Identification of long-range regulatory elements in the protocadherin-α gene cluster

Scott Ribich*, Bosiljka Tasic†, and Tom Maniatis‡

Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138

Contributed by Tom Maniatis, November 3, 2006 (sent for review October 2, 2006)

The clustered protocadherins (Pcdh) are encoded by three closely linked gene clusters (*Pcdh-α*, *-β*, and *-γ*) that span nearly 1 million base pairs of DNA. The *Pcdh-α* gene cluster encodes a family of 14 distinct cadherin-like cell surface proteins that are expressed in neurons and are present at synaptic junctions. Individual *Pcdh-α* mRNAs are assembled from one of 14 ''variable'' (V) exons and three ''constant'' exons in a process that involves both differential promoter activation and alternative pre-mRNA splicing. In individual neurons, only one (and rarely two) of the *Pcdh α*1–12 promoters is independently and randomly activated on each chromosome. Thus, in most cells, this unusual form of monoallelic expression leads to the expression of two different *Pcdh-α* 1–12 V exons, one from each chromosome. The two remaining V exons in the cluster (*Pcdh-α*C1 and αC2) are expressed biallelically in every neuron. The mechanisms that underlie promoter choice and monoallelic expression in the *Pcdh-α* gene cluster are not understood. Here we report the identification of two long-range cis-regulatory elements in the *Pcdh-α* gene cluster, HS5–1 and HS7. We show that HS5–1 is required for maximal levels of expression from the *Pcdh α*1–12 and αC1 promoters, but not the *Pcdh-α*C2 promoter. The nearly cluster-wide requirement of the HS5–1 element is consistent with the possibility that the monoallelic expression of *Pcdh-α* V exons is a consequence of competition between individual V exon promoters for the two regulatory elements.

gene regulation | transcription enhancer | monoallelic expression

A fundamental problem in neurobiology is how individual neurons of a given cell type acquire unique cellular identities. For instance, individual neurons of the same type can engage in highly specific synaptic interactions, and they can express unique subsets of genes. An excellent example of this type of neuronal diversification is provided by individual olfactory neurons, which express only one of >1,000 olfactory receptor genes and establish connections based on the receptor expressed. This is accomplished by a complex process involving stochastic promoter choice, monoallelic expression, and a feedback mechanism that stabilizes promoter choice (1). Another example is the highly conserved clustered Pcdh genes (*Pcdhs-α*, *-β*, and *-γ*), which encode diverse cadherin-like cell surface molecules that are present at synaptic junctions (2–7). Single-cell analyses reveal that individual Purkinje cells express distinct subsets of *Pcdh* mRNAs (2, 8, 9). If different Pcdh proteins function cooperatively at the synapse, they could provide an extraordinarily diverse array of distinct synaptic tags and thus may play an important role in establishing the identities of individual neurons.

The organization of the *Pcdh-α* gene cluster results in the generation of 14 different Pcdh proteins in the mouse. Multiple first exons [variable region (V) exons] are arranged sequentially and each encode six cadherin-like ectodomains, a transmembrane domain, and a small portion of the cytoplasmic domain. They are followed by three constant region exons, which encode the remainder of the cytoplasmic domain (Fig. 1*A*). *Pcdh-α* mRNAs are generated by splicing a single V exon to the three constant region exons. The two *Pcdh-α* V exons located closest to the constant region exons, which are named αC1 and αC2, are more related to each other and to the three similarly positioned V exons in the *Pcdh-γ* cluster than to the *Pcdh α*1–12 V exons (Fig. 1*A*) (3).

Each V exon is preceded by a promoter, and all *Pcdh-α* promoters except the αC2 promoter contain a conserved sequence element (CSE) (6, 10). The CSE is 22-nt long and contains an invariant tetranucleotide CGCT (CGCT box) that is required for promoter activity. Activation of a particular promoter leads to transcription of a long premRNA containing all of the downstream V exons and the three constant exons. Generation of the mature *Pcdh-α* mRNA involves splicing the V exon proximal to the 5′ end of the pre-mRNA to the first constant region exon (10, 11).

Although *Pcdh-α* mRNAs are expressed throughout the nervous system, individual *Pcdh-α* isoforms do not appear to be expressed in specific brain layers or brain nuclei (2). In fact, RNA *in situ* analyses have shown that single neurons of the same cell type (olfactory bulb periglomerular cells in one case and cerebellar Purkinje cells in another) express different *Pcdh-α* mRNAs (2, 8). Moreover, single-cell PCR analyses of Purkinje cells show that individual neurons express apparently random sets of *Pcdh α*1–12 mRNAs, with the majority of cells expressing only two (8). Remarkably, the same *Pcdh α*1–12 isoform can be expressed from both chromosomes but rarely is, suggesting that the choice of active promoters is random within each chromosome and independent between chromosomes. The mechanisms of this unusual example of monoallelic expression are not known. By contrast, expression of *Pcdh-α*C1 and αC2 V exons is biallelic, and both mRNAs appear to be expressed in every neuron (9). Thus, the *Pcdh α*1–12 and *Pcdh-α*C1 and αC2 promoters appear to be regulated by distinct mechanisms. The apparently random and predominately monoallelic expression of individual *Pcdh α*1–12 mRNAs suggests that promoter identity and cell-specific transacting factors cannot be the sole determinants of their promoter activation.

In an effort to understand the mechanisms of monoallelic expression of the *Pcdh-α* genes, we have identified long-range cis-regulatory DNA elements in the *Pcdh-α* gene cluster. We show that two of these DNA sequences display enhancer activity in reporter assays, and one of them is necessary for high-level expression from all *Pcdh α*1–12 promoters, as well as the *Pcdh-α*C1 promoter. By contrast, the element is not required for expression from the *Pcdh-α*C2 promoter.

Author contributions: S.R. and B.T. designed research; S.R. and B.T. performed research; S.R., B.T., and T.M. analyzed data; and S.R., B.T., and T.M. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: CIS, conserved intergenic/intronic sequence; Pcdh, protocadherin; HS, hypersensitivity site; V, variable region; SV40, simian virus 40; E*n*, embryonic day *n*.

*Present address: Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115.

†Present address: Department of Biological Sciences, Stanford University, Stanford, CA 94305.

‡To whom correspondence should be addressed. E-mail: maniatis@mcb.harvard.edu.

This article contains supporting information (SI) online at www.pnas.org/cgi/content/full/0609445104/DC1.

© 2006 by The National Academy of Sciences of the USA

BIOCHEMISTRY

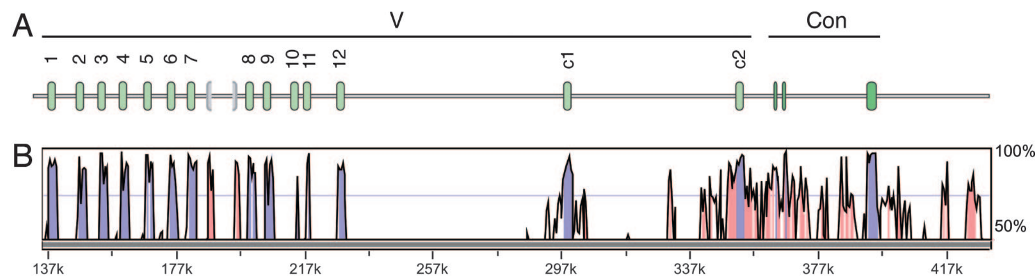


Fig. 1. CISs in the mouse and human *Pcdh-α* gene clusters. (*A*) Genomic organization of the mouse *Pcdh-α* gene cluster. The *Pcdh-α* gene cluster spans ≈350 kb of mouse chromosome 18. Individual V exons (denoted by V, light-green ovals) are individually spliced to the three constant-region exons (denoted by Con, dark-green ovals). Gray ovals indicate relics of a *Pcdh-α* V exon. (*B*) Sequence comparison between mouse and human for portions of the *Pcdh-α* cluster using VISTA. The *y* axis indicates percent mouse/human sequence identity. The *x* axis indicates position along the cluster as measured in kilobase pairs. Blue shading denotes exonic sequences, and red shading denotes CISs that are >75% identical between mouse and human DNA. A total of 65 CISs >100 bp with >75% identity were identified.

Results

Identification of Putative *cis*-Regulatory Elements in the *Pcdh-α* Gene Cluster. Regulatory elements such as enhancers and locus control regions are highly conserved between different mammalian species (12–14). To identify intergenic regions containing putative DNA elements that regulate *Pcdh-α* expression, we compared the genomic DNA sequences of the mouse and human *Pcdh-α* clusters. Vista analysis (15, 16) was performed on mouse and human *Pcdh-α* genomic sequences, including 138 kb upstream and 77 kb downstream of the mouse gene cluster and 80 kb upstream and 82 kb downstream of the human gene cluster. The borders for this analysis were determined by the location of the genes immediately upstream and downstream from the *Pcdh-α* gene cluster. The upstream genes correspond to *HARSL* for the mouse and *Zmat2* for human, and the downstream gene corresponds to *Pcdh-β1* in both mouse and human. This analysis revealed 65 conserved intergenic/intronic sequences (CISs) containing >100 bp with >75% identity [Fig. 1 and supporting information (SI) Table 1]. Six CISs were located between the α7 and α8 V exons and correspond to a relic mouse *Pcdh-α* V exon that was created by the insertion of an intracisternal A-particle retrotransposon present in laboratory mouse strains (17). Another conserved DNA sequence was identified in the promoter of α1, two more within the promoter of αC2, and one in the promoter of *Pcdh-β1* (the first exon in the downstream *Pcdh-β* cluster). The remainder of the conserved sequences are distributed in the 3′ end of the cluster; 43 are located in the introns between αC1 and the third constant region exon; the remaining 12 CISs are located downstream of the cluster. No CISs were found immediately upstream of the cluster or in any of the introns between α1 and αC1.

Regulatory DNA sequences can be identified as DNase I hypersensitive sites (HSs), which are thought to be a consequence of the remodeling or removal of nucleosomes by bound proteins, leaving the surrounding DNA accessible for degradation by DNase I (18). Given the complexity of brain tissue and that individual neurons can express different sets of *Pcdh-α* isoforms (2, 8), we made use of cell lines as homogenous sources of nuclei for DNase I hypersensitivity assays. We used the mouse neuroblastoma cell lines Neuro-2a and CAD, which express defined subsets of *Pcdh-α*, and the mouse leukemia cell line BB88, which does not express *Pcdh-α* (SI Fig. 5). We conducted DNase I hypersensitivity assays on the 55 CISs in Neuro-2a cells, covering ≈120 kb of the *Pcdh-α* cluster. We identified 15 HSs distributed over 77 kb of the 3′ end of the cluster and designated them HS15 through HS1, with HS15 being the most 5′, and HS1 being the most 3′ (Fig. 2 and SI Figs. 6–9, and SI Table 2). All HSs except HS9 were found within 200 bp of a CIS discovered by Vista analysis. HSs 15, 14, and 13 were all found within 2.2 kb of the αC2 translational start site and likely correspond to its promoter and associated regulatory elements. HSs 12, 11, and 10 are in the intron between the αC2 V exon and the first constant region exon. HSs 9, 8, and 7 are in the intron between the second and third constant region exons. HS6 is located 3.5 kb down-

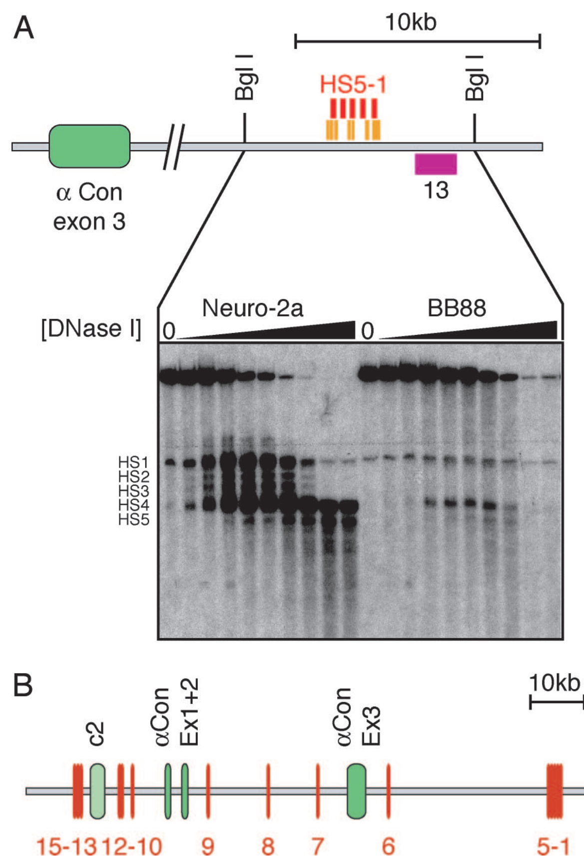


Fig. 2. Identification of putative cis-regulatory elements by DNase I hypersensitivity. (*A*) (*Upper*) a diagram of the region downstream of the third constant-region exon indicating the positions of restriction enzyme sites, CIS (orange boxes), and Southern blot probes (purple boxes). Red boxes indicate identified DNase I HSs. (*Lower*) DNase I hypersensitivity assays on the region 30 kb downstream of the third constant region exon flanked by BglI sites on Neuro-2a and BB88 nuclei. Triangles indicate increasing concentrations of DNase I. Southern probe no. 13 was used. Five HSs were identified (HS5–1), which are prominent in Neuro-2a but weak in BB88. (*B*) Schematic showing the location of the 15 identified HSs in Neuro-2a cells. All but HS12 were also present in CAD cells. Except for the weak HS1 and HS4, none were present in BB88 cells.

stream of the third constant region exon. HSs 5, 4, 3, 2, and 1 (hereafter referred to as HS5–1) are clustered ≈30 kb downstream of the third constant region exon. In BB88 cells, only HS1 and HS4 were identifiable and at a dramatically reduced intensity compared with Neuro-2a cells (Fig. 2 and SI Figs. 6–9).

To determine whether the hypersensitive sites identified in Neuro-2a cells are a general feature of *Pcdh-α*-expressing cells, we examined CAD cells, which express a different subset of *Pcdh-α* mRNAs. As shown in SI Table 2 and SI Figs. 10 and 11, all 15 hypersensitive sites identified in Neuro-2a cells were also present in CAD cells, with the exception of HS12, which was only weakly detected in Neuro-2a cells and absent in CAD cells.

HS7 and HS5–1 Are Tissue-Specific Enhancers. We assayed the 15 identified HSs for enhancer activity in transient transfection experiments. Individual HSs or groups of HSs were inserted downstream of the firefly luciferase gene driven by the simian virus 40 (SV40), *Pcdh-α*11, or *Pcdh-α*9 promoter. These constructs were transiently transfected into Neuro-2a or CAD cells, and relative luciferase activity was measured (Fig. 3*A*). The HS7 and HS5–1 sequences consistently enhanced transcription from each promoter at a level comparable to or greater than that observed with the SV40 enhancer, indicating that HS7 and HS5–1 can function as transcriptional enhancers. The remainder of the HSs (HS15, 14, 13, 12, 11, 10, 9, 8, and 6) failed to display any enhancer activity in these assays.

We found that both HS1 and HS7 contain CSEs similar to those found in all *Pcdh-α* promoters except the αC2 promoter (SI Fig. 12) (6). This finding is consistent with the possibility that HS1 and HS7 are regulatory elements, because cis-regulatory elements and the promoters they act upon often share binding sites for the same regulatory proteins (19–21). None of the other HSs contain a CSE-like sequence.

The reporter assays demonstrated that HS7 and HS5–1 can function as enhancers in transient transfections. To determine whether they can also function as such *in vivo*, we assayed their ability to drive reporter gene expression in transgenic mice. We cloned HS7 and HS5–1 downstream of an *hsp68* minimal promoter driving *lacZ* (Fig. 3 *B* and *C*). Upon integration into the genome, the *hsp68* minimal promoter is active only when in proximity of an enhancer (22). If HS7 and HS5–1 function as enhancers in the *Pcdh-α* locus, each should promote a consistent tissue-specific and developmentally regulated pattern of transgene expression in multiple transgenic founders (13, 23).

Endogenous *Pcdh-α* expression is first detected in the nervous system of the developing mouse embryo at embryonic day (E)12.5, peaks at postnatal day (P)0, and then moderately declines to a steady-state level that is maintained in the adult (2, 4, 24). We chose to analyze transgene expression at E14.5 and E15.5, when *Pcdh-α* expression is high, yet the embryo is still permeable to whole-mount β-galactosidase staining. At this stage, *Pcdh-α* expression is detected throughout the nervous system, with prominent expression in most areas of the developing CNS (2, 24–26). However, the expression is not limited to the CNS and is present in sensory and sympathetic ganglia, olfactory epithelium, retina, middle and inner ear, whisker follicles, and the mouth (24, 27, 28).

In all five transgenic lines containing the *hsp68-lacZ*-HS5–1 transgene, β-galactosidase staining was detected in the developing cerebral cortex and olfactory bulb. In some of the lines, staining was also present in other areas of the nervous system, including the olfactory epithelium, midbrain, developing cerebellum, spinal cord, whisker follicles, and mouth (Fig. 3*B*, SI Table 3, and SI Fig. 13). Only one line displayed aberrant expression, presumably because of position effect at the site of integration, likely because of the influence of local enhancers surrounding the insertion site (22, 23). Thus, the HS5–1 enhancer is capable of activating reporter gene expression in most

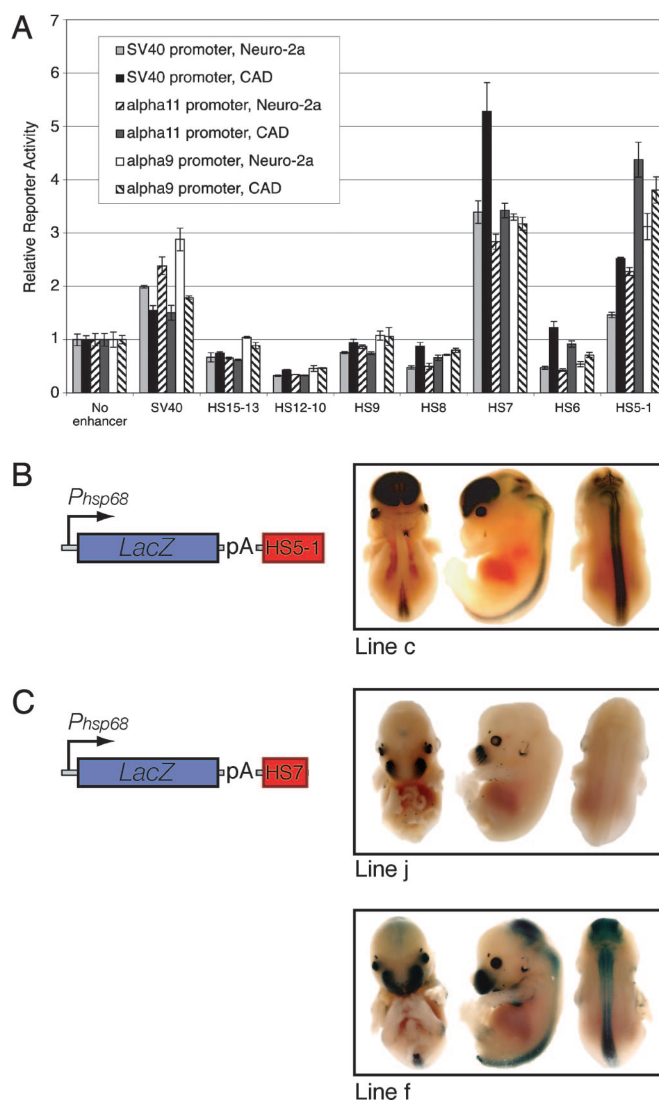


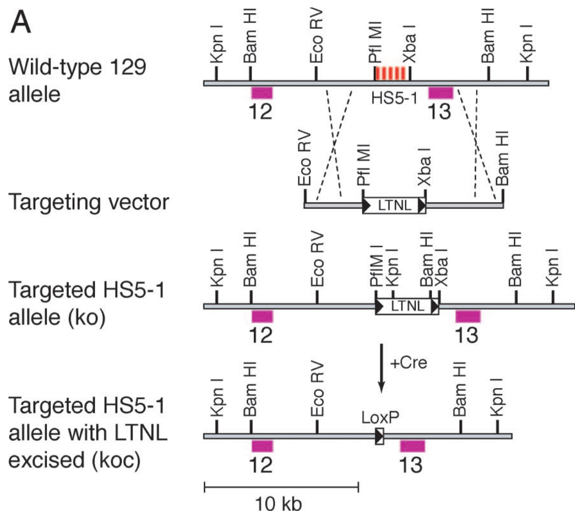
Fig. 3. HS7 and HS5–1 are tissue-specific enhancers. (*A*) Reporter assays for constructs containing HSs cloned downstream of the SV40, *Pcdh-α*11, or α9 promoter for both Neuro-2a and CAD cells. The *y* axis represents relative reporter activation compared with the construct containing only the SV40, α11, or α9 promoter. (*B*) A schematic representing the construct used to test HS5–1 for enhancer activity in transgenic mice and the expression in a representative transgenic line in mouse embryos at E14.5. pHsp68, minimal promoter from heat-shock protein 68; pA, poly(A) signal. (*C*) A schematic representing the construct used to test HS7 for enhancer activity in transgenic mice and the expression in two representative transgenic lines in mouse embryos at E14.5. Abbreviations are the same as in *B*.

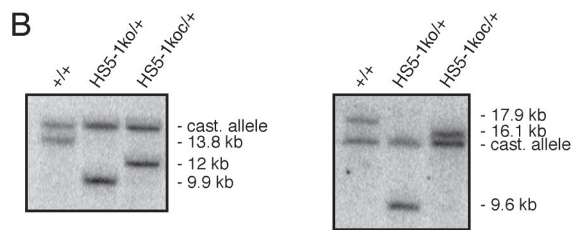
if not all areas that express *Pcdh-α*, but it does so most consistently in the olfactory bulb and the cortex.

In all but one of the 14 transgenic lines containing the *hsp68-lacZ*-HS7 transgene, β-galactosidase staining was prominently seen in the whisker follicles (Fig. 3*C*, SI Table 3, and SI Fig. 14). Most of these lines also displayed expression in eyebrow follicles, ears, and eyes. In seven of the lines, prominent staining was also seen in the spinal cord, and in five of these staining was present in the midbrain. Only three lines displayed staining in the cerebral cortex or olfactory bulb. We conclude that HS7 by itself is capable of activating reporter gene expression in most areas of *Pcdh-α* expression but does so most consistently in the whisker and eyebrow follicles, ears, and eyes.

As a negative control, we tested the *hsp68-lacZ* transgene

BIOCHEMISTRY





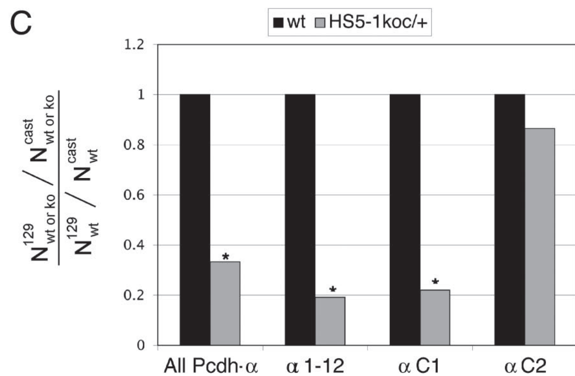


Fig. 4. HS5–1 are necessary for high-level expression of mRNAs containing any *Pcdh* α V exon (α1-αC1), except αC2. (*A*) A schematic of the strategy used to delete HS5–1 in mouse ES cells. (*Top* to *Bottom*) The genomic organization of the part of the *Pcdh-α* locus containing HS5–1; the targeting vector with the LNTL cassette; the *Pcdh-α* cluster after homologous recombination (HS5–1ko); the *Pcdh-α* cluster after Cre-mediated excision of the LTNL cassette (HS5–1koc). Purple rectangles represent Southern blot probes used in *B*. Red blocks denote HS5–1. LTNL is the selection cassette containing the thymidine kinase and neomycin resistance genes flanked by loxP sites (black triangles). (*B*) Verification of the proper recombination of the locus with the targeting vector and subsequent Cre-mediated excision of the selection cassette using Southern blotting. (*Left*) Verification of the proper recombination of the left arm using BamHI and the Southern probe no. 12. (*Right*) Verification of the proper recombination of the right arm using KpnI and the Southern probe no. 11. The targeting cassette recombined with the 129 chromosome, as indicated. In both cases, the *M. castaneus* allele has a different size because of strain-specific polymorphisms. (*C*) Comparison of expression of *Pcdh-α* mRNAs from differentiated wild-type and HS5–1koc/+ ES cells. Transcripts from all *Pcdh-α*, α1–12, αC1, or αC2 were amplified by RT-PCR and cloned, and individual clones were sequenced to determine the chromosome of origin. The ratio of 129 to *M. castaneus* clones was then normalized to 1 for wild-type (wt) cells to account for natural differences in mRNA expression between the two chromosomes. For identity and number of clones sequenced, see SI Table 5. *, $P < 0.001$. For αC2 analysis, $P > 0.2$. The *P* values were calculated by using the χ^2 test.

containing no enhancer. Of the five different transgenic pup founders analyzed, all displayed some position effect. However, none shared a common expression pattern, indicating that the intrinsic activity of the *hsp68* minimal promoter and the insertional specificity of the transgene are negligible (SI Fig. 15).

These results indicate that the two enhancers identified *in silico* and *in vitro* can function as tissue-specific and developmentally specific enhancers *in vivo*. Although both enhancers are capable of activating reporter gene expression in most if not all areas of *Pcdh-α* expression, HS5–1 shows preference for the CNS, especially the cortex and olfactory bulb, whereas HS7 shows preference for sensory organs, specifically whisker and eyebrow follicles, ears, and eyes.

HS5–1 Is Necessary for High-Level Expression of all *Pcdh-α* mRNAs, Except *Pcdh-α*C2. To test whether HS5–1 is necessary for expression of all *Pcdh-α* mRNAs, we deleted this DNA sequence from the endogenous locus in mouse ES cells. We performed the deletion in 129/Sv × *Musculus castaneus* F_1 ES cells (29) and analyzed the expression in heterozygous ES cells induced to differentiate into CNS-like neurons. The unmodified *M. castaneus* chromosome served as an internal control.

The HS5–1 targeting vector was prepared from 129/Sv genomic DNA. The HS5–1 region was replaced by a LTNL cassette, which consists of the neomycin phosphotransferase and thymidine kinase selection markers flanked by LoxP sites (30) (Fig. 4 *A* and *B*). As expected, the recombination occurred on the 129 chromosome (Fig. 4*B*), thereby generating the HS5–1 knockout allele (HS5–1ko). The LTNL cassette was subsequently excised by transient transfection of a Cre-producing plasmid, thereby generating the HS5–1koc allele. Proper targeting and Cre-mediated recombination were verified by both Southern blotting and PCR (Fig. 4 *A* and *B* and data not shown).

To compare *Pcdh-α* expression from the 129 and *M. castaneus* chromosomes, we identified sequence polymorphisms in each *Pcdh-α* exon (SI Table 4). To examine the effects of the deletion in neuronal cells, we used an *in vitro* differentiation protocol that generates a large percentage (>70%) of CNS-like postmitotic neurons (31). Postmitotic neuronal markers were identified in the *in vitro* differentiated cells by both RT-PCR and immunocytochemistry (SI Fig. 16). To determine the effect of the HS5–1 deletion on *Pcdh-α* gene expression, RNA was extracted from day 10 differentiated F_1 ES cells and RT-PCR analysis of the following *Pcdh-α* transcripts performed: constant region (all *Pcdh-α*), α1–12, αC1, and αC2. *Pcdh* α1–12 mRNAs were amplified by using a common forward primer that recognizes α1-α12 and a reverse primer in the constant region. *Pcdh* αC1 and αC2 mRNAs were each amplified by using specific forward primers and a constant-region reverse primer. The resulting products were cloned and a large number of individual clones sequenced to determine the chromosome of origin. By normalizing to the *M. castaneus* chromosome, it was possible to determine the effect of the HS5–1 deletion on the 129 chromosome.

Deletion of HS5–1 resulted in a 3-fold decrease ($P < 0.001$, χ^2 test) of overall *Pcdh-α* mRNA expression (Fig. 4*C* and SI Table 5). α1–12 and αC1 expression was reduced 5.2- and 4.5-fold (both $P < 0.001$, χ^2 test), respectively. Remarkably, *Pcdh-α*C2 expression was not significantly altered ($P > 0.2$). Likewise, no effect on *Pcdh-β* expression was observed (SI Table 5). We conclude that HS5–1 is necessary for wild-type levels of transcription for all *Pcdh-α* mRNAs (α1–12 and αC1), except for αC2.

Discussion

Here we report the identification and functional characterization of two long-range regulatory elements in the mouse *Pcdh-α* gene cluster, HS5–1 and HS7. These elements were identified through comparative genomics and DNase I hypersensitivity assays, and

they display enhancer activity in transient transfections and in transgenic mice. Both HS5–1 and HS7 contain a single CSE, an element found in all *Pcdh*-α promoters except the *Pcdh*-αC2 promoter. *In vivo* tests of HS7 and HS5–1 activities in transgenic mice revealed that they activate a reporter gene in tissues with endogenous *Pcdh*-α expression (the CNS and sensory organs) at an appropriate developmental stage, thereby demonstrating that HS5–1 and HS7 are tissue and developmental stage-specific enhancers. Although each enhancer is capable of activating expression in most, if not all, regions where *Pcdh*-α mRNAs are expressed, HS5–1 does so most consistently in the CNS (cortex and olfactory bulb), whereas HS7 preferentially activates the reporter in sensory organs (whisker and eyebrow follicles, ear and eye).

Deletion of HS5–1 by homologous recombination decreased the expression of *Pcdh* α1–12 and *Pcdh*-αC1 from the mutant chromosome in differentiated ES cells by 5.2- and 4.5- fold, respectively. Thus, HS5–1 is required for maximal levels of expression from all of the *Pcdh*-α promoters except *Pcdh*-αC2. The remaining transcriptional activity of the *Pcdh* α1–12 and *Pcdh*-αC1 promoters may be a result of their activation by HS7 and/or additional regulatory elements within the gene cluster. The additive effects of separate enhancers have been demonstrated for the mouse β-globin locus control regions (32). The lack of an effect of the HS5–1 deletion on *Pcdh*-αC2 expression could be because of a high intrinsic activity of the αC2 promoter alone, the presence of one or more *Pcdh*-αC2-specific enhancers in the gene cluster, and/or incompatibility of *Pcdh*-αC2 promoter and HS5–1 enhancer. It is also possible that HS7 compensates for the loss of HS5–1 more efficiently for the *Pcdh*-αC2 promoter than for other *Pcdh*-α promoters. At present, we are unable to distinguish between these possibilities. It is worth noting, however, that *Pcdh*-αC2 has a developmentally delayed onset of expression (4), and that the *Pcdh*-αC2 promoter does not contain a CSE (6). Thus, it appears that the mechanisms that lead to the activation of the *Pcdh*-αC2 promoter are distinct from those that regulate the other promoters in the *Pcdh*-α gene cluster.

Based on single-cell RT-PCR experiments, the activation of individual *Pcdh* α1–12 promoters appears to be stochastic and displays an unusual type of monoallelic expression (in most cells, only 1 of the 12 promoters is active on each chromosome) (8). Thus, random combinations of at least two of the *Pcdh* α1–12 proteins are expressed in individual neurons. By contrast, both *Pcdh*-αC1 and -αC2 are expressed in a biallelic manner (9). At present, it is not clear why the HS5–1 action results in biallelic expression in one case (*Pcdh*-αC1), whereas it results in random monoallelic expression in the case of *Pcdh* α1–12. At least two mechanisms have been proposed to account for other examples of random monoallelic expression. In one mechanism, a promoter or an essential regulatory element assembles a functional transcription complex with a low probability (for example, because of limiting levels of transcription components), thereby resulting in monoallelic expression (33, 34). In another mechanism, an essential regulatory element that is shared among two or more promoters is capable of interacting with only one promoter at a time. The most striking example of this mechanism occurs among the olfactory receptor gene family, where an enhancer element located on one chromosome can activate only one allele of >1,000 olfactory receptor genes located on the same or different chromosomes (1). Moreover, in this case, the enhancer-dependent selection of the active promoter is combined with inactivation of the enhancer element on the second chromosome by DNA methylation, leading to monoallelic expression of only a single allele of an individual olfactory receptor in one olfactory receptor neuron. Although it is possible that certain *Pcdh*-α promoters compete for HS7 and HS5–1 activity, imprinting or active allelic exclusion of HS5–1, HS7, or individual *Pcdh* α1–12 promoters is unlikely, because both chromosomes express at least one of *Pcdh* α1–12 mRNAs and sometimes even the same *Pcdh* α1–12 mRNA within a single cell.

Extraordinary diversity can be generated at the cell surface by expression of one or a few genes from a larger set of related genes in a mutually exclusive manner (1). In the case of the clustered Pcdhs, the random expression of the many individual isoforms in the *Pcdh*-α, -β, and -γ gene clusters could, in principle, combinatorially generate an extensive diversity of protein expression at the synapse. The studies reported here provide insights into the organization of the intergenic regulatory sequences required to generate this diversity.

Methods

Nucleic Acid Preparation. Plasmid DNA was isolated by using Qiagen (Valencia, CA) miniprep and maxiprep kits. Total RNA was isolated by using TRIzol Reagent (Invitrogen, Santa Clara, CA). BAC DNA was purchased from Children's Hospital Oakland Research Institute, then prepared using cesium-chloride method (35). Primers were synthesized by Operon (Huntsville, AL); see SI Table 6 for primer sequences.

Comparative Genomics. The mouse *Pcdh*-α locus sequence was compiled from sequences corresponding to GenBank accession nos. AC020967–69 and AC020971–74. The human *Pcdh*-α locus sequence was compiled from sequences corresponding to GenBank accession nos. AC005366, AC004776, AC005609, AC005618, AC005752, AC005754, AC008468, AC010223, AC025436, and AC074130. Repetitive elements were identified and masked by the RepeatMasker program (36). CIS (>100 bp and >75% identity) were identified by using the program Vista (15, 16).

DNase I Hypersensitivity Assays. The DNase I hypersensitivity assays were done as described (37). The complete protocol is available in SI *Supporting Text*.

Transient Reporter Assays. Luciferase assay constructs were based on the pGL3 series of plasmids from Promega (Madison, WI). The details on plasmid construction are available in SI *Supporting Text*. Neuro-2a and CAD cells at 80% confluence were transfected with Lipofectamine 2000 (Invitrogen) in a 96-well plate with 210 ng of plasmid DNA containing 37 fmol of the reporter plasmid to be tested, 10 ng of pRL-Tk (Promega), and any amount of pBluescript to supplement the total amount of DNA to 210 ng of total. Luciferase activity was assayed 24 h later by using the Dual-Glo System (Promega) and an Analyst AD luminometer (LJL Biosystems, Sunnyvale, CA).

Transgenic Mice. The phsp68-lacZ construct was obtained from Marcelo Nobrega (U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA). HS5–1 and HS7 were amplified from pGL3-HS5–1-SV40 and pGL3-HS7-SV40, respectively, by using High-Fidelity Polymerase (Roche, Indianapolis, IN). Primers used were HS5–1-F-N1 and HS5–1-R-N1 (for HS5–1) and HS7-F-N1 and HS7-R-N1 (for HS7). Both PCR products were ligated into the NotI site of phsp68-lacZ to create phsp68-lacZ-HS5–1 and phsp68-lacZ-HS7. The transgene was subsequently excised by digestion with SalI and purified by using QIAquick Gel Extraction kit (Qiagen). The *hsp68-lacZ*, *hsp68-lacZ*-HS5–1, or *hsp68-lacZ*-HS7 transgene DNAs were then individually injected by the Harvard Molecular and Cellular Biology Genome Manipulation Facility (Boston, MA) into fertilized oocytes. Embryos were either harvested at E14.5 or allowed to come to term to create founder lines. Embryos were stained for β-galactosidase expression for 12–24 h, as described (13).

Generation of HS5–1 Knockout ES Cells. The HS5–1 knockout cassette was created by subcloning the EcoR/V/BamHI fragment of BAC RP22-307G19 that contained HS5–1 into pBlue-

BIOCHEMISTRY

scriptSK(−). A HindIII portion of this DNA was then separately subcloned into pBSSK(−). HS5–1 was excised from this fragment with PflM/I/XbaI and replaced with a PflM/I/XbaI fragment containing the LTNL cassette (30). This HindIII fragment was then reinserted back into the original EcoRV/BamHI fragment to create the completed targeting vector. F_1 129/Sv × *M. castaneus* cells were obtained from Kevin Eggan (Harvard University, Boston, MA) (29). A full description of ES cell culturing and modification is available in SI *Supporting Text*.

Characterization of HS5–1koc/+ ES Cells. Wild-type and HS5–1koc/+ F_1 129/Sv × *M. castaneus* ES cells were differentiated as described (31). To test expression upon ES cell differentiation, total RNA was reverse-transcribed by using Random Decamers (Operon) and SuperScript III (Invitrogen), according to the manufacturer's instructions. The cDNA corresponding to 100 ng of total RNA was amplified by using High-Fidelity Polymerase (Roche) according to the manufacturer's instructions. To generate the *Pcdh-α* constant region, *Pcdh* α1–12, αC1, and αC2 RT-PCR products for sequencing, 29, 29, 34, and 32 PCR cycles were used, respectively. Primers used were MACon geno Ex3 and Cast-Acon-Ex3-R (*Pcdh-α* constant region), MAV_1643F and MACR1C (α1–12), MC1_1793F and MACR1C (αC1), and MC2_1860F and MACR1C (αC2). All primer pairs except MACon geno Ex3 and Cast-Acon-Ex3-R spanned introns. MACon geno Ex and Cast-Acon-Ex3-R did not generate a PCR product in a no-RT control. The subsequent products were cloned into pCR4-TOPO (Invitrogen), and individual clones were sequenced at the Harvard Molecular and Cellular Biology Sequencing Facility (Applied Biosystems, Foster City, CA). χ^2 test for statistical significance was used.

We thank William Anderson and Jay Rajagopal for technical assistance with ES cells and Paula Cramer for assistance with the transgenic mice and discussions. We also thank Sean Buchanan, Benjamin tenOever, and other members of the Maniatis laboratory for advice and discussion. This work was supported by the National Institutes of Health (Grant 5RO1GM042231).

1. Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R (2006) *Cell* 126:403–413.
2. Kohmura N, Senzaki K, Hamada S, Kai N, Yasuda R, Watanabe M, Ishii H, Yasuda M, Mishina M, Yagi T (1998) *Neuron* 20:1137–1151.
3. Wu Q, Maniatis T (1999) *Cell* 97:779–790.
4. Takei Y, Hamada S, Senzaki K, Mutoh T, Sugino H, Yagi T (2001) *Genomics* 72:321–330.
5. Sano K, Tanihara H, Heimark RL, Obata S, Davidson M, St John T, Taketani S, Suzuki S (1993) *EMBO J* 12:2249–2256.
6. Wu Q, Zhang T, Cheng JF, Kim Y, Grimwood J, Schmutz J, Dickson M, Noonan JP, Zhang MQ, Myers RM, Maniatis T (2001) *Genome Res* 11:389–404.
7. Sugino H, Hamada S, Yasuda R, Tuji A, Matsuda Y, Fujita M, Yagi T (2000) *Genomics* 63:75–87.
8. Esumi S, Kakazu N, Taguchi Y, Hirayama T, Sasaki A, Hirabayashi T, Koide T, Kitsukawa T, Hamada S, Yagi T (2005) *Nat Genet* 37:171–176.
9. Kaneko R, Kato H, Kawamura Y, Esumi S, Hirayama T, Hirabayashi T, Yagi T (2006) *J Biol Chem* 281:30551–30560.
10. Tasic B, Nabholz CE, Baldwin KK, Kim Y, Rueckert EH, Ribich SA, Cramer P, Wu Q, Axel R, Maniatis T (2002) *Mol Cell* 10:21–33.
11. Wang X, Su H, Bradley A (2002) *Genes Dev* 16:1890–1905.
12. Frazer KA, Elnitski L, Church DM, Dubchak I, Hardison RC (2003) *Genome Res* 13:1–12.
13. Nobrega MA, Ovcharenko I, Afzal V, Rubin EM (2003) *Science* 302:413.
14. Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W, Rubin EM, Frazer KA (2000) *Science* 288:136–140.
15. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) *Nucleic Acids Res* 32:W273–W279.
16. Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I (2000) *Bioinformatics* 16:1046–1047.
17. Sugino H, Toyama T, Taguchi Y, Esumi S, Miyazaki M, Yagi T (2004) *Gene* 337:91–103.
18. Li Q, Harju S, Peterson KR (1999) *Trends Genet* 15:403–408.
19. Purucker M, Bodine D, Lin H, McDonagh K, Nienhuis AW (1990) *Nucleic Acids Res* 18:7407–7415.
20. Johnson KD, Christensen HM, Zhao B, Bresnick EH (2001) *Mol Cell* 8:465–471.
21. Lee GR, Fields PE, Griffin TJ, Flavell RA (2003) *Immunity* 19:145–153.
22. Kothary R, Clapoff S, Darling S, Perry MD, Moran LA, Rossant J (1989) *Development (Cambridge, UK)* 105:707–714.
23. Tewari R, Gillemans N, Harper A, Wijgerde M, Zafarana G, Drabek D, Grosveld F, Philipsen S (1996) *Development (Cambridge, UK)* 122:3991–3999.
24. Carroll P, Gayet O, Feuillet C, Kallenbach S, de Bovis B, Dudley K, Alonso S (2001) *Mol Cell Neurosci* 17:611–623.
25. Senzaki K, Ogawa M, Yagi T (1999) *Cell* 99:635–647.
26. Morishita H, Murata Y, Esumi S, Hamada S, Yagi T (2004) *NeuroReport* 15:2595–2599.
27. Visel A, Thaller C, Eichele G (2004) *Nucleic Acids Res* 32:D552–D556.
28. Heymann R, Kallenbach S, Alonso S, Carroll P, Mitsiadis TA (2001) *Mech Dev* 106:181–184.
29. Eggan K, Akutsu H, Loring J, Jackson-Grusby L, Klemm M, Rideout WM, III, Yanagimachi R, Jaenisch R (2001) *Proc Natl Acad Sci USA* 98:6209–6214.
30. Mombaerts P, Wang F, Dulac C, Chao SK, Nemes A, Mendelsohn M, Edmondson J, Axel R (1996) *Cell* 87:675–686.
31. Abe Y, Kouyama K, Tomita T, Tomita Y, Ban N, Nawa M, Matsuoka M, Niikura T, Aiso S, KitaY, Iwatsubo T, Nishimoto I (2003) *J Neurosci* 23:8513–8525.
32. Bender MA, Roach JN, Halow J, Close J, Alami R, Bouhassira EE, Groudine M, Fiering SN (2001) *Blood* 98:2022–2027.
33. Rhoades KL, Singh N, Simon I, Glidden B, Cedar H, Chess A (2000) *Curr Biol* 10:789–792.
34. Guo L, Hu-Li J, Paul WE (2005) *Immunity* 23:89–99.
35. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning* (Cold Spring Harbor Laboratory Press, Plainview, NY), 2nd Ed.
36. Smit A, Hubley R, Green P (1996–2004) RepeatMasker-Open-3.0 (www.repeatmasker.org).
37. Reitman M, G Felsenfeld. (1990) *Mol Cell Biol* 10:2774–2786.