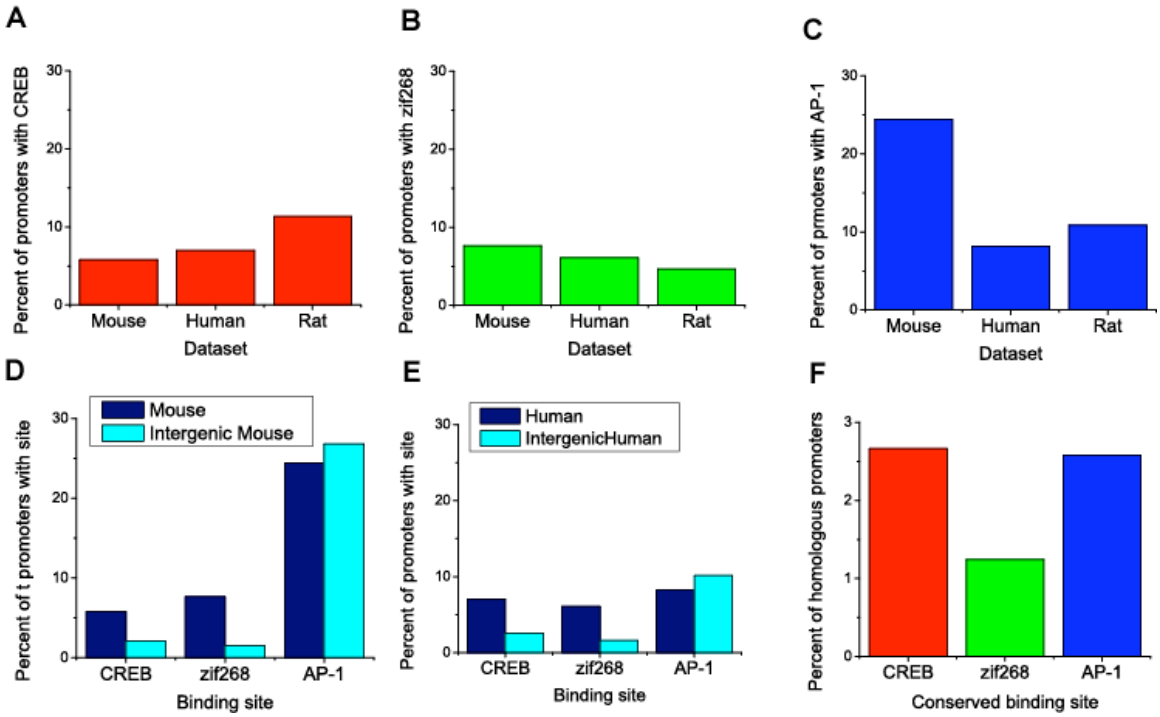

Figure 3. Relative frequency of transcription factor binding sites within promoter regions across species. (A) – (E) gives the number of percentage of promoters with a transcription factor binding site. For (F), the values shown are (# of promoters with a binding site x 100)/(# of human and mouse homologous pairs). Panels show (A) CREB, (B) zif268, (C) AP-1 frequency in promoters. A comparison of (D) mouse and (E) human binding site frequencies in promoter and intergenic regions is given. (F) shows the frequency of conserved activity-dependent transcription factor binding sites in the homolog dataset.

Estimation of search quality

To develop a rigorous estimate of the quality of our hits, calculated the positive

predictive value. This was accomplished by examining the location and conservation of binding sites. This is a conservative estimate of the percentage of identified transcription factor binding sites that we expect to be functional binding sites. In individual species, we compared predicted transcription factor binding site frequencies within gene promoters to those for control sequences selected from intergenic regions (see Methods). Our analysis showed that CREB and zif268 binding sites were more likely to be found in our 1,200 bp promoter region (Figure 3d and 3e). A more detailed analysis shows that the frequency of these binding site increases relative to how close to transcription start site you are (Figure 4a and 4b). This translated to a positive predictive value of for CREB 0.64 in human and mouse. Zif268 had a slightly higher positive predictive value of 0.81 in mouse and 0.74 in human (Table 1).

Comparative genomics was also used a method of obtaining the positive predictive value. To identify the homologous mouse/human pairs, we used the Homologene database (Homologene; Wheeler et al. 2005). This dataset was found to have a similar frequency of transcription factor binding sites (Table 1). We then examined the conservation of each type of transcription factor binding for each mouse/human pair. In the set of mouse-human homologues, 2.7% have a conserved CREB site, 1.2% have a conserved zif268 site, and 2.6% of promoters have a conserved AP-1 site (Figure 3f). These binding sites that existed in a mouse-human homologue pair were considered to be true positives. The positive predictive value was calculated by comparing this number to the number of binding sites one would expect to find if there was no evolutionary pressure for site conservation. Unlike using the location specificity, the positive predictive value was higher for CREB than zif268. We predicted it as 0.83 for CREB

and 0.56 in zif268. This difference is most likely to due change in zif268 targets between species.

Location specificity within the promoter

Based on the observation that more binding sites were found the promoter regions for CREB and zif268, we conducted a more complete analysis of binding site location (Figure 4). The frequency of CREB (Figure 4a) and zif268 (Figure 4b) binding sites were observed to increase close to transcription start, especially the –100 bp region. In the –900 to –600 bp region, the frequency was no different from the intergenic regions. The same location specificity exists in the conserved dataset (Fig. 4d and 4e). The high quality of the conserved dataset for CREB sites is supported by an observable increase in the peak at –100 bp as opposed –900 bp.

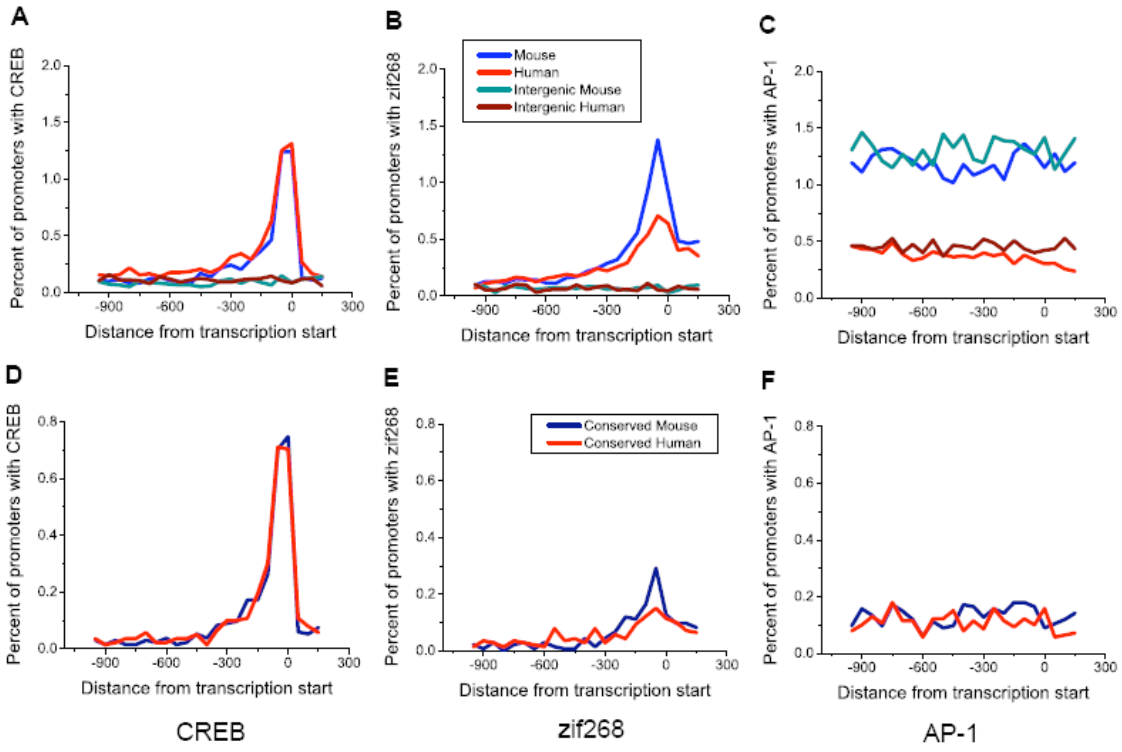

Figure 4. Histograms of CREB and zif268 binding are grouped by position relative to transcription start, showing pronounced location specificity within the promoter. All histograms were created using a bin size of 50 bp. The total number of binding sites in each 50 bp region was divided by the total number of promoters for that dataset. Shown are both human and mouse and promoters. For the intergenic dataset, the area shown is from -51,200 bp to -50,000 bp relative to transcription start. The binding sites analyzed are (A) CREB, (B) zif268, and (C) AP-1. Histograms for the conserved human and mouse datasets are given in for (D) CREB, (E) zif268, and (F) AP-1 transcription factor binding sites. The percentage of promoters with a binding site is calculated relative to the total number of homologous pairs. The mouse dataset is the location of mouse binding sites when the homologous gene also has a binding site of the same type. The human dataset is the location of human binding sites when the homologous gene also has a binding site of the same type.

Unlike CREB and AP-1, the location specificity of zif268 was different between mouse and human. The most likely explanation is the higher GC content in human. Certain GC rich sequences that would be recognized in mouse would be recognized in humans. Both transcription factor binding sites and GC rich regions are likely to be close to transcription start (Vinogradov 2005). It remains a mystery whether the decrease in the strength of the peak in human is a significant reduction of false positives or true positives. In the human, mouse, and conserved datasets, AP-1 did not have any location specificity (Fig 4c and 4f). One hypothesis is that there are more false positives for AP-1 that overwhelm the true hits. This could be due to a poor data that the binding site model was constructed from. Another explanation is that it is not optimal for AP-1 to be located within a specific region of the promoter.

Overlap of CREB, zif268, and AP-1 target genes

Co-regulation of promoters, or two transcription factor binding sites regulating the same genes, is a relatively frequent occurrence. However, it has not been studied with regards to activity-dependent gene regulation and these transcription factors. If co-regulation does occur, you would expect the two transcription factor binding sites to be often found in the same promoter region. The amount of overlap between zif268 and

CREB targets was greater than would be expected by chance (Table 2, highlighted in yellow), suggesting that at least a subset of target genes may be regulated by both CREB and zif268.

Table 3. Significance of CREB and zif268 binding site colocalization

	CREB and zif268	CREB and AP-1	zif268 and AP-1
Mouse	0.023[a] [b]	0.236	0.013
	Positive[c]	Negative	Negative
Human	0.180	0.141	0.007
	Positive	Negative	Negative
Rat	<.001	0.538	0.708
	Positive	Positive	None

[a] The p-value was calculated using Fisher's exact test (see methods).

[b] The positve association between CREB and zif268 is highlighted for mouse, human, and rat.

[c] A positive direction of the association means that the two binding sites are are more likely to be found on the same promoter than if they were randomly distributed.

We found a significant negative association between the transcription factors zif268 and AP-1. This is particularly peculiar because of AP-1 low positive predictive value and estimated binding site quality. The most straightforward explanation is that zif268 and AP-1 are purposely activating distinct subsets of genes. Another hypothesis is that zif268, as GC rich binding site, is more likely to be found in GC rich promoters, while AP-1, an AT rich binding site, is more likely to be found in AT rich promoters.

Analysis of specific target genes: CREB and zif268 targets

The analysis of the CREB/zif268 overlap led to identify the targets of this possible co-regulation. To leave only the highest quality targets, we decided to include on those genes with both a conserved CREB and zif268 binding site between mouse and human. FosB (Figure 5A), Jund1 (Figure 5B), and Maff (Figure 5C) are all members of

the AP-1 family of transcription factors. The Skil (Figure 5D) transcription factor is a member of the SKI/SNO/DAC family which are known in some cases to attach to the same protein complex as AP-1 (Xu et al. 2000). The observation that a specific group of transcription factors can be regulated by both CREB and zif268 implicate these genes in transcriptional networks of activity-regulated gene expression. Neuronal pentraxin 1 is an interesting in neural-specific target. It is known to be expressed in response BDNF, a chemical stimulant of neurons important in plasticity (Ring et al. 2006), as well as brain injury (Hossain et al. 2004).

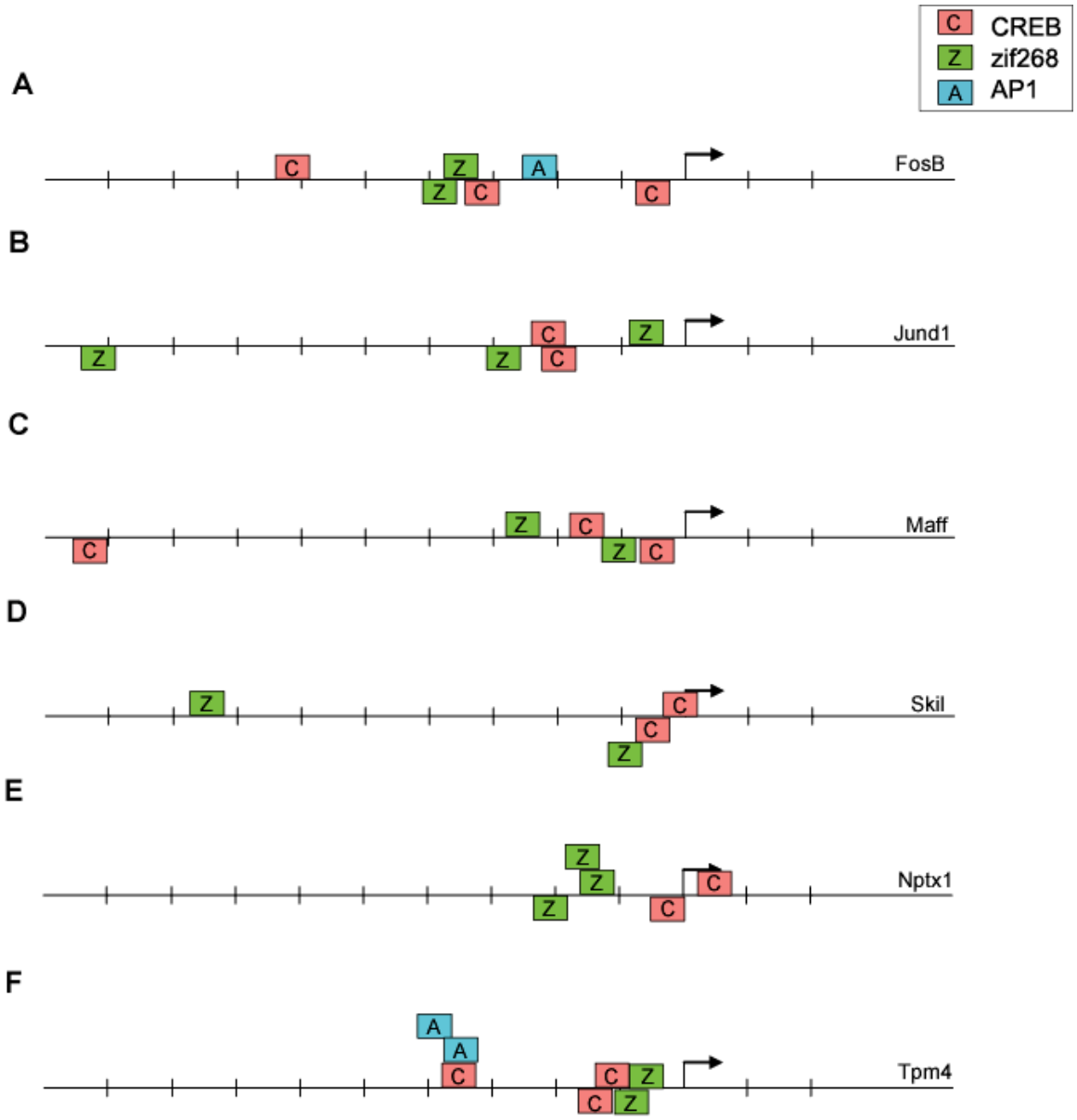

Figure 5. Binding site location in CREB/zif268 double hits. Promoter regions for the genes on the line to the left are from -1000 bp to 200 bp relative to transcription start, which is denoted by the arrow. The red blocks above represent CREB binding sites, the green blocks represent zif268 binding sites, and the blue blocks represent AP-1 binding sites. Mouse transcription factor binding sites are found on top of the line while human sites are below. A non-allignment method was used to identify promoter regions, so homologous binding sites might not be at the same location. Gene symbols shown are those for mouse. Promoter regions shown are for (A) *FBJ osteosarcoma oncogene B/FosB*, (B) *Jun proto-oncogene related gene d1* (C) *v-maf musculoaponeurotic fibrosarcoma oncogene family, protein F (avian)*, (D) *SKI-like*, (E) *neuronal pentraxin 1*, and (F) *tropomyosin 4*.

Because our analysis of this co-regulation was encouraging, we attempted to develop broader roles of CREB and zif268 by examining the function of their targets. The GoSTAT web resource allowed us to look for functions in a set of genes that are

over and under-represented (Beissbarth and Speed 2004). This program was used to compare the set of CREB, zif268, and AP-1 targets to the set of all genes studied in a species (Table 3). Interestingly, genes with CREB consensus sites showed significant overrepresentation for targets involved in transcription and RNA processing, but underrepresented in more obvious important neural targets, receptors and channels. However, it could be the case that the few targets predicted are functional important. RNA processing genes, such as those the facilitate RNA splicing, have important effects of neuron-neuron communication (Zhong et al. 2006). Along with CREB, zif268 targets were strongly enriched for transcription factors, but only in mouse. This lack of overrepresented function in human may correspond to the location specific decrease in the peak close to transcription start for zif268. The overrepresentation of transcription factors for both binding sites is consistent with the prior analysis of the putative high-quality hits generated by searching for CREB/zif268 co-regulated genes among the Homologene set. To further support this hypothesis, the CREB/zif268 targets in human and mouse were analyzed for over and underrepresented functions. Transcription factor binding sites were overrepresented in both species. Only one class, immune response, was significantly enriched for AP-1, a fact that underscores the significance of the observations for CREB and zif268.

Table 4. Over and underrepresented functions in transcription factor binding site candidates

Human	p-value[a]	Mouse	p-value[a]
Overrepresented Category [b]			
CREB			
Nucleus	5.97E-06	Nucleus	8.62E-07
Transcription	0.000966	Dephosphorylation	0.0014
Primary Metabolism	0.00158	Biopolymer Metabolism	0.00312
RNA Processing	0.00229	Nucleotide Binding	0.0111
Regulation of Physiological Process	0.00274	RNA Processing	0.0236
Mitosis	0.0114	Protein Dimerization Activity	0.0309
Microtubuole Cytoskeleton	0.0143	Cell Cycle	0.0495
RNA Splicing	0.0265		
Phosphate Metabolism	0.0492		
zif268			
None		Nucleus	0.00117
		Zinc Ion Binding	0.00117
		Transcription, DNA Dependent	0.004
		Leading Edge	0.00926
		Cellular Physiological Process	0.0139
		Cellular Metabolism	0.019
		Cell Development	0.0425
AP-1			
Immune Response	0.0132	None	
CREB/zif268			
Protein Dimerization Activity	0.00235	Protein Dimerization Activity	0.0134
Transcription Factor Binding	0.00677	Membrane-bound Organelle	0.0134
Transcription Corepressor Activity	0.0122	Regulation of Epidermis Development	0.0134
Activation of MAPKK Activity	0.0122	Protein Serine/Threonine Phosphotase	0.0144
		Regulation of Transcription	0.0433
Underrepresented Category [b]			
CREB			
Intrinsic to Membrane	8.10E-07	Organismal Physiological Process	0.000289
Receptor Activity	0.00158	Intrinsic to Membrane	0.0083
Ion Transporter Activity	0.00821	Neurophysiological Process	0.0083
Sensory Perception	0.0469	Transmembrane Receptor Activity	0.0086
		Sensory Perception	0.0129
zif268			
None		Olfactory Receptor Activity	0.00117
		G-protein Coupled Receptor Activity	0.00277
		Organismal Physiological Process	0.0132
		Transmembrane Receptor Activity	0.019
		Neurophysiological Process	0.0275
		Defense Response	0.0481
AP-1			
None		None	
CREB/zif268			
None		None	

[a] This shows p-values for all gene ontologies signaficantly over and under-represented using the goSTAT application at (http://gostat.wehi.edu.au/) (Biessbarth and Speed).

[b] Similar categories are listed together. More information about the gene ontology catagories can be found at (http://www.informatics.jax.org/searches/GO_form.shtml).

Experimental Validation of Targets

Experimental validation of the transcription factor binding site targets is difficult because our method was developed in response to a perceived weakness the current

experimental methods. We believe that the changes in gene expression measured by microarrays are mostly due to networks of transcription. However, one would expect to observe an increase (or decrease) in the predicted CREB and zif268 targets in response to increased activity. To test this hypothesis, we looked at the percentage of targets in rat genes differentially expressed after seizure compared with the entire rat promoter dataset (Figure 6). No binding site is extremely overrepresented in the rat seizure data. The CREB and zif268 targets both increase until the 240 hour time point. This could be noise, but it could also indicate a binding site specific in crease mRNA post-seizure.

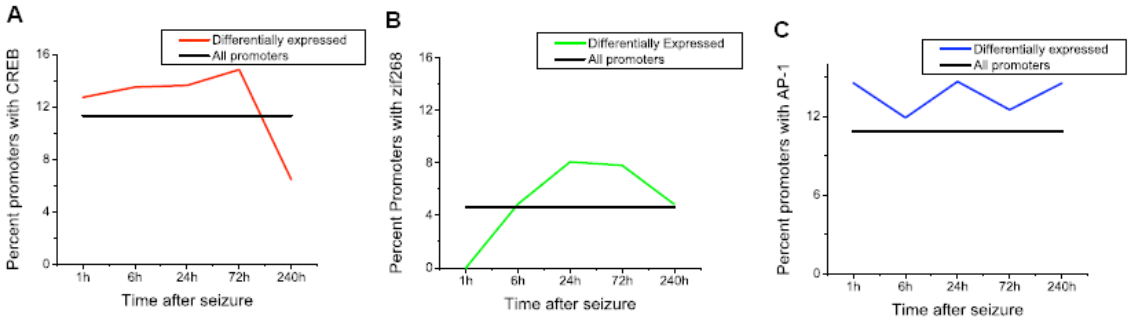

Figure 6. Binding site frequencies in genes responding to seizure. The rat microarray data used was obtained from the NCBI Gene Expression Omnibus. Seizure was induced in rat at p15 using kainite. The mice were sacrificed at 1 hour, 6 hours, 24 hours, 72 hours, and 240 hours post-seizure. Hippocampal tissue was examined for three control and three experimental trials for each time point. Significance was calculated as a function of the database by comparing the intensity across trials (Barrett et al. 2005). The differentially expressed genes in include those that are downregulated. These promoters of the differentially expressed genes were scanned for (A) CREB, (B) zif268, and (C) AP-1 transcription factor binding sites. The dark line shows the frequency of the transcription factor binding site in all rat promoters

Discussion

We have carried out our computational approach to identifying the plasticity transcriptome by identifying the targets of the transcription factors CREB, zif268, and AP-1. Location of the binding sites was used to get a measure of binding site quality. The comparative genomics approach gave an independent measure of quality as well as a method of decreasing the frequency of false positives. Further verification of transcription factor binding sites was carried out in rat by comparing gene activity to our

targets list. This study is the first computationally rigorous sequence-based analysis of genes that are likely to be regulated by neural activity. Our methods are not dependent on unreliable tissue collection or limited to a handful of testable targets which may bias most studies of the transcriptome. It has the added advantage of being easy to run and cost-effective. This analysis predicts CREB and zif268 are targeting another level of regulation and, in general, not genes important to plasticity.

New methods for evaluating target quality

It is not surprising that the transcription factor binding sites are most likely found in promoter regions. This location specificity has been anecdotally noted in previous work (Elemento and Tavazoie 2005). The incredibly strong location specificity of CREB and zif268 was unexpected. A huge difference between the number of sites found at –100bp compared to –600bp suggests that there is evolutionary pressure for a binding to be located very close to transcription start. It follows that these sites are more likely to be functional than sites found farther away. Perhaps better annotated transcription starts and perfect binding sites models could reveal a distinct window in which a transcription factor must operate. Another possibility is that there is a peak region for binding site function surrounded by sites that only have a small influence on the gene.

Transcriptional networks triggered by neural activity

Although we understood that CREB and zif268 have a role in activity-dependent gene expression, we were not sure what that role was. This study developed two related functions for CREB and zif268 in relation to the plasticity transcriptome. First, guided

by their co-occurrence in the same promoter, they seem to be co-regulating member of the AP-1 family of transcription factors. Second, on a broader scale, they seem to be targeting transcription factors and other gene involved in regulatory networks. This knowledge could direct biologists to study other transcription factors as targets of CREB and zif268 instead of the more interesting important neural candidates that are singled out when using the individual gene approach to identifying the plasticity transcriptome.

Future improvements

Our success in validating transcription factor binding site targets by using both location specificity and comparatives genomics leaves many avenues of research open. Homology methods for improving specificity of binding site prediction (Elemento and Tavazoie 2005) have already been explored. However, to our knowledge, location specificity of binding sites has not been formally used in motif discovery. More research could also be conducted on binding site co-occurrence (Bulyk et al. 2004; Luscombe et al. 2004). Our analysis currently uses location specificity, comparative genomics, and response to seizure independently to verify targets. A Bayesian model would be able to take each of these factors into account and improve the search for binding sites.

Conclusion

This research is an important step in the use of sequence based methods to study the plasticity transcriptome. Our initial goal was to simply provide a list of possible candidates. Through our attempts to rigorously validate these candidates, we were able to discover interesting properties such as location specificity, binding site associations,

and possible networks of gene regulation. Each of these factors suggests that activity-dependent gene expression requires a very complex model to understand. The weakness of this sequence-based method, along with the current experimental methods, is that it is unable to understand the intricate system that the transcriptome is a part of. The strength of the sequenced-based methods is it able to discover where in the system the complexity lies in an unbiased way.

Acknowledgements

This work was supported by Merck Computational Biology and Chemistry Summer Program Fellowship (ARP), the Carnegie Mellon Undergraduate Research Initiative (ARP), the HHMI Summer Undergraduate Research Program (ARP), the Alfred P. Sloan Foundation (ALB), and U.S. National Science Foundation Award DBI-0346981 (RS). We thank Christopher Burge for helpful comments on this work.

References

Agresti, A. 1992. A survery of exact inference for contingency tables. *Statistical Science* 7: 131-177.

Barrett, T., T.O. Suzek, D.B. Troup, S.E. Wilhite, W.C. Ngau, P. Ledoux, D. Rudnev, A.E. Lash, W. Fujibuchi, and R. Edgar. 2005. NCBI GEO: mining millions of expression profiles--database and tools. *Nucleic Acids Res* 33: D562-566.

Befort, K., L. Karchewski, C. Lanoue, and C.J. Woolf. 2003. Selective up-regulation of the growth arrest DNA damage-inducible gene Gadd45 alpha in sensory and motor neurons after peripheral nerve injury. *Eur J Neurosci* 18: 911-922.

Beissbarth, T. and T.P. Speed. 2004. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* 20: 1464-1465.

Benbrook, D.M. and N.C. Jones. 1994. Different binding specificities and transactivation of variant CRE's by CREB complexes. *Nucleic Acids Res* 22: 1463-1469.

Bulyk, M.L. 2003. Computational prediction of transcription-factor binding site locations. *Genome Biol* 5: 201.

Bulyk, M.L., A.M. McGuire, N. Masuda, and G.M. Church. 2004. A motif co-occurrence approach for genome-wide prediction of transcription-factor-binding sites in Escherichia coli. *Genome Res* 14: 201-208.

Cha-Molstad, H., D.M. Keller, G.S. Yochum, S. Impey, and R.H. Goodman. 2004. Cell-type-specific binding of the transcription factor CREB to the cAMP-response element. *Proc Natl Acad Sci U S A* 101: 13572-13577.

Corriveau, R.A., G.S. Huh, and C.J. Shatz. 1998. Regulation of class I MHC gene expression in the developing and mature CNS by neural activity. *Neuron* 21: 505-520.

Costigan, M., K. Befort, L. Karchewski, R.S. Griffin, D. D'Urso, A. Allchorne, J. Sitarski, J.W. Mannion, R.E. Pratt, and C.J. Woolf. 2002. Replicate high-density rat genome oligonucleotide microarrays reveal hundreds of regulated genes in the dorsal root ganglion after peripheral nerve injury. *BMC Neurosci* 3: 16.

Del Rio, J.A. and C. Barlow. 2002. Genomics and neurological phenotypes: applications for seizure-induced damage. *Prog Brain Res* 135: 149-160.

Elemento, O. and S. Tavazoie. 2005. Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach. *Genome Biol* 6: R18.

Elliott, R.C. and D.H. Lowenstein. 2004. Gene expression profiling of seizure disorders. *Neurochem Res* 29: 1083-1092.

Flood, W.D., R.W. Moyer, A. Tsykin, G.R. Sutherland, and S.A. Koblar. 2004. Nxf and Fbxo33: novel seizure-responsive genes in mice. *Eur J Neurosci* 20: 1819-1826.

Gibbs, R.A. G.M. Weinstock M.L. Metzker D.M. Muzny E.J. Sodergren S. Scherer G. Scott D. Steffen K.C. Worley P.E. Burch G. Okwuonu S. Hines L. Lewis C. DeRamo O. Delgado S. Dugan-Rocha G. Miner M. Morgan A. Hawes R. Gill Celera R.A. Holt M.D. Adams P.G. Amanatides H. Baden-Tillson M. Barnstead S. Chin C.A. Evans S. Ferriera C. Fosler A. Glodek Z. Gu D. Jennings C.L. Kraft T. Nguyen C.M. Pfannkoch C. Sitter G.G. Sutton J.C. Venter T. Woodage D. Smith H.M. Lee E. Gustafson P. Cahill A. Kana L. Doucette-Stamm K. Weinstock K. Fechtel R.B. Weiss D.M. Dunn E.D. Green R.W. Blakesley G.G. Bouffard P.J. De Jong K. Osoegawa B. Zhu M. Marra J. Schein I. Bosdet C. Fjell S. Jones M. Krzywinski C. Mathewson A. Siddiqui N. Wye J. McPherson S. Zhao C.M. Fraser J. Shetty S. Shatsman K. Geer Y. Chen S. Abramzon W.C. Nierman P.H. Havlak R. Chen K.J. Durbin A. Egan Y. Ren X.Z. Song B. Li Y. Liu X. Qin S. Cawley K.C. Worley A.J. Cooney L.M. D'Souza K. Martin J.Q. Wu M.L. Gonzalez-Garay A.R. Jackson K.J. Kalafus M.P. McLeod A. Milosavljevic D. Virk A. Volkov D.A. Wheeler Z. Zhang J.A. Bailey E.E. Eichler E. Tuzun E. Birney E. Mongin A. Ureta-Vidal C. Woodwark E. Zdobnov P. Bork M. Suyama D. Torrents M. Alexandersson B.J. Trask J.M. Young H. Huang H. Wang H. Xing S. Daniels D. Gietzen J. Schmidt K. Stevens U. Vitt J. Wingrove F. Camara M. Mar Alba J.F. Abril R. Guigo A. Smit I. Dubchak E.M. Rubin O. Couronne A. Poliakov N. Hubner D. Ganten C. Goesele O. Hummel T. Kreitler Y.A. Lee J. Monti H. Schulz H. Zimdahl H. Himmelbauer H. Lehrach H.J. Jacob S.

Bromberg J. Gullings-Handley M.I. Jensen-Seaman A.E. Kwitek J. Lazar D. Pasko P.J. Tonellato S. Twigger C.P. Ponting J.M. Duarte S. Rice L. Goodstadt S.A. Beatson R.D. Emes E.E. Winter C. Webber P. Brandt G. Nyakatura M. Adetobi F. Chiaromonte L. Elnitski P. Eswara R.C. Hardison M. Hou D. Kolbe K. Makova W. Miller A. Nekrutenko C. Riemer S. Schwartz J. Taylor S. Yang Y. Zhang K. Lindpaintner T.D. Andrews M. Caccamo M. Clamp L. Clarke V. Curwen R. Durbin E. Eyras S.M. Searle G.M. Cooper S. Batzoglou M. Brudno A. Sidow E.A. Stone J.C. Venter B.A. Payseur G. Bourque C. Lopez-Otin X.S. Puente K. Chakrabarti S. Chatterji C. Dewey L. Pachter N. Bray V.B. Yap A. Caspi G. Tesler P.A. Pevzner D. Haussler K.M. Roskin R. Baertsch H. Clawson T.S. Furey A.S. Hinrichs D. Karolchik W.J. Kent K.R. Rosenbloom H. Trumbower M. Weirauch D.N. Cooper P.D. Stenson B. Ma M. Brent M. Arumugam D. Shteynberg R.R. Copley M.S. Taylor H. Riethman U. Mudunuri J. Peterson M. Guyer A. Felsenfeld S. Old S. Mockrin and F. Collins. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428: 493-521.

Guan, Z., S. Saraswati, B. Adolfsen, and J.T. Littleton. 2005. Genome-wide transcriptional changes associated with enhanced activity in the Drosophila nervous system. *Neuron* 48: 91-107.

GuhaThakurta, D. and G.D. Stormo. 2001. Identifying target sites for cooperatively binding factors. *Bioinformatics* 17: 608-621.

Hai, T. and T. Curran. 1991. Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity. *Proc Natl Acad Sci U S A* 88: 3720-3724.

Hakimi, M.A., D.A. Bochar, J. Chenoweth, W.S. Lane, G. Mandel, and R. Shiekhattar. 2002. A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc Natl Acad Sci U S A* 99: 7420-7425.

Herdegen, T. and J.D. Leah. 1998. Inducible and constitutive transcription factors in the mammalian nervous system: control of gene expression by Jun, Fos and Krox, and CREB/ATF proteins. *Brain Res Brain Res Rev* 28: 370-490.

Hong, E.J., A.E. West, and M.E. Greenberg. 2005. Transcriptional control of cognitive development. *Curr Opin Neurobiol* 15: 21-28.

Hossain, M.A., J.C. Russell, R. O'Brien, and J. Laterra. 2004. Neuronal pentraxin 1: a novel mediator of hypoxic-ischemic injury in neonatal brain. *J Neurosci* 24: 4187-4196.

Hunsberger, J.G., A.H. Bennett, E. Selvanayagam, R.S. Duman, and S.S. Newton. 2005. Gene profiling the response to kainic acid induced seizures. *Brain Res Mol Brain Res* 141: 95-112.

Impey, S., S.R. McCorkle, H. Cha-Molstad, J.M. Dwyer, G.S. Yochum, J.M. Boss, S. McWeeney, J.J. Dunn, G. Mandel, and R.H. Goodman. 2004. Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* 119: 1041-1054.

Israsena, N., M. Hu, W. Fu, L. Kan, and J.A. Kessler. 2004. The presence of FGF2 signaling determines whether beta-catenin exerts effects on proliferation or neuronal differentiation of neural stem cells. *Dev Biol* 268: 220-231.

James, A.B., A.M. Conway, and B.J. Morris. 2005. Genomic profiling of the neuronal target genes of the plasticity-related transcription factor -- Zif268. *J Neurochem* 95: 796-810.

Laifenfeld, D., E. Klein, and D. Ben-Shachar. 2002. Norepinephrine alters the expression of genes involved in neuronal sprouting and differentiation: relevance for major depression and antidepressant mechanisms. *J Neurochem* 83: 1054-1064.

Lander, E.S. L.M. Linton B. Birren C. Nusbaum M.C. Zody J. Baldwin K. Devon K. Dewar M. Doyle W. FitzHugh R. Funke D. Gage K. Harris A. Heaford J. Howland L. Kann J. Lehoczky R. LeVine P. McEwan K. McKernan J. Meldrim J.P. Mesirov C. Miranda W. Morris J. Naylor C. Raymond M. Rosetti R. Santos A. Sheridan C. Sougnez N. Stange-Thomann N. Stojanovic A. Subramanian D. Wyman J. Rogers J. Sulston R. Ainscough S. Beck D. Bentley J. Burton C. Clee N. Carter A. Coulson R. Deadman P. Deloukas A. Dunham I. Dunham R. Durbin L. French D. Grafham S. Gregory T. Hubbard S. Humphray A. Hunt M. Jones C. Lloyd A. McMurray L. Matthews S. Mercer S. Milne J.C. Mullikin A. Mungall R. Plumb M. Ross R. Shownkeen S. Sims R.H. Waterston R.K. Wilson L.W. Hillier J.D. McPherson M.A. Marra E.R. Mardis L.A. Fulton A.T. Chinwalla K.H. Pepin W.R. Gish S.L. Chissoe M.C. Wendl K.D. Delehaunty T.L. Miner A. Delehaunty J.B. Kramer L.L. Cook R.S. Fulton D.L. Johnson P.J. Minx S.W. Clifton T. Hawkins E. Branscomb P. Predki P. Richardson S. Wenning T. Slezak N. Doggett J.F. Cheng A. Olsen S. Lucas C. Elkin E. Uberbacher M. Frazier R.A. Gibbs D.M. Muzny S.E. Scherer J.B. Bouck E.J. Sodergren K.C. Worley C.M. Rives J.H. Gorrell M.L. Metzker S.L. Naylor R.S. Kucherlapati D.L. Nelson G.M. Weinstock Y. Sakaki A. Fujiyama M. Hattori T. Yada A. Toyoda T. Itoh C. Kawagoe H. Watanabe Y. Totoki T. Taylor J. Weissenbach R. Heilig W. Saurin F. Artiguenave P. Brottier T. Bruls E. Pelletier C. Robert P. Wincker D.R. Smith L. Doucette-Stamm M. Rubenfield K. Weinstock H.M. Lee J. Dubois A. Rosenthal M. Platzer G. Nyakatura S. Taudien A. Rump H. Yang J. Yu J. Wang G. Huang J. Gu L. Hood L. Rowen A. Madan S. Qin R.W. Davis N.A. Federspiel A.P. Abola M.J. Proctor R.M. Myers J. Schmutz M. Dickson J. Grimwood D.R. Cox M.V. Olson R. Kaul C. Raymond N. Shimizu K. Kawasaki S. Minoshima G.A. Evans M. Athanasiou R. Schultz B.A. Roe F. Chen H. Pan J. Ramser H. Lehrach R. Reinhardt W.R. McCombie M. de la Bastide N. Dedhia H. Blocker K. Hornischer G. Nordsiek R. Agarwala L. Aravind J.A. Bailey A. Bateman S. Batzoglou E. Birney P. Bork D.G. Brown C.B. Burge L. Cerutti H.C. Chen D. Church M. Clamp R.R. Copley T. Doerks S.R. Eddy E.E. Eichler T.S. Furey J. Galagan J.G. Gilbert C. Harmon Y. Hayashizaki D. Haussler H. Hermjakob K. Hokamp W. Jang L.S. Johnson T.A. Jones S. Kasif A. Kaspryzk S. Kennedy W.J. Kent P. Kitts E.V. Koonin I. Korf D. Kulp D. Lancet T.M. Lowe A. McLysaght T. Mikkelsen J.V. Moran N.

Mulder V.J. Pollara C.P. Ponting G. Schuler J. Schultz G. Slater A.F. Smit E. Stupka J. Szustakowski D. Thierry-Mieg J. Thierry-Mieg L. Wagner J. Wallis R. Wheeler A. Williams Y.I. Wolf K.H. Wolfe S.P. Yang R.F. Yeh F. Collins M.S. Guyer J. Peterson A. Felsenfeld K.A. Wetterstrand A. Patrinos M.J. Morgan P. de Jong J.J. Catanese K. Osoegawa H. Shizuya S. Choi and Y.J. Chen. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Lee, K.H., C.J. Ryu, H.J. Hong, J. Kim, and E.H. Lee. 2005. CDNA microarray analysis of nerve growth factor-regulated gene expression profile in rat PC12 cells. *Neurochem Res* 30: 533-540.

Lukasiuk, K. and A. Pitkanen. 2004. Large-scale analysis of gene expression in epilepsy research: is synthesis already possible? *Neurochem Res* 29: 1169-1178.

Luo, Y., J.M. Long, E.L. Spangler, D.L. Longo, D.K. Ingram, and N.P. Weng. 2001. Identification of maze learning-associated genes in rat hippocampus by cDNA microarray. *J Mol Neurosci* 17: 397-404.

Luscombe, N.M., M.M. Babu, H. Yu, M. Snyder, S.A. Teichmann, and M. Gerstein. 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 431: 308-312.

Nedivi, E., D. Hevroni, D. Naot, D. Israeli, and Y. Citri. 1993. Numerous candidate plasticity-related genes revealed by differential cDNA cloning. *Nature* 363: 718-722.

Ota, T. Y. Suzuki T. Nishikawa T. Otsuki T. Sugiyama R. Irie A. Wakamatsu K. Hayashi H. Sato K. Nagai K. Kimura H. Makita M. Sekine M. Obayashi T. Nishi T. Shibahara T. Tanaka S. Ishii J. Yamamoto K. Saito Y. Kawai Y. Isono Y. Nakamura K. Nagahari K. Murakami T. Yasuda T. Iwayanagi M. Wagatsuma A. Shiratori H. Sudo T. Hosoiri Y. Kaku H. Kodaira H. Kondo M. Sugawara M. Takahashi K. Kanda T. Yokoi T. Furuya E. Kikkawa Y. Omura K. Abe K. Kamihara N. Katsuta K. Sato M. Tanikawa M. Yamazaki K. Ninomiya T. Ishibashi H. Yamashita K. Murakawa K. Fujimori H. Tanai M. Kimata M. Watanabe S. Hiraoka Y. Chiba S. Ishida Y. Ono S. Takiguchi S. Watanabe M. Yosida T. Hotuta J. Kusano K. Kanehori A. Takahashi-Fujii H. Hara T.O. Tanase Y. Nomura S. Togiya F. Komai R. Hara K. Takeuchi M. Arita N. Imose K. Musashino H. Yuuki A. Oshima N. Sasaki S. Aotsuka Y. Yoshikawa H. Matsunawa T. Ichihara N. Shiohata S. Sano S. Moriya H. Momiyama N. Satoh S. Takami Y. Terashima O. Suzuki S. Nakagawa A. Senoh H. Mizoguchi Y. Goto F. Shimizu H. Wakebe H. Hishigaki T. Watanabe A. Sugiyama M. Takemoto B. Kawakami M. Yamazaki K. Watanabe A. Kumagai S. Itakura Y. Fukuzumi Y. Fujimori M. Komiyama H. Tashiro A. Tanigami T. Fujiwara T. Ono K. Yamada Y. Fujii K. Ozaki M. Hirao Y. Ohmori A. Kawabata T. Hikiji N. Kobatake H. Inagaki Y. Ikema S. Okamoto R. Okitani T. Kawakami S. Noguchi T. Itoh K. Shigeta T. Senba K. Matsumura Y. Nakajima T. Mizuno M. Morinaga M. Sasaki T. Togashi M. Oyama H. Hata M. Watanabe T. Komatsu J. Mizushima-Sugano T. Satoh Y. Shirai Y. Takahashi K. Nakagawa K. Okumura T. Nagase N. Nomura H. Kikuchi Y. Masuho R. Yamashita K. Nakai T. Yada Y.

Nakamura O. Ohara T. Isogai and S. Sugano. 2004. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* 36: 40-45.

Qiu, P. 2003. Recent advances in computational promoter analysis in understanding the transcriptional regulatory network. *Biochem Biophys Res Commun* 309: 495-501.

Rakhade, S.N., B. Yao, S. Ahmed, E. Asano, T.L. Beaumont, A.K. Shah, S. Draghici, R. Krauss, H.T. Chugani, S. Sood, and J.A. Loeb. 2005. A common pattern of persistent gene activation in human neocortical epileptic foci. *Ann Neurol* 58: 736-747.

Ring, R.H., J. Alder, M. Fennell, E. Kouranova, I.B. Black, and S. Thakker-Varia. 2006. Transcriptional profiling of brain-derived-neurotrophic factor-induced neuronal plasticity: a novel role for nociceptin in hippocampal neurite outgrowth. *J Neurobiol* 66: 361-377.

Stormo, G.D. 1990. Consensus patterns in DNA. *Methods Enzymol* 183: 211-221.

Sun, Y.M., D.J. Greenway, R. Johnson, M. Street, N.D. Belyaev, J. Deuchars, T. Bee, S. Wilde, and N.J. Buckley. 2005. Distinct Profiles of REST Interactions with Its Target Genes at Different Stages of Neuronal Development. *Mol Biol Cell* 16: 5630-5638.

Suzuki, Y., R. Yamashita, M. Shirota, Y. Sakakibara, J. Chiba, J. Mizushima-Sugano, A.E. Kel, T. Arakawa, P. Carninci, J. Kawai, Y. Hayashizaki, T. Takagi, K. Nakai, and S. Sugano. 2004. Large-scale collection and characterization of promoters of human and mouse genes. *In Silico Biol* 4: 429-444.

Swirnoff, A.H. and J. Milbrandt. 1995. DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol Cell Biol* 15: 2275-2287.

Tang, Y., A. Lu, B.J. Aronow, and F.R. Sharp. 2001. Blood genomic responses differ after stroke, seizures, hypoglycemia, and hypoxia: blood genomic fingerprints of disease. *Ann Neurol* 50: 699-707.

Valerio, A., M. Ferrario, F.O. Martinez, M. Locati, V. Ghisi, L.G. Bresciani, A. Mantovani, and P. Spano. 2004. Gene expression profile activated by the chemokine CCL5/RANTES in human neuronal cells. *J Neurosci Res* 78: 371-382.

Vanderluit, J.L., K.L. Ferguson, V. Nikoletopoulou, M. Parker, V. Ruzhynsky, T. Alexson, S.M. McNamara, D.S. Park, M. Rudnicki, and R.S. Slack. 2004. p107 regulates neural precursor cells in the mammalian brain. *J Cell Biol* 166: 853-863.

Vavouri, T. and G. Elgar. 2005. Prediction of cis-regulatory elements using binding site matrices--the successes, the failures and the reasons for both. *Curr Opin Genet Dev* 15: 395-402.

Venter, J.C. M.D. Adams E.W. Myers P.W. Li R.J. Mural G.G. Sutton H.O. Smith M. Yandell C.A. Evans R.A. Holt J.D. Gocayne P. Amanatides R.M. Ballew D.H. Huson J.R. Wortman Q. Zhang C.D. Kodira X.H. Zheng L. Chen M. Skupski G. Subramanian P.D. Thomas J. Zhang G.L. Gabor Miklos C. Nelson S. Broder A.G. Clark J. Nadeau V.A. McKusick N. Zinder A.J. Levine R.J. Roberts M. Simon C. Slayman M. Hunkapiller R. Bolanos A.

Delcher I. Dew D. Fasulo M. Flanigan L. Florea A. Halpern S. Hannenhalli S. Kravitz S. Levy C. Mobarry K. Reinert K. Remington J. Abu-Threideh E. Beasley K. Biddick V. Bonazzi R. Brandon M. Cargill I. Chandramouliswaran R. Charlab K. Chaturvedi Z. Deng V. Di Francesco P. Dunn K. Eilbeck C. Evangelista A.E. Gabrielian W. Gan W. Ge F. Gong Z. Gu P. Guan T.J. Heiman M.E. Higgins R.R. Ji Z. Ke K.A. Ketchum Z. Lai Y. Lei Z. Li J. Li Y. Liang X. Lin F. Lu G.V. Merkulov N. Milshina H.M. Moore A.K. Naik V.A. Narayan B. Neelam D. Nusskern D.B. Rusch S. Salzberg W. Shao B. Shue J. Sun Z. Wang A. Wang X. Wang J. Wang M. Wei R. Wides C. Xiao C. Yan A. Yao J. Ye M. Zhan W. Zhang H. Zhang Q. Zhao L. Zheng F. Zhong W. Zhong S. Zhu S. Zhao D. Gilbert S. Baumhueter G. Spier C. Carter A. Cravchik T. Woodage F. Ali H. An A. Awe D. Baldwin H. Baden M. Barnstead I. Barrow K. Beeson D. Busam A. Carver A. Center M.L. Cheng L. Curry S. Danaher L. Davenport R. Desilets S. Dietz K. Dodson L. Doup S. Ferriera N. Garg A. Gluecksmann B. Hart J. Haynes C. Haynes C. Heiner S. Hladun D. Hostin J. Houck T. Howland C. Ibegwam J. Johnson F. Kalush L. Kline S. Koduru A. Love F. Mann D. May S. McCawley T. McIntosh I. McMullen M. Moy L. Moy B. Murphy K. Nelson C. Pfannkoch E. Pratts V. Puri H. Qureshi M. Reardon R. Rodriguez Y.H. Rogers D. Romblad B. Ruhfel R. Scott C. Sitter M. Smallwood E. Stewart R. Strong E. Suh R. Thomas N.N. Tint S. Tse C. Vech G. Wang J. Wetter S. Williams M. Williams S. Windsor E. Winn-Deen K. Wolfe J. Zaveri K. Zaveri J.F. Abril R. Guigo M.J. Campbell K.V. Sjolander B. Karlak A. Kejariwal H. Mi B. Lazareva T. Hatton A. Narechania K. Diemer A. Muruganujan N. Guo S. Sato V. Bafna S. Istrail R. Lippert R. Schwartz B. Walenz S. Yooseph D. Allen A. Basu J. Baxendale L. Blick M. Caminha J. Carnes-Stine P. Caulk Y.H. Chiang M. Coyne C. Dahlke A. Mays M. Dombroski M. Donnelly D. Ely S. Esparham C. Fosler H. Gire S. Glanowski K. Glasser A. Glodek M. Gorokhov K. Graham B. Gropman M. Harris J. Heil S. Henderson J. Hoover D. Jennings C. Jordan J. Jordan J. Kasha L. Kagan C. Kraft A. Levitsky M. Lewis X. Liu J. Lopez D. Ma W. Majoros J. McDaniel S. Murphy M. Newman T. Nguyen N. Nguyen M. Nodell S. Pan J. Peck M. Peterson W. Rowe R. Sanders J. Scott M. Simpson T. Smith A. Sprague T. Stockwell R. Turner E. Venter M. Wang M. Wen D. Wu M. Wu A. Xia A. Zandieh and X. Zhu. 2001. The sequence of the human genome. *Science* 291: 1304-1351.

Vinogradov, A.E. 2005. Dualism of gene GC content and CpG pattern in regard to expression in the human genome: magnitude versus breadth. *Trends Genet* 21: 639-643.

Vo, N., M.E. Klein, O. Varlamova, D.M. Keller, T. Yamamoto, R.H. Goodman, and S. Impey. 2005. A cAMP-response element binding protein-induced microRNA regulates neuronal morphogenesis. *Proc Natl Acad Sci U S A* 102: 16426-16431.

Waterston, R.H. K. Lindblad-Toh E. Birney J. Rogers J.F. Abril P. Agarwal R. Agarwala R. Ainscough M. Alexandersson P. An S.E. Antonarakis J. Attwood R. Baertsch J. Bailey K. Barlow S. Beck E. Berry B. Birren T.

Bloom P. Bork M. Botcherby N. Bray M.R. Brent D.G. Brown S.D. Brown C. Bult J. Burton J. Butler R.D. Campbell P. Carninci S. Cawley F. Chiaromonte A.T. Chinwalla D.M. Church M. Clamp C. Clee F.S. Collins L.L. Cook R.R. Copley A. Coulson O. Couronne J. Cuff V. Curwen T. Cutts M. Daly R. David J. Davies K.D. Delehaunty J. Deri E.T. Dermitzakis C. Dewey N.J. Dickens M. Diekhans S. Dodge I. Dubchak D.M. Dunn S.R. Eddy L. Elnitski R.D. Emes P. Eswara E. Eyras A. Felsenfeld G.A. Fewell P. Flicek K. Foley W.N. Frankel L.A. Fulton R.S. Fulton T.S. Furey D. Gage R.A. Gibbs G. Glusman S. Gnerre N. Goldman L. Goodstadt D. Grafham T.A. Graves E.D. Green S. Gregory R. Guigo M. Guyer R.C. Hardison D. Haussler Y. Hayashizaki L.W. Hillier A. Hinrichs W. Hlavina T. Holzer F. Hsu A. Hua T. Hubbard A. Hunt I. Jackson D.B. Jaffe L.S. Johnson M. Jones T.A. Jones A. Joy M. Kamal E.K. Karlsson D. Karolchik A. Kasprzyk J. Kawai E. Keibler C. Kells W.J. Kent A. Kirby D.L. Kolbe I. Korf R.S. Kucherlapati E.J. Kulbokas D. Kulp T. Landers J.P. Leger S. Leonard I. Letunic R. Levine J. Li M. Li C. Lloyd S. Lucas B. Ma D.R. Maglott E.R. Mardis L. Matthews E. Mauceli J.H. Mayer M. McCarthy W.R. McCombie S. McLaren K. McLay J.D. McPherson J. Meldrim B. Meredith J.P. Mesirov W. Miller T.L. Miner E. Mongin K.T. Montgomery M. Morgan R. Mott J.C. Mullikin D.M. Muzny W.E. Nash J.O. Nelson M.N. Nhan R. Nicol Z. Ning C. Nusbaum M.J. O'Connor Y. Okazaki K. Oliver E. Overton-Larty L. Pachter G. Parra K.H. Pepin J. Peterson P. Pevzner R. Plumb C.S. Pohl A. Poliakov T.C. Ponce C.P. Ponting S. Potter M. Quail A. Reymond B.A. Roe K.M. Roskin E.M. Rubin A.G. Rust R. Santos V. Sapojnikov B. Schultz J. Schultz M.S. Schwartz S. Schwartz C. Scott S. Seaman S. Searle T. Sharpe A. Sheridan R. Shownkeen S. Sims J.B. Singer G. Slater A. Smit D.R. Smith B. Spencer A. Stabenau N. Stange-Thomann C. Sugnet M. Suyama G. Tesler J. Thompson D. Torrents E. Trevaskis J. Tromp C. Ucla A. Ureta-Vidal J.P. Vinson A.C. Von Niederhausern C.M. Wade M. Wall R.J. Weber R.B. Weiss M.C. Wendl A.P. West K. Wetterstrand R. Wheeler S. Whelan J. Wierzbowski D. Willey S. Williams R.K. Wilson E. Winter K.C. Worley D. Wyman S. Yang S.P. Yang E.M. Zdobnov M.C. Zody and E.S. Lander. 2002. Initial sequencing and comparative analysis of the mouse genome. ***Nature*** **420: 520-562.**

Wheeler, D.L., T. Barrett, D.A. Benson, S.H. Bryant, K. Canese, D.M. Church, M. DiCuccio, R. Edgar, S. Federhen, W. Helmberg, D.L. Kenton, O. Khovayko, D.J. Lipman, T.L. Madden, D.R. Maglott, J. Ostell, J.U. Pontius, K.D. Pruitt, G.D. Schuler, L.M. Schriml, E. Sequeira, S.T. Sherry, K. Sirotkin, G. Starchenko, T.O. Suzek, R. Tatusov, T.A. Tatusova, L. Wagner, and E. Yaschenko. 2005. Database resources of the National Center for Biotechnology Information. ***Nucleic Acids Res*** **33: D39-45.**

Wilson, D.N., H. Chung, R.C. Elliott, E. Bremer, D. George, and S. Koh. 2005. Microarray analysis of postictal transcriptional regulation of neuropeptides. ***J Mol Neurosci*** **25: 285-298.**

Wingender, E. 2004. TRANSFAC, TRANSPATH and CYTOMER as starting points for an ontology of regulatory networks. ***In Silico Biol*** **4: 55-61.**

Wingender, E., X. Chen, R. Hehl, H. Karas, I. Liebich, V. Matys, T. Meinhardt, M. Pruss, I. Reuter, and F. Schacherer. 2000. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res* 28: 316-319.

Xu, W., K. Angelis, D. Danielpour, M.M. Haddad, O. Bischof, J. Campisi, E. Stavnezer, and E.E. Medrano. 2000. Ski acts as a co-repressor with Smad2 and Smad3 to regulate the response to type beta transforming growth factor. *Proc Natl Acad Sci U S A* 97: 5924-5929.

Yao, W.D., R.R. Gainetdinov, M.I. Arbuckle, T.D. Sotnikova, M. Cyr, J.M. Beaulieu, G.E. Torres, S.G. Grant, and M.G. Caron. 2004. Identification of PSD-95 as a regulator of dopamine-mediated synaptic and behavioral plasticity. *Neuron* 41: 625-638.

Yin, J.C., J.S. Wallach, M. Del Vecchio, E.L. Wilder, H. Zhou, W.G. Quinn, and T. Tully. 1994. Induction of a dominant negative CREB transgene specifically blocks long-term memory in Drosophila. *Cell* 79: 49-58.

Zhong, X., J.R. Liu, J.W. Kyle, D.A. Hanck, and W.S. Agnew. 2006. A profile of alternative RNA splicing and transcript variation of CACNA1H, a human T-channel gene candidate for idiopathic generalized epilepsies. *Hum Mol Genet* 15: 1497-1512.