Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press



Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE

Eivind Valen, Giovanni Pascarella, Alistair Chalk, et al.

Genome Res. published online December 11, 2008
Access the most recent version at doi:10.1101/gr.084541.108

P<P Published online December 11, 2008 in advance of the print journal.

License

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



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

Copyright © 2009 by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE

Eivind Valen,[1] Giovanni Pascarella,[2] Alistair Chalk,[3] Norihiro Maeda,[4] Miki Kojima,[4] Chika Kawazu,[4] Mitsuyoshi Murata,[4] Hiromi Nishiyori,[4] Dejan Lazarevic,[2,8] Dario Motti,[2] Troels Torben Marstrand,[1] Man-Hung Eric Tang,[1] Xiaobei Zhao,[1] Anders Krogh,[1] Ole Winther,[1] Takahiro Arakawa,[4] Jun Kawai,[4] Christine Wells,[3] Carsten Daub,[5] Matthias Harbers,[7] Yoshihide Hayashizaki,[4] Stefano Gustincich,[2] Albin Sandelin,[1,9] and Piero Carninci[4,6,9]

[1]*The Bioinformatics Centre, Department of Biology and Biotech Research and Innovation Centre, University of Copenhagen, Ole Maaloes vej 5, DK-2200, Denmark;* [2]*The Giovanni Armenise-Harvard Foundation Laboratory, Sector of Neurobiology, International School for Advanced Studies (SISSA), Basovizza, 34012 Trieste, Italy;* [3]*The National Centre for Adult Stem Cell Research, The Eskitis Institute for Cell and Molecular Therapies, Griffith University, Nathan QLD 4111, Australia;* [4]*LSA Technology Development Group, Omics Science Center, RIKEN Yokohama Institute, Yokohama, Kanagawa 230-0045 Japan;* [5]*LSA Bioinformatics Team, Omics Science Center, RIKEN Yokohama Institute, Yokohama, Kanagawa 230-0045 Japan;* [6]*Functional Genomics Technology Team, Omics Science Center, RIKEN Yokohama Institute, Yokohama, Kanagawa 230-0045 Japan;* [7]*DNAFORM Inc., Yokohama, Kanagawa 230-0046, Japan;* [8]*CBM Scrl–Consorzio per il Centro di Biomedicina Molecolare, Basovizza, 34012 Trieste, Italy*

Finding and characterizing mRNAs, their transcription start sites (TSS), and their associated promoters is a major focus in post-genome biology. Mammalian cells have at least 5–10 magnitudes more TSS than previously believed, and deeper sequencing is necessary to detect all active promoters in a given tissue. Here, we present a new method for high-throughput sequencing of 5′ cDNA tags—DeepCAGE: merging the Cap Analysis of Gene Expression method with ultra-high-throughput sequence technology. We apply DeepCAGE to characterize 1.4 million sequenced TSS from mouse hippocampus and reveal a wealth of novel core promoters that are preferentially used in hippocampus: This is the most comprehensive promoter data set for any tissue to date. Using these data, we present evidence indicating a key role for the *Arnt2* transcription factor in hippocampus gene regulation. DeepCAGE can also detect promoters used only in a small subset of cells within the complex tissue.

[Supplemental material is available online at www.genome.org. CAGE tag sequences have been submitted to the DNA Data Bank of Japan (http://www.ddbj.nig.ac.jp/) under accession nos. AGAAA0000001–AGAAA0552486. Processed CAGE data sets are freely available at http://people.binf.ku.dk/albin/supplementary_data/hcamp/.]

Transcription initiation is one of the most fundamental cellular processes. The identification of transcription start sites (TSS) leads to the detection of the associated core promoters. Historically, precise definition of TSS has been laborious and addressed one gene at a time. Therefore, few genes have had their start sites mapped in detail. We, and others, have presented techniques that can identify TSS on a genome-wide scale—typically, by generating full-length cDNAs and then sequencing short tags at their 5′ ends (Ng et al. 2005; Kodzius et al. 2006). The largest study to date using such methods (Carninci et al. 2006) was carried out by taking advantage of the cap analysis of gene expression (CAGE) technology (Kodzius et al. 2006). In the FANTOM3 project, tag libraries were sequenced from 22 tissues with an average ~48,500 tags per library in mouse. This study gave new insights into how transcription initiation works, as reviewed (Muller et al. 2007; Sandelin et al. 2007), and suggested that mammalian cells have many more core promoters than previously appreciated. In a previous study (Gustincich et al. 2006), we have also shown that mouse brain regions have a higher number of active TSS than other tissues, presumably leading to a higher diversity of distinct transcripts in these tissues. Since the average mammalian cell is estimated to express at least 350,000 mRNAs (Jackson et al. 2000) and the brain is a highly complex tissue with many distinct regions, which in themselves have high cellular heterogeneity, it is evident that the promoters sampled so far are just skimming the surface of the brain transcriptional complexity.

Among the different regions of the brain, the hippocampal formation (hippocampus, dentate gyrus, and subiculum) has been the subject of intense studies due to its essential role in the formation of new episodic and long-term memories (Bird and Burgess 2008). In particular, the hippocampus has been a major experimental system to unveil the role of synaptic plasticity. The hippocampus has also been implicated in a number of neurological and psychiatric disorders including epilepsy and Alzheimer's disease. Additionally, the subgranular zone of the dentate gyrus is one of the sites of adult neurogenesis (Zhao et al. 2008).

The study of the hippocampal formation has relied on its distinctive and readily identifiable structure at both gross and histological levels (Paxinos 2004). The hippocampus is divided

[9]**Corresponding authors.**
E-mail albin@binf.ku.dk; fax 45-3532-1281.
E-mail carninci@riken.jp; fax 81-48-4624868.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.084541.108.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

into fields CA1–CA3 comprising the pyramidal cell layer, where the pyramidal cells are present, and a heterogeneous group of diverse GABAergic interneurons (Parra et al. 1998; Maccaferri and Lacaille 2003). The dentate gyrus comprises the molecular, principal, and polymorphic layers where the granule cells are the principal cells and the pyramidal basket cells are the most prominent class of interneurons. In a few cases, the use of cell-type-specific knockout mice for ligands or receptors of neuroactive molecules has made it possible to integrate physiological, anatomical, and molecular data from synapses to understand animal behavior (Tsien et al. 1996; Nakazawa et al. 2004). However, our understanding of the physiological organization of the hippocampus has been hampered by difficulties in describing the complete repertoire of neuronal cell types and their properties (Parra et al. 1998).

Thus, the identification of specific promoters that initiate transcription and drive gene expression in hippocampus, or even in specific subsets of cells within the tissue, will be important for developing new lines of mice with cell- or tissue-specific gene isoforms labeled and/or knocked out, or by the use of neuronal cell ablation (Watanabe et al. 1998; Tonegawa et al. 2003). The identification of novel promoters, particularly those that drive expression in a small number of neurons, may also lead to the description of rare neuronal types that have not been previously characterized.

To this end, we present DeepCAGE, a method combining CAGE technology with a high-throughput tag sequencer, the GS20 sequencer (454 Life Sciences [Roche]). We use DeepCAGE to create a comprehensive resource of hippocampal TSS and core promoters. Since with this method we can reach an unprecedented sampling depth (2 million tags, of which 1.4 million can be unambiguously mapped to the genome), we explored this data set to find promoters preferentially used in the hippocampus, their effects on the proteome, correlation with spatial expression data, and to understand the transcriptional regulatory program in the hippocampus.

Results

DeepCAGE sequencing

We adapted the CAGE method (Shiraki et al. 2003; Kodzius et al. 2006) to the 454 Life Sciences (Roche) GS20 sequencer as described in Methods (Fig. 1). Briefly, total RNA from C57B6/J pooled mice hippocampi was purified and used as a template in the first-strand cDNA reaction primed by random primers to capture both the poly(A)$^+$ and poly(A)$^-$ RNA species. To extend cDNA synthesis through GC-rich regions in the 5′ UTR, we carried out the reverse transcription reaction at high temperature in the presence of trehalose and sorbitol (Carninci et al. 2002). cDNAs reaching the cap site were then selected by cap-trapping. They were then ligated to a linker having a recognition site for the class-IIs restriction endonuclease *MmeI* just next to the start of the cDNAs corresponding to the 5′-end of the original RNAs. This linker was used to prime second-strand cDNA synthesis. Subsequently, *MmeI* digestion cleaved 20 ~ 21 bp within the double-stranded cDNA, releasing CAGE tags. After ligation of a second linker to the 3′ end opened by MmeI digestion, CAGE tags were PCR amplified, purified, and further amplified before restriction and concatenation for direct sequencing (see Methods). The DeepCAGE technology does not require cloning in bacteria.

The key step for direct sequencing on the 454 device is the introduction of specific primer sites at the ends of the concatamers by mixing the CAGE tags at a ratio of 20:1 with a mixture of the 454 linkers "A" and "B." Since these linkers (Margulies et al. 2005) can ligate DNA only on one side, they terminate the concatenation reaction and provide ends suitable for sequencing when A and B appear on the opposite sides of the concatamer, regardless of the orientation of the CAGE tags and insert size, which was optimized to be ~500 bp.

After a first test run, we produced two large-scale reactions, achieving in total $\sim 2 \times 10^6$ CAGE tags. After sequencing, tags were mapped to the mouse genome (mm8 assembly) using an algorithm based on our previous studies (see Methods); in total, 1.4×10^6 tags map with high stringency (see Supplemental Table S1). The same mapping protocol was applied to all other CAGE tag libraries that are part of the data set used in this study. All CAGE tags have been submitted to the DNA Data Bank of Japan (DDBJ) under accession numbers AGAAA0000001–AGAAA0552486.

Data sets and resource preparation

Similar to most other high-throughput genomic technologies (The ENCODE Consortium 2007), CAGE data are most useful together with other data sets. In this study, we compare the hippocampus CAGE data to seven other CAGE data sets—each corresponding to a different tissue—with varying sequencing depth, including three brain tissues: visual cortex, somatosensory cortex, and cerebellum (see Supplemental Table S1). We also use the FANTOM3 cDNA set for associating promoters with genes and gene annotation (Carninci et al. 2005). As a resource for the community, the CAGE data sets (including "tissue-specific" promoter sets discussed below) prepared here are freely available as data tracks and sequence files at http://people.binf.ku.dk/albin/supplementary_data/hcamp/, from where they can be directly uploaded to the UCSC browser for visualization, or downloaded for analysis by power users. Additional statistics such as the fraction of tags mapping to known 5′-ends are shown in Supplemental Figure S1.

Exploration of tissue preferences for core promoters

As shown previously, for analysis of core promoters, it is helpful to group CAGE tags that map close to each other on the genome. Using the method of Carninci et al. (2006), tags from any tissue were grouped into a tag cluster if their genome mapping coordinates overlap on the same strand.

To explore the overall tissue preference of the CAGE clusters, we selected all CAGE tag clusters that have more than 30 tags per million (TPM), when counting all tissues. For clarity, TPM normalization is commonly used in tag-based studies and can be described as normalizing all tag counts so that the total count of mapped tags within a library equals 10^6 tags. The reason for this conservative cutoff is that a certain number of tags are needed to assess tissue distributions. For simplicity, we refer to these clusters as core promoters in this study, in the same way as in Carninci et al. (2006) and Sandelin et al. (2007).

This analysis identified 18,948 core promoters. To explore both what fraction of promoters are expressed primarily in one or a subset of tissues and what tissues have similar promoter usage, we hierarchically clustered these core promoters in terms of their expression and visualized the results as a heatmap (Fig. 2A; Eisen et al. 1998). We observe that:

1. The brain tissues cluster together in terms of promoter usage; in particular, visual and somatosensory have the most similar

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

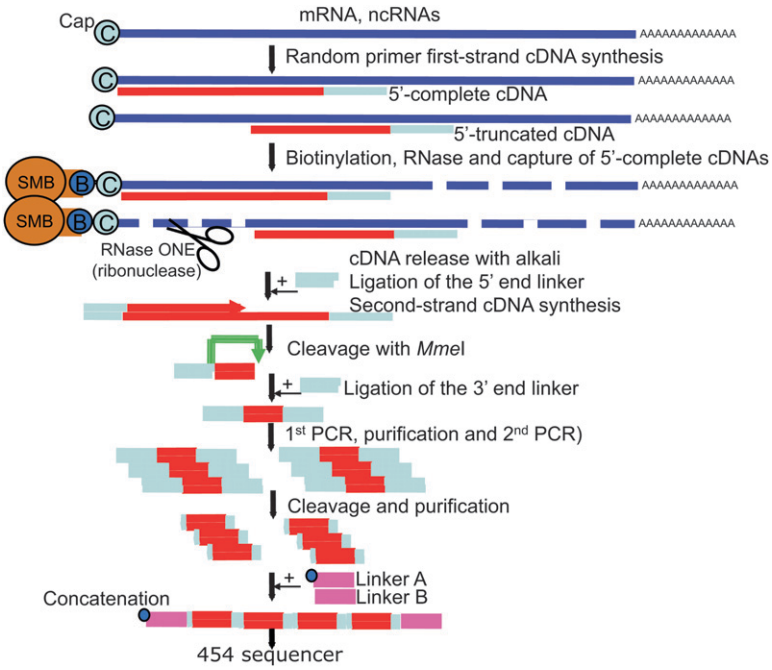


Figure 1. Preparation of DeepCAGE cDNA libraries. cDNA is produced with reverse transcriptase using random priming to maximize chances to reach the cap sites and to include non-polyadenylated RNAs. The cap site is biotinylated, followed by cleavage of single-strand RNA (incomplete cDNA). After recovery of the cDNA, a linker is ligated on the 5′-end, which carries the MmeI restriction site, which cleaves 20/21 nt inside the 5′-end of the cDNA. After a second linker ligation and PCR amplification, restricted digested tags are purified and finally concatenated together with the "454" linkers A and B. (SMB) Streptavidin-coated magnetic beads; (B) biotin.

usage, with few CAGE tag clusters being preferentially used in only one of these tissues.

2. The cortex tissues have substantial "smearing." Many promoters are used, but they are also shared between at least two tissues. This is a property seen also for macrophage and lung tissues, which might be due to the large number of macrophage cells present in lung tissue.
3. Conversely, there is a large set of promoters that are used mostly in the hippocampus: This promoter set has very little smearing, a feature shared with less complex tissues such as liver. The cerebellum is somewhere in between the hippocampus and the cortex tissues in terms of smearing and specificity.

Note that this analysis only captures general properties of the sets—there are always individual exceptions to any "rules" implied, but the goal is to see the overall tendencies. The results from this data exploration suggest that both the hippocampus and cerebellum have more core promoters that are biased toward the tissue in question than other sampled brain tissues. We explore these in the next section.

Definition of preferentially expressed promoters

We then identified the core promoters that are significantly biased toward individual tissues. Historically, in molecular biology literature, promoters with this property are often called tissue specific: We will avoid this term here since it implies exclusive expression in a given tissue, which is impossible to ascertain since we cannot sample all tissues with infinite depth. We can only assess the properties of the tissues sampled, and the goal is to find tag clusters that are *preferentially* used in one of the sampled set of tissues compared to the rest of tissues. We refer to such clusters as preferentially expressed promoters (PEPs), used in the following context: A liver PEP is a promoter that is preferentially used in liver tissue. We term a core promoter preferentially expressed for a certain tissue if (1) the number of tags from this tissue is greater than the sum of all other tags from other tissues, normalized for library size; and (2) this overrepresentation is statistically significant (see Methods).

Using this method, out of the 18,948 core promoters assessed, 6536 (34%) are preferentially used in a single tissue. Hippocampus, cerebellum, and liver have the most PEPs, while visual and somatosensory cortex have the fewest (Fig. 2B). Interestingly, only 8% of the hippocampus PEPs analyzed have only hippocampus tags, although this fraction is dependent on the expression cutoff (30 TPMs) used in the selection of tag clusters to start with: Smaller cutoffs give larger number promoters only detected in hippocampus (Supplemental Fig. S2).

As noted previously (Gustincich et al. 2006), promoters used preferentially in brain generally have multiple TSS and higher CpG content compared to promoters used preferentially in other tissues, which often have a single peak distribution of TSS, governed by a TATA-box (Sandelin et al. 2007). We found that hippocampus PEPs share the properties of the other brain PEPs in this regard—broader, CpG-rich promoters with fewer TATA patterns (Supplemental Figs. S3–S5). We also note that for promoters that are used strongly in many brain tissues, the hippocampus tag usage at the nucleotide level generally correlates well with tag distributions from the other brain tissues (Supplemental Figs. S6 and S7). There are exceptions to this; promoters where the tag distribution shape differs between tissues have been explored previously by Kawaji et al. (2006).

We then assessed where the PEPs from the various tissues were located in terms of overlap with known genes (see Methods). Figure 2B shows that the hippocampus, apart from having the largest amount of PEPs overall, also has the greatest number of PEPs located in intronic and intergenic space. This indicates that there are many strong promoters preferentially expressed in the hippocampus that have no known corresponding gene. This observation is not simply a sample size effect, as the cerebellum has almost as many PEPs with just ~18% of the sequencing depth compared to the hippocampus.

RACE validates distal upstream promoters

While PEPs falling within genes can be considered candidate alternative promoters for the same gene, PEPs in intergenic space are more likely to be the promoters of novel transcripts. We selected 10 intergenic hippocampus PEPs for RACE validation. Importantly, the selection was not in any way based on additional information that would be a validation in itself, such as EST

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

sequences. Out of the 10 cases, eight had a PCR product, and the sequenced product validated a hippocampus PEPs in six of these cases. Of the failed cases, one had supporting evidence from other sources (overlap of 5′-ends of spliced ESTs) (data not shown). The outcome is comparable to that of RACE validation of intergenic transcribed regions from tiling array data in the ENCODE project (50%–70% success rate) (The ENCODE Project Consortium 2007); however, this should not be viewed as a true sensitivity measure of

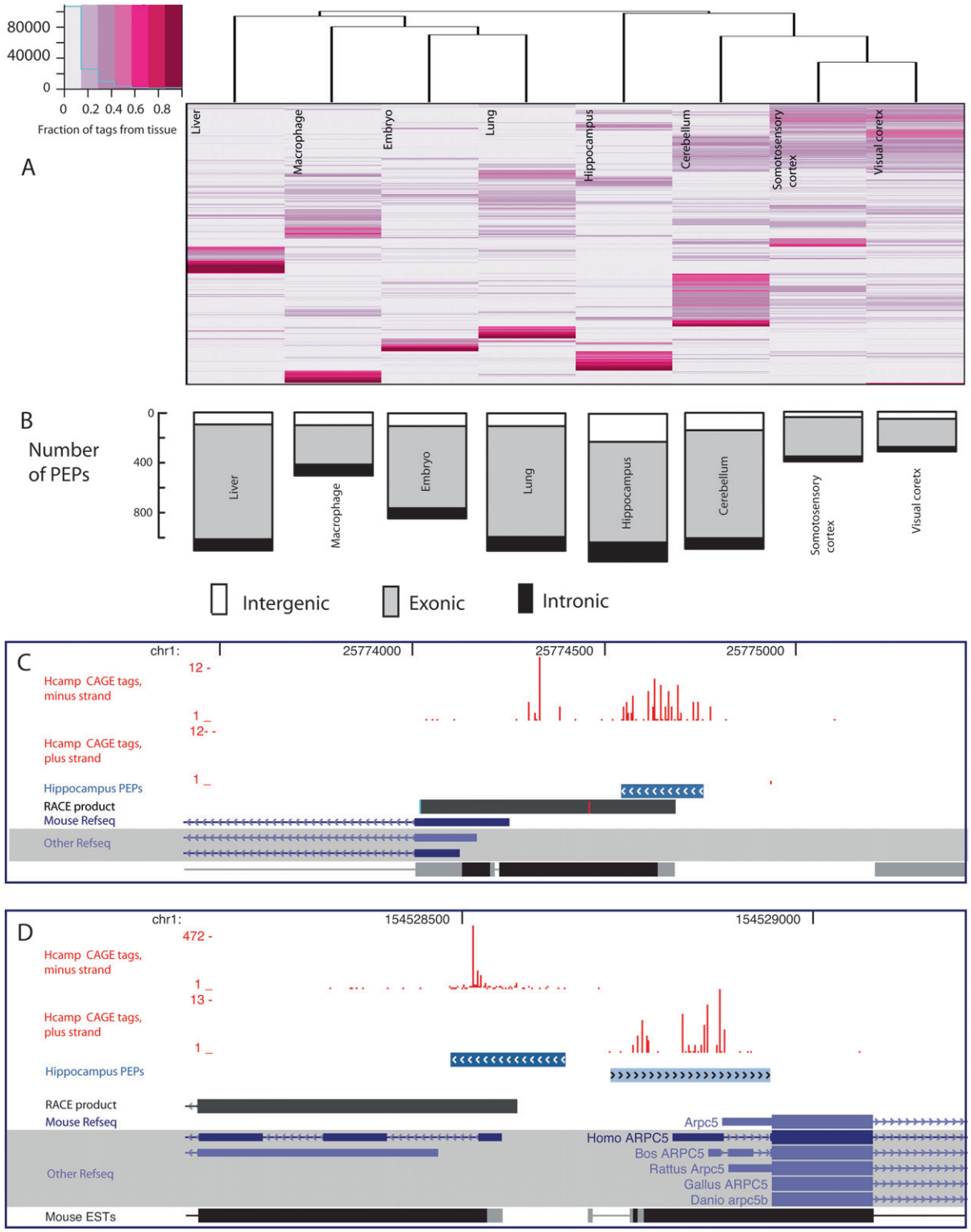


Figure 2. (Legend on next page)

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

DeepCAGE, for three principal reasons: (1) Even in perfect circumstances, RACE does not have perfect sensitivity; (2) as we are focusing on novel core promoters, we do not know the exon structure of the downstream product, which makes primer design nontrivial; and (3) many of the promoters have high GC content, which makes amplification challenging. Nevertheless, these results show that promoters inferred by DeepCAGE can be detected by other methods, as already shown previously in extensive validation experiments of the original FANTOM3 CAGE study (see Supplemental Material of Carninci et al. 2006).

In the annotation process, we noticed a considerable number of cases in which intergenic hippocampus PEPs were located relatively close to the 5′-end of a known gene. These cases are likely novel alternative upstream promoters. An example is shown in Figure 2C: CAGE and RACE data show a novel hippocampus PEP upstream of the mouse *Bai3* (brain-specific angiogenesis inhibitor 3) gene. An extreme case is shown in Figure 2D: CAGE identifies a hippocampus PEP that is upstream of the *Arpc5* (actin-related protein 2/3 complex, subunit 5) gene, but on the other strand, forming a bidirectional promoter. RACE validation as well as human orthologous transcripts and EST evidence show that the novel promoter is likely a distal upstream alternative promoter of the *Rgl1* (ral guanine nucleotide dissociation stimulator-like 1) gene, whose RefSeq-annotated start site (which is also a hippocampus PEP) (data not shown) is a remarkable ~141 kb downstream from the novel promoter. Therefore, while we focused on intergenic promoters to find novel transcripts, we often identified novel promoters that provide new ways to regulate the transcription of known genes.

Brain tissues use different alternative promoters within the same gene

As shown in Figure 2B, the majority of hippocampus PEPs are located inside genes, overlapping annotated exons. It is likely that many of these are alternative promoters for the same gene, since many of them are supported by full-length cDNAs (see examples in Figs. 3 and 4). Alternative promoters are interesting for three reasons: (1) They allow a gene to have multiple, distinct, regulatory inputs; (2) alternative promoter locations can affect the protein content of the gene product similarly to alternative splicing; and (3) it is important for molecular approaches in neurobiology to selectively knock down gene isoforms that are preferentially used in a given tissue.

We first identified all genes containing one or more PEPs from hippocampus, somatosensory cortex, visual cortex, and cerebellum inside exons. Then, we counted the number of genes with multiple distinct PEPs from the different tissues (Fig. 3A). The *Dlgap1* gene (guanylate kinase-associated protein [GKAP] or SAPAP [synapse-associated protein 90-postsynaptic density-95-associated protein]) (Fig. 3B) is exceptional since it has four core promoters that are preferentially used in hippocampus, somatosensory cortex, visual cortex, and cerebellum, respectively. All of these PEPs overlap corresponding 5′-ends from full-length cDNAs (Fig. 3B); in this case, the CAGE verifies these 5′-ends and assigns tissue expression constraints. *Dlgap1* is a scaffolding postsynaptic density protein at excitatory synapsis that contains 14-amino-acid repeats at the N terminus involved in protein–protein interactions and that are affected by different promoter usage (Kim et al. 1997; Romorini et al. 2004); the CAGE data indicate that all the PEPs are upstream of these repeats except for the cerebellum, indicating that cerebellum transcripts do not include the repeats. Thus, the selection of alternative promoters has in this case a clear functional consequence.

We then sought to systematically identify potential changes in protein domain composition caused by usage of hippocampus PEPs. Using cDNA data, we predicted protein domains to genomic positions and determined in how many cases a hippocampus PEP falls within a gene but downstream from a protein domain within the same gene, which then would give a protein product that is lacking the domain in question. Using conservative criteria (see Methods), we found 50 such genes (see Supplemental Material). Three examples (*Pclo*, *Bai1*, and *Myo10*), showing dramatic protein domain content diversity, are shown in Figure 4.

Transcription factor binding sites analysis on specific core promoters

An advantage with the CAGE is that tags give high-resolution mappings of active TSS, which can be used to pinpoint core promoters for computational sequence analyses (Wasserman and Sandelin 2004). We first analyzed the −1000 to +200 region surrounding PEPs from the tissues in Supplemental Table S1 for significantly overrepresented motif matches from the JASPAR database (Vlieghe et al. 2006). Our results are largely consistent with previous studies of promoters used primarily in single tissues—for instance, homeobox motifs are overrepresented in embryo PEPS, ETS motifs in macrophage PEPs, and so on (Supplemental Tables S3 and S4). Since CAGE data may also be interpreted as promoter usage (the number of tags mapped to a loci), we investigated what transcription factor genes are strongly expressed in hippocampus, and whether their predicted transcription factor binding sites have a clear preference to the promoters that are preferentially used in the same tissue.

Figure 5A plots the fraction of hippocampus tags in transcription factor genes compared to other tissues versus the overall hippocampus expression of the same genes (CAGE TPMs) (see Methods). Only a handful of transcription factor genes stand out as very highly expressed in hippocampus, including *Arnt2*, *Sp3*, and *Aes*, and only some of these have a clear preference for

Figure 2. Exploration and validation of identified core promoters. (*A*) Exploration of tissue usage in all core promoters having more than 30 tags per million using hierarchical clustering, with CAGE tag expression data from the actual core promoters. Preferential usage for a certain tissue (fraction of tags belonging to the tissue in question) is color-coded as shown in the legend. Rows represent individual promoters (the row dendrogram is omitted because of the large number of rows), while columns are the different tissues. (*B*) Number of core promoters used preferentially in just one tissue (PEPs, as defined in Methods) and the locations of these core promoters relative to known genes. (*C*,*D*) Examples of discovery and validation of novel intergenic promoters expressed preferentially in hippocampus. All hippocampus tags are shown as red bar plots (one for each strand). Clusters of tags that are preferentially expressed in hippocampus (hippocampus PEPs) are shown as blue fragments where the color intensity is proportional to the fraction of tags in the cluster from hippocampus vs. other tissues. RACE products are in black. (*C*) The validation of a proximal alternative promoter to the *Bai3* gene. (*D*) The validation of a promoter upstream of the *Arpc5* gene but on the opposite strand; RACE as well as a human orthologous transcript show that this is a distal upstream promoter for the *Rgl1* gene.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

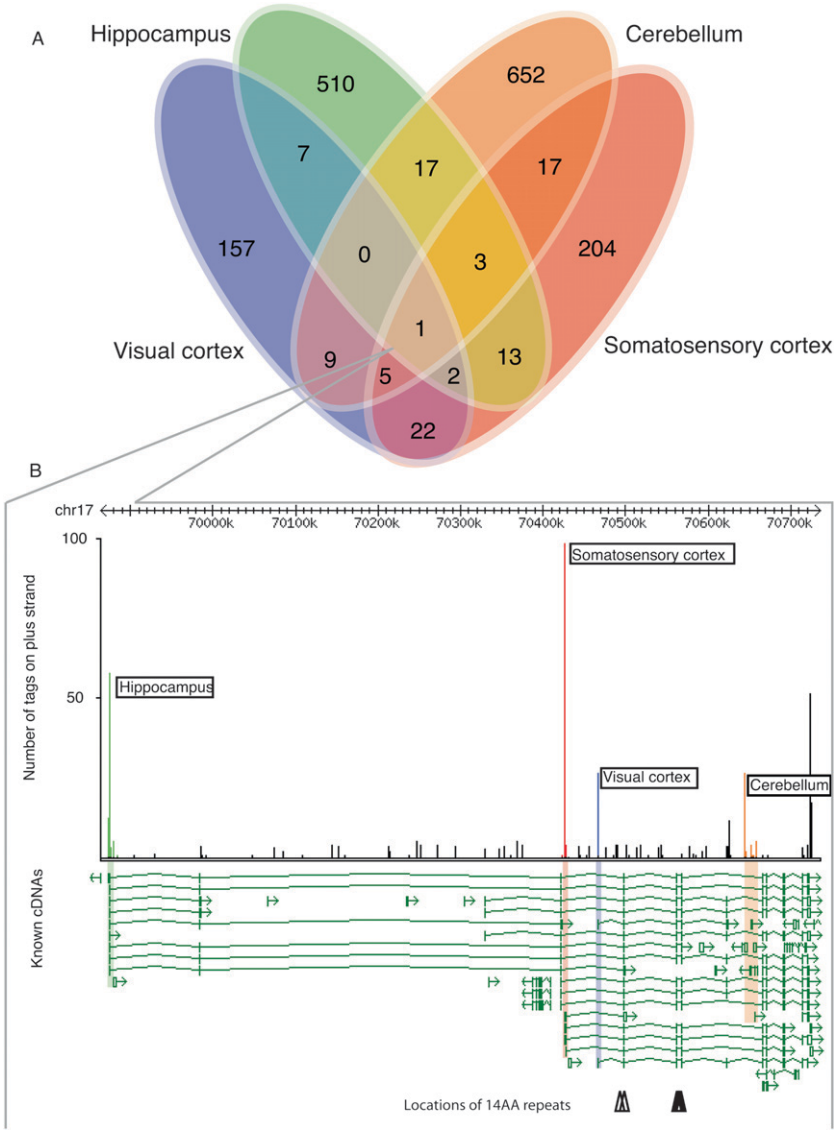


Figure 3. Alternative promoters preferentially expressed in different brain tissues. (*A*) The Venn diagram shows the number of genes having at least one preferentially expressed promoter (PEP) from any of the four sampled brain tissues, or any combination PEPs of the four tissues. For example, there are 15 genes that have at least one hippocampus PEP and one cerebellum PEP. (*B*) The *Dlgap1* gene has four PEPs, one from each brain tissue. All of these are overlapping known 5′-ends inferred by cDNAs. There are five 14-amino-acid-long repeats in the N-terminal end of the corresponding protein that affects the ability to interact with other proteins. Repeats are indicated as triangles. Note that the cerebellum PEP is downstream from these and will give a protein that lacks these repeats.

hippocampus. We compared these highly expressed transcription factors with mouse in situ hybridization experiments from the Allen Brain Atlas (Lein et al. 2007, see image 5B-I). Overall, there is a high correspondence between both preferential expression in hippocampus and overall strength of expression between CAGE and the in situ hybridization data. The general pattern is that in situ data show high correlation with the CAGE data if the number of CAGE tags is high, while the gene is sometimes not visibly expressed in situ if the CAGE tag count is low; this is likely because a certain number of transcripts are necessary to get a visible signal in in situ hybridization, whereas CAGE technology may sample smaller numbers of transcripts in a cell.

For most of the highly expressed transcription factors, we have no corresponding computational model for how they bind DNA. However, since many transcription factors from the same structural class bind similar target sequences (Sandelin and Wasserman 2004), observed overrepresentation of hits with a given model might be due to binding sites from a different factor with similar binding preferences. As an example, the predicted sites for the well-studied bHLH-PAS *Arnt* gene are overrepresented in hippocampus PEPs (Supplemental Table S3), but the *Arnt* gene is lowly expressed in the whole brain as measured by both CAGE and in situ data (data not shown). Interestingly, its paralog *Arnt2* is primarily expressed in brain. According to the CAGE data, *Arnt2* is highly, and preferentially, expressed in hippocampus (Fig. 5A). Furthermore, in situ images of *Arnt2* confirm a distinctive expression of *Arnt2* in the C1 region of the hippocampus (Fig. 5B). This leads to the hypothesis that the *Arnt* predicted sites are, in fact, sites for *Arnt2*, which would make *Arnt2* a major factor in hippocampus transcription regulation.

Promoters used in restricted cell types in hippocampus

We have focused above on relatively strong promoters having more than 30 TPMs in order to study the distribution of tags from different tissues in a statistically valid way. The significance of transcripts that are present in a tissue with low frequency has been met with suspicion, and the observations were often labeled either as methodological or transcriptional noise. Although both of these are still a possibility, we have explored the expression properties of known genes having substantially less than 30 TPM by analyzing their spatial expression patters using in situ hybridization data. In Figure 6, we compare the number of tags hitting a known gene and the corresponding in situ images. Note that in almost all these cases, the tags hit the annotated 5′-end of the gene (Supplemental Fig. S8).

From Figure 6, we note that rare tags frequently identify genes whose expression is restricted only to a reduced and well-defined cell population within the tissue, which may have specific physiological roles. Within these cells, the gene is highly expressed, but since the CAGE experiment is performed on all the cells in the tissue, the total signal is averaged out. Development of next-generation technologies to prepare CAGE libraries from individual neuronal populations will further address this variability. Nevertheless, the ability of DeepCAGE to detect transcripts used in only a handful of cells in a complex tissue shows the utility of the method.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

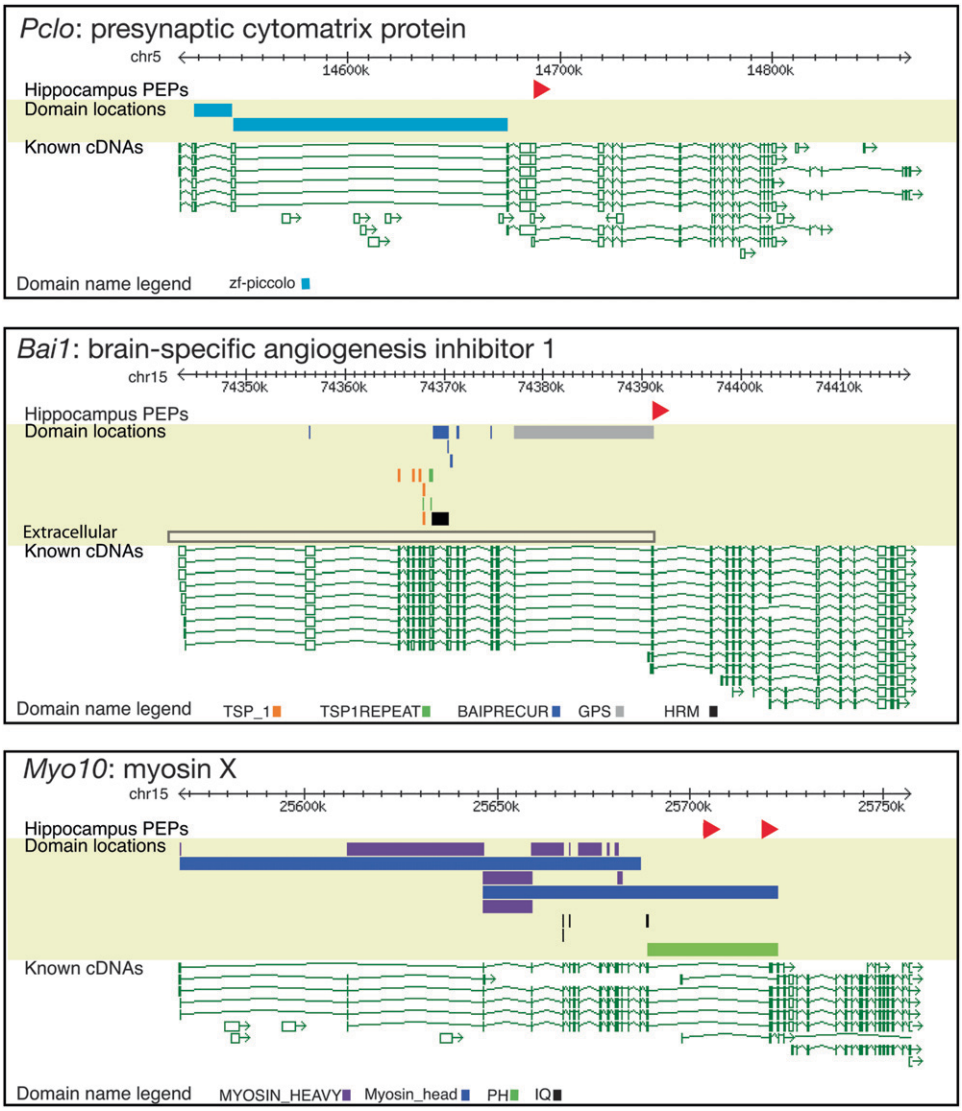


Figure 4. Examples of changes of domain content for genes by use of hippocampus PEPs. Hippocampus preferentially expressed promoter (PEP) locations are shown as red triangles. Locations of predicted protein domains are shown as colored blocks (note that domains spanning more than exons are extended over the intron(s). In all of these cases, at least one domain is upstream of the PEP, which means that this domain is not included in the isoform expressed in hippocampus. Known cDNA locations are shown *below*: transcription is right-to-left. *BAI1* is a membrane protein whose N-terminal domain is extracellular, with a transmembrane region just downstream from the GPS domain (data not shown). The extracellular part can be cleaved off at the GPS domain, releasing a tumor-suppressing peptide; however, the hippocampus PEP is just downstream from the GPS domain, presumably giving a *BAI1* variant that is attached to the membrane but without the extracellular domains, which lack the tumor-suppression capability (Kaur et al. 2005). Similarly, the PEPs in *Myo10* confirm a previous study showing the neuronal expression of an isoform lacking the Myosin head domain (Sousa et al. 2006). In *Pclo*, the zf-piccolo domain cannot be included when using the hippocampus PEP.

Discussion

We are on the verge of a new era, in which sequencing technology can be used to infer biological function on a comprehensive scale, and the power of sequencing centers will be available to normal laboratories. Here, we have modified the CAGE protocol for the 454 Life Science instrument and demonstrate the usefulness of deep sequencing to discover new promoters in complex tissues.

We have identified a large number of core promoters that are preferentially used within hippocampus. Our results indicate that of the tissues we tested, the hippocampus has the largest number of such promoters, closely followed by cerebellum. These results may be due to two different factors: cell type diversity of the tissue and sequencing depth. The cerebellum is one of the least complex among brain tissues, while the hippocampus and the cortex tissues are very complex, with a plethora of different cell types. This is also shown in Figure 6, where it is evident that small, distinct cell populations within the hippocampus express a given gene, while most other cells do not. The methods we use to measure transcription cannot quantitatively measure the diversity of cells within a tissue and consequently neither the transcription dynamics within single cells. This strongly motivates further developments to assess the expression rates of genes within smaller cell populations. Such approaches in combination with the in situ data

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

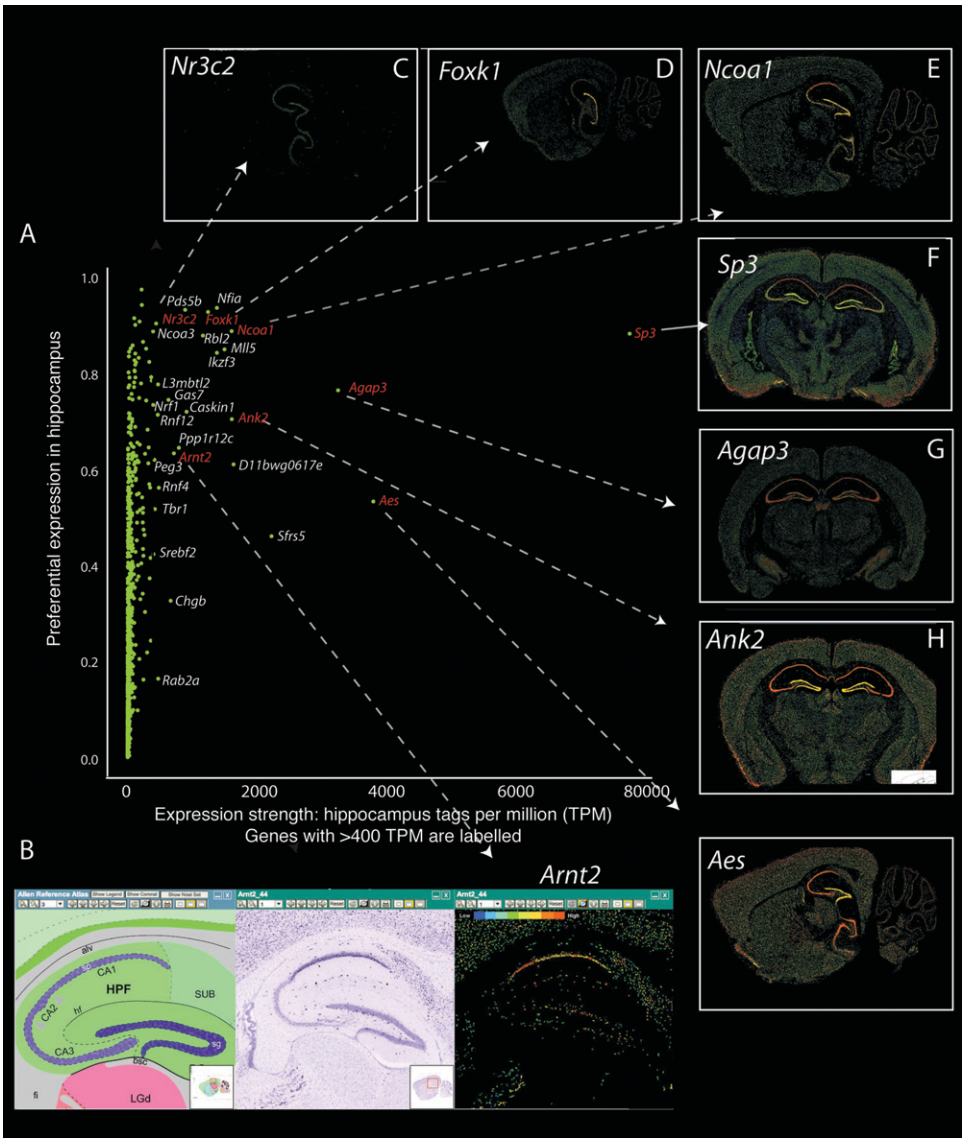


Figure 5. Transcription factor genes with preferential expression in hippocampus. (*A*) The relation between expression strength (number of hippocampus CAGE tags/million) vs. the ''tissue specificity'' (fraction of hippocampus tags vs. all brain tags, normalized for library size) of all known transcription factor genes. Note that only a few transcription factor genes are both highly expressed in and highly specific for hippocampus. For example, both in situ hybridization and CAGE data show that the *Nr3c2* gene (the mineralocorticoid receptor) is not highly expressed in brain, but almost exclusively expressed in hippocampus, while the *Aes* gene is expressed in the whole brain with a preference for hippocampus. (*C-H* and *Aes*, *right*) In situ images for some of these factors. (*B*) In situ hybridization images (Lein et al. 2007) showing the hippocampus expression of the *Arnt2* gene. The three images show the Allen reference atlas, the ISH in situ image, and corresponding expression heatmaps.

from the Allen Brain Atlas may result in a new molecular taxonomy of the different types of hippocampal cells providing a framework for a complete description of the components of hippocampal cell's network (Gray et al. 2004; Ma 2006; Sugino et al. 2006).

The subset of novel hippocampus promoters that are not overlapping known transcripts could indicate noncoding RNAs that have as yet not been sampled in mouse; Mercer et al. (2008) showed the existence of other long noncoding RNAs that are expressed primarily in brain. On the other hand, we also find that many of the "novel" promoters, in fact, are new upstream promoters for known genes.

However, most of these novel promoters fall within known genes, and we find that many genes have different core promoters that are used preferentially by different brain tissues, which may give partially different RNAs and protein products. In extension, identification of cell-type-specific alternative promoters for genes encoding for proteins responsible for neuronal and synaptic activity (channels, receptors, etc.) may provide increased specificity for drug treatments for epilepsy and other hippocampal-related neuropsychiatric disorders. Although this approach may seem far from our current technology, the use of antigene RNAs (agRNA) or peptide nucleic acids (agPNA) as well as of Locked Nucleic Acids (LNA) that target specific promoters has been proposed and demonstrated in vitro (Janowski et al. 2005a,b). This is particularly relevant since new promising strategies for delivery of nucleic-acid-based modifiers of gene expression into the brain have been

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

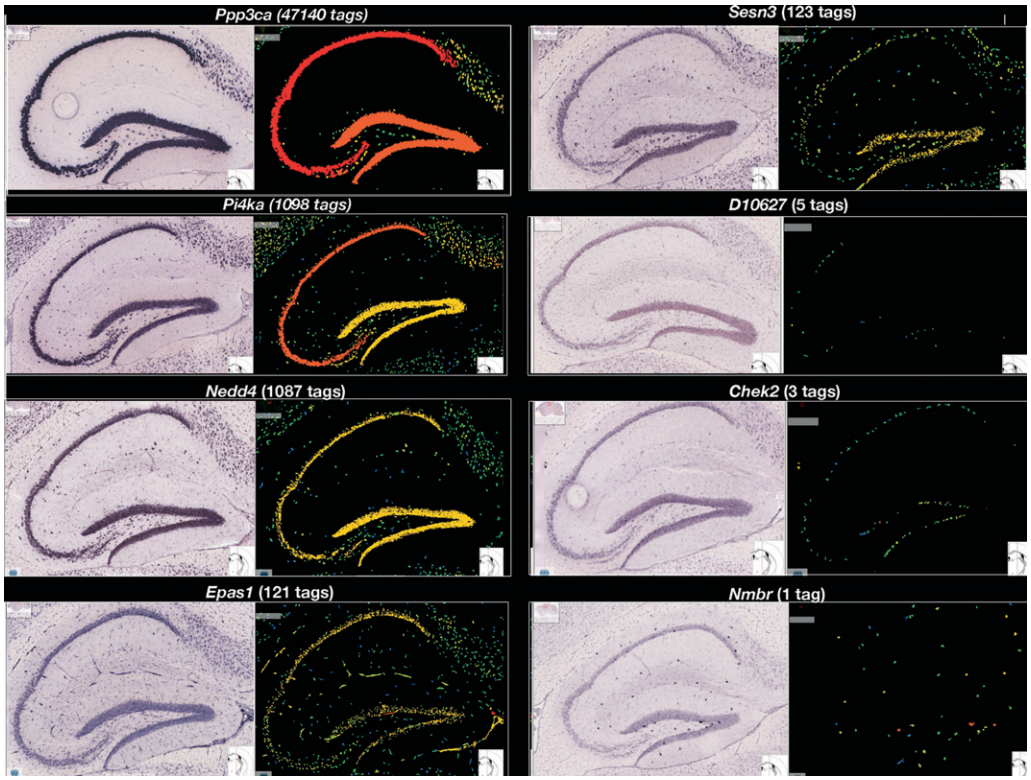


Figure 6. CAGE identifies promoter activity from small subpopulations of hippocampal cells. Examples of correspondence between CAGE tags and signal detected by in situ hybridization, ordered from relatively high expression (from the *top left* quadrant), expressed as the number of CAGE tags from hippocampus mapping to the gene, to low expression (the *lowest right* quadrant). On the *left* the original in situ signal is shown; on the *right* the in situ hybridization signal is quantified with pseudo-colors where red corresponds to high expression. In situ hybridization images were obtained from the Allen Brain Institute (Lein et al. 2007). Notice that the signal or tags corresponding to less than $5/1.4 \times 10^6$ mapped tags correspond to RNAs that are expressed only in a specific subset of cells.

recently proven (Kumar et al. 2007). To this end, the data set we present here, enabled by DeepCAGE, is to date the most comprehensive brain-centric promoter-exploration resource.

Methods

Preparation and sequencing of CAGE libraries

The preparation of the CAGE library is adapted from Shiraki et al. (2003) and Kodzius et al. (2006), to work with the 454 Life Sciences sequencer. The schema is represented in Figure 1. A detailed protocol of the CAGE library preparation, starting from trizol-extracted RNAs, is available in the Supplemental Material.

Once the CAGE library is prepared, we test various ratios of beads to CAGE library ratios, using usually an excess of DNA over beads (1:4 to 1:16 ratio beads:DNA) in the 454 GS20 protocol. During the calibration of the instrument, small-scale runs (1/8 of small kit runs) are used to calibrate the best DNA/beads ratio, followed by one or more runs of 454 large-scale sequencing kits (further details at http://www.454.com/).

In silico mapping of CAGE tags

Sequenced tags CAGE tags were mapped to mouse chromosomes and the mitochondrial genome (Genome build: mm8) using the BLAST/Vmatch alignment programs, and the longest full-matched (meaning no mismatches in the middle) positions were selected. These tags were referred to as "single-mapped" tags. Tags that map to multiple locations on the genome (with the same length) were called "multi-mapped," and tags that did not map (mapped <18 bp long) were called "unmapped." These multi-mapped and unmapped tags were passed to the rescue stage to increase the number of "single-mapped" tags (see Supplemental Material), since many promoters share identical subsequences (Faulkner et al. 2008). Rescued tags were incorporated into the single-mapped tag collection, and other tags were discarded. In the rest of the analysis, we use only the single-mapped tags; note that the same mapping procedure was applied to all CAGE libraries in the study.

Mouse hippocampus RNA preparation and 5′-RACE PCR validation of target intergenic core promoters

Adult C57/Bl wild-type mice (n = 5) were sacrificed by CO_2 inhalation, and hippocampal regions were rapidly dissected and snap-frozen in liquid nitrogen. Total RNA was extracted with TRIzol reagent (Invitrogen) following the manufacturer's protocol; the RNA sample was treated with DNase (Ambion), aliquoted in RNase free LoBind tubes (Eppendorf), and stored at −80°C.

RACE-ready cDNA was obtained with the Generacer kit (Invitrogen) following the manufacturer's protocol with no modifications starting from 5 μg of hippocampus total RNA. 5′-RACE was carried out using Platinum Taq DNA polymerase High Fidelity (Invitrogen); each PCR product was cloned in a TOPO TA vector (Invitrogen) and transformed in OneShot Top10 chemically competent *Escherichia coli* cells. Five colonies from each plate were selected for growth, DNA extraction (DNA Mini Kit; QIAGEN), and sequencing.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Oligonucleotide primers for the validation of intergenic core promoters were hand-designed according to guidelines from the Generacer kit manual and checked with PerlPrimer for possible primer-dimer formation. Primers used to validate the core promoters are shown in Supplemental Table S1.

Generation of tag clusters

A tag cluster (TC) was defined as the maximum set of tags where all 5′-ends are <20 bp from the closest neighbor, and on the same strand. We chose 20 bp because it is approximately the length of a CAGE tag, and thus we know with certainty that the transcript starting at the tag's 5′-end at least spans this region. This is the same definition as used in Carninci et al. (2006).

Exploration of tissue preferences of tag clusters

We first normalized the expression of each TC to TPMs for each tissue:

$$n_{c,i} = r_{c,i} \times 10^6 / r_i$$

$$n_c = \sum_i n_{c,i}$$

where $r_{c,j}$ is the number of non-normalized tags in cluster c for tissue i, r_i is the total number of non-normalized tags for tissue i, and n_c is the total TPM for the cluster. Only TCs with n_c >30 TPM were considered. We then normalized the total expression of each such tag cluster to sum to 1:

$$t_{c,i} = n_{c,i} / n_c$$

where $t_{c,i}$ is the normalized contribution of tissue i to cluster c. We then hierarchically clustered the set of promoters in terms of expression in each tissue (the $t_{c,i}$ values) using Euclidian distance measure and complete linkage as the clustering method (the defaults of the dist() and hclust() functions in R, respectively). The reordering of columns (tissues) and rows (tag clusters) was visualized using the heatmap.2() function in the gplots R package.

Generation of preferentially expressed promoters (PEPs)

To call a TC c preferentially expressed in a tissue i, we considered:

1. $t_{c,i}$ values as defined above. We required one such value to be >0.5, since with this cutoff, a TC will be preferentially expressed only in one tissue.
2. The assessment whether this over- or underrepresentation was significant, that is, unlikely to have arisen from random sampling from the underlying tags. This can be expressed as a binomial overrepresentation test. We required that the TC in question presented a P-value <0.05 in a one-tailed binomial overrepresentation test (R function: binom.test).
3. We only assessed core promoters n_c > 30 TPMs. The tag number constraint is not strictly necessary; we introduced the additional constraint to reduce the number of statistical tests (as tests with few tags will always be insignificant) and to focus on strong promoters.

Mapping PEPs to genes and introns

PEPs were considered to belong to a gene if they had at least one tag on the same strand within the boundary of its transcript (using the RIKEN cDNA database) including a 50-bp slack at the 5′-end of the gene. If a PEP had no such overlap, it was considered intergenic. PEPs belonging to genes were further divided into exonic if the PEP overlaps with an exon, or otherwise as intronic.

Domain annotation and PEPs

Domains were annotated using RIKEN cDNA annotation (corresponding to Interpro domain locations). To determine whether transcription initiation at the hippocampus PEPs changed the domain product, we used the gene mappings from above. Then we checked whether any domain in a gene containing a hippocampus PEP was upstream of this PEP and downstream from the annotated transcription start site. Usage of this PEP would result in the domain being lost and consequently in a different protein product.

TFBS overrepresentation analysis

We searched all sequence sets with the JASPAR matrices (Vlieghe et al. 2006) using the ASAP tool (Marstrand et al. 2008) with the following setting: uniform background model, a pseudo-count of 1, and threshold value of 0.7 relative to the matrix-specific scoring range. For all matrices, we calculated a P-value representing the overrepresentation using the binomial test as described in van Helden et al. (1998). For the tables, the P-value threshold is <0.01. As a background set, we used all core promoters with more than 30 TPM.

In situ comparison

For comparing expression and tissue preference of transcription factors between CAGE and in situ experiments available from the Allen Brain Atlas (Lein et al. 2007), we calculated the hippocampus strength versus tissue preference T for each transcription factor gene (using the RIKEN TF database [Kanamori et al. 2004]):

$$T = \frac{\sum_c n_{c,hippocampus}}{\sum_c \sum_i n_{c,i}}$$

where we sum over all TCs c that are within the boundary of the gene of interest, and all the brain tissues i. $n_{c,i}$ is the TPM count for respective TC and tissue. This was then visually compared with corresponding in situ images, downloaded from http://www.brain-map.org/.

Acknowledgments

P.C. and Y.H. are supported by the National Project on Protein Structural and Functional Analysis from MEXT and the National Project on Genome Network Analysis and the RIKEN Genome Exploration Research Project from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government. P.C. is also supported by a grant by the EU 6th Framework Program (NFG project). Authors affiliated with the Bioinformatics Centre are supported by a grant from the Novo Nordisk Foundation. The European Research Council has provided financial support to A.S. under the EU 7th Framework Programme (FP7/2007-2013)/ERC grant agreement 204135. S.G. is supported by a career development grant from "The Giovanni Armenise-Harvard Foundation." We thank Susan Sunkin for help with in situ images and scientific discussion, Kazuho Ikeo and Toshitsugi Okayama for developing and sharing the CAGE mapping procedures before publication, and Akira Hasegawa for technical support.

References

Bird, C.M. and Burgess, N. 2008. The hippocampus and memory: Insights from spatial processing. *Nat. Rev. Neurosci.* **9:** 182–194.

Carninci, P., Shiraki, T., Mizuno, Y., Muramatsu, M., and Hayashizaki, Y. 2002. Extra-long first-strand cDNA synthesis. *Biotechniques* **32:** 984–985.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. 2005. The transcriptional landscape of the mammalian genome. *Science* **309:** 1559–1563.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38:** 626–635.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95:** 14863–14868.

The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816.

Faulkner, G.J., Forrest, A.R., Chalk, A.M., Schroder, K., Hayashizaki, Y., Carninci, P., Hume, D.A., and Grimmond, S.M. 2008. A rescue strategy for multimapping short sequence tags refines surveys of transcriptional activity by CAGE. *Genomics* **91:** 281–288.

Gray, P.A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D.I., Tsung, E.F., Cai, Z., et al. 2004. Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306:** 2255–2257.

Gustincich, S., Sandelin, A., Plessy, C., Katayama, S., Simone, R., Lazarevic, D., Hayashizaki, Y., and Carninci, P. 2006. The complexity of the mammalian transcriptome. *J. Physiol.* **575:** 321–332.

Jackson, D.A., Pombo, A., and Iborra, F. 2000. The balance sheet for transcription: An analysis of nuclear RNA metabolism in mammalian cells. *FASEB J.* **14:** 242–254.

Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Hardy, D., Shames, D.S., Minna, J.D., and Corey, D.R. 2005a. Inhibiting gene expression at transcription start sites in chromosomal DNA with antigene RNAs. *Nat. Chem. Biol.* **1:** 216–222.

Janowski, B.A., Kaihatsu, K., Huffman, K.E., Schwartz, J.C., Ram, R., Hardy, D., Mendelson, C.R., and Corey, D.R. 2005b. Inhibiting transcription of chromosomal DNA with antigene peptide nucleic acids. *Nat. Chem. Biol.* **1:** 210–215.

Kanamori, M., Konno, H., Osato, N., Kawai, J., Hayashizaki, Y., and Suzuki, H. 2004. A genome-wide and nonredundant mouse transcription factor database. *Biochem. Biophys. Res. Commun.* **322:** 787–793.

Kaur, B., Brat, D.J., Devi, N.S., and Van Meir, E.G. 2005. Vasculostatin, a proteolytic fragment of brain angiogenesis inhibitor 1, is an antiangiogenic and antitumorigenic factor. *Oncogene* **24:** 3632–3642.

Kawaji, H., Frith, M.C., Katayama, S., Sandelin, A., Kai, C., Kawai, J., Carninci, P., and Hayashizaki, Y. 2006. Dynamic usage of transcription start sites within core promoters. *Genome Biol.* **7:** R118. doi: 10.1186/gb-2006-7-12-r118.

Kim, E., Naisbitt, S., Hsueh, Y.P., Rao, A., Rothschild, A., Craig, A.M., and Sheng, M. 1997. GKAP, a novel synaptic protein that interacts with the guanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules. *J. Cell Biol.* **136:** 669–678.

Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., et al. 2006. CAGE: Cap analysis of gene expression. *Nat. Methods* **3:** 211–222.

Kumar, P., Wu, H., McBride, J.L., Jung, K.E., Kim, M.H., Davidson, B.L., Lee, S.K., Shankar, P., and Manjunath, N. 2007. Transvascular delivery of small interfering RNA to the central nervous system. *Nature* **448:** 39–43.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445:** 168–176.

Ma, Q. 2006. Transcriptional regulation of neuronal phenotype in mammals. *J. Physiol.* **575:** 379–387.

Maccaferri, G. and Lacaille, J.C. 2003. Interneuron diversity series: Hippocampal interneuron classifications–making things as simple as possible, not simpler. *Trends Neurosci.* **26:** 564–571.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.

Marstrand, T.T., Frellsen, J., Moltke, I., Thiim, M., Valen, E., Retelska, D., and Krogh, A. 2008. ASAP: A framework for overrepresentation statistics for transcription factor binding sites. *PLoS One* **3:** e1623. doi: 10.1371/journal.pone.0001623.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. 2008. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci.* **105:** 716–721.

Muller, F., Demeny, M.A., and Tora, L. 2007. New problems in RNA polymerase II transcription initiation: Matching the diversity of core promoters with a variety of promoter recognition factors. *J. Biol. Chem.* **282:** 14685–14689.

Nakazawa, K., McHugh, T.J., Wilson, M.A., and Tonegawa, S. 2004. NMDA receptors, place cells and hippocampal spatial memory. *Nat. Rev. Neurosci.* **5:** 361–372.

Ng, P., Wei, C.L., Sung, W.K., Chiu, K.P., Lipovich, L., Ang, C.C., Gupta, S., Shahab, A., Ridwan, A., Wong, C.H., et al. 2005. Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nat. Methods* **2:** 105–111.

Parra, P., Gulyas, A.I., and Miles, R. 1998. How many subtypes of inhibitory cells in the hippocampus? *Neuron* **20:** 983–993.

Paxinos, G. 2004. *The rat nervous system*. Elsevier Academic Press, San Diego, CA.

Romorini, S., Piccoli, G., Jiang, M., Grossano, P., Tonna, N., Passafaro, M., Zhang, M., and Sala, C. 2004. A functional role of postsynaptic density-95-guanylate kinase-associated protein complex in regulating Shank assembly and stability to synapses. *J. Neurosci.* **24:** 9391–9404.

Sandelin, A. and Wasserman, W.W. 2004. Constrained binding site diversity within families of transcription factors enhances pattern discovery bioinformatics. *J. Mol. Biol.* **338:** 207–215.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. 2007. Mammalian RNA polymerase II core promoters: Insights from genome-wide studies. *Nat. Rev. Genet.* **8:** 424–436.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci.* **100:** 15776–15781.

Sousa, A.D., Berg, J.S., Robertson, B.W., Meeker, R.B., and Cheney, R.E. 2006. Myo10 in brain: Developmental regulation, identification of a headless isoform and dynamics in neurons. *J. Cell Sci.* **119:** 184–194.

Sugino, K., Hempel, C.M., Miller, M.N., Hattox, A.M., Shapiro, P., Wu, C., Huang, Z.J., and Nelson, S.B. 2006. Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat. Neurosci.* **9:** 99–107.

Tonegawa, S., Nakazawa, K., and Wilson, M.A. 2003. Genetic neuroscience of mammalian learning and memory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358:** 787–795.

Tsien, J.Z., Huerta, P.T., and Tonegawa, S. 1996. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell* **87:** 1327–1338.

van Helden, J., Andre, B., and Collado-Vides, J. 1998. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.* **281:** 827–842.

Vlieghe, D., Sandelin, A., De Bleser, P.J., Vleminckx, K., Wasserman, W.W., van Roy, F., and Lenhard, B. 2006. A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res.* **34:** D95–D97.

Wasserman, W.W. and Sandelin, A. 2004. Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* **5:** 276–287.

Watanabe, D., Inokawa, H., Hashimoto, K., Suzuki, N., Kano, M., Shigemoto, R., Hirano, T., Toyama, K., Kaneko, S., Yokoi, M., et al. 1998. Ablation of cerebellar Golgi cells disrupts synaptic integration involving GABA inhibition and NMDA receptor activation in motor coordination. *Cell* **95:** 17–27.

Zhao, C., Deng, W., and Gage, F.H. 2008. Mechanisms and functional implications of adult neurogenesis. *Cell* **132:** 645–660.

Received August 8, 2008; accepted in revised form December 3, 2008.