

# Monoallelic expression of protocadherin genes

Andrew Chess

Random monoallelic expression is known to affect a variety of autosomal genes involved in specifying cell identity. Now, the neuronally expressed protocadherins can be added to this list.

Although most autosomal genes are expressed biallelically, a number of important genes are randomly monoallelically expressed<sup>1</sup>, similar to the genes subject to random X-chromosome inactivation in female cells. Among these are the genes encoding immunoglobulins, T cell receptors, interleukins and odorant receptors. These gene families share a key characteristic: specification of cellular identity. Now, the protocadherin (*Pcdh*) gene family can be added to the list of monoallelically expressed genes, as reported by Esumi *et al.* on page 171 of this issue<sup>2</sup>. This observation suggests that stochastic regulation of the *Pcdh* genes helps to specify individual neuronal cell identity.

## Unusual organization

The *Pcdh* genes are a diverse group encoding cell surface proteins with N-terminal cadherin-like extracellular domains<sup>3</sup>. Protocadherins localize to synapses, and their diversity could thus be involved in the complexities of forming distinct synapses. Three *Pcdh* gene families (*Pcdha*, *Pcdhb* and *Pcdhc*) reside in a single large cluster located on chromosome 18 in mouse<sup>4</sup> and on chromosome 5q in humans<sup>5</sup>, in a region implicated in genetic predisposition to schizophrenia<sup>6</sup>. The genomic organization of these genes is quite unusual. The organization of the *Pcdha* and *Pcdhc* genes resembles that of the immunoglobulin and T-cell receptor gene clusters<sup>5</sup>, with more than a dozen tandemly arrayed variable region (V) exons followed by a set of three constant (C) exons. For each *Pcdha* and *Pcdhc* gene, the entire extracellular domain (including six cadherin-like domains) as well as the transmembrane domain and a portion of the cytoplasmic domain are encoded by the V exons, with the remainder of the cytoplasmic domain encoded by the C exons. Each *Pcdha* gene therefore has the same C terminus, and each *Pcdhc* gene has the same C terminus, but the  $\alpha$  and  $\gamma$  termini are distinct. The initial discovery of this genomic organization raised the possibility

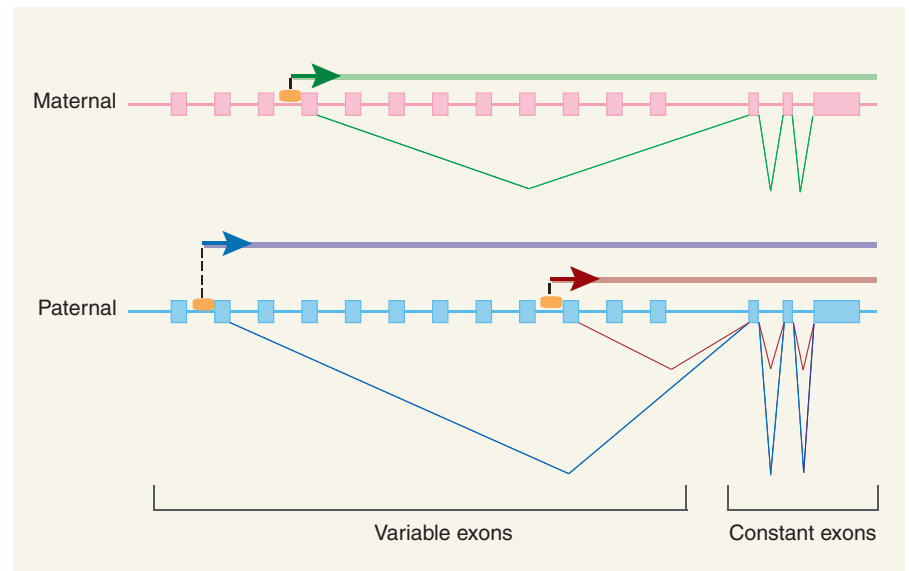
that DNA rearrangement was responsible for generating the various isoforms, but this possibility was discounted by studies showing that promoter choice determines the expression of each particular V exon and its correct splicing to the C exons<sup>7</sup>.

For most genes that specify different neuronal types, including most cadherin-like genes, patterns of expression are restricted to specific brain regions defined by the location of cell bodies or the sites of neuronal projections. By contrast, the individual *Pcdh* genes are expressed in overlapping punctate patterns in different regions of the brain<sup>3</sup>, as if they were regulated independently and with a stochastic component such that each is sprinkled among the neurons in a given region. A similar punctate pattern of expression is also observed for the odorant receptor gene family<sup>8</sup>. In the case of the odorant receptors, each neuron seems to express a single gene that serves to define the odors to which the neuron can respond<sup>9</sup> and also dictates axon guidance properties of the neuron<sup>10</sup>, features crucial for cellular identity of olfactory neurons.

## Random monoallelic *Pcdh* expression

Studying individual isolated Purkinje cells by RT-PCR, Esumi *et al.* found strong evidence for monoallelic expression of *Pcdha* genes<sup>2</sup>. For each *Pcdha* gene examined, some cells expressed the maternal allele, others expressed the paternal allele, still others expressed both the maternal and paternal, and others expressed neither allele. This suggests that the alleles are independently regulated, as has been suggested for other randomly monoallelically expressed genes.

Genes are typically regulated by the transcription factor milieu in which they are immersed, but this is not the case for randomly monoallelically expressed genes. Despite their exposure to the same transcription factor environment, the two sequence-identical alleles are differentially regulated such that one allele can be active in the absence of activation of the other allele. Though not directly proposed by Esumi *et al.*, a stochastic gene regulation model emerges from their study, in which a limiting number of 'activating complexes' exists in each nucleus. Each individual com-



**Figure 1** Random monoallelic expression of *Pcdha* genes. Of the 24 possible V exons (12 from each allele), three have been chosen, as indicated by a yellow blob present in a promoter region. The primary transcripts are indicated with arrows above the chosen V exons followed by straight lines. The splicing that occurs to generate the mature mRNA is drawn below.

Andrew Chess is at the Center for Human Genetic Research, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, 02114, USA. e-mail: achess@partners.org

plex can bind stably to a single piece of DNA (*i.e.*, one of the promoters upstream of a given V exon). This immediately gives rise to the punctate pattern of expression observed, as the limiting number of complexes leads directly to only a subset of *Pcdh* genes being active in each neuron. This type of mechanism would also lead to random monoallelic expression, with each neuron expressing the maternal allele, the paternal allele, both alleles or neither allele of each *Pcdh* gene (Fig. 1).

### Cell-specific identity

The precise regulation of spatial and temporal expression of genes allows specification of different parts of the nervous system, just as it allows specification of different parts of a developing organism. Distinguishing very similar neurons can also be accomplished by the differential activities of various transcription factors. The stochastic monoallelic expression of *Pcdh* genes provides an additional layer of cellular specification to otherwise similar cells. The reason for the existence of different forms of *Pcdh* may lie in their collective ability to give distinct identities to adjacent cells that are otherwise transcriptionally identical. This would be a type of individual cell 'self versus nonself' distinction, potentially useful to neurons as they form specific synapses<sup>11</sup>.

Another notable example of neuronal self versus nonself distinction has emerged from recent single-cell analyses of the alterna-

tive splicing of the *Drosophila melanogaster* Down syndrome cell adhesion molecule gene (*Dscam*), an immunoglobulin superfamily gene involved in specifying neuronal connections. The alternative splicing of *Dscam* is particularly notable, with 38,016 possible splice forms allowing 19,008 distinct extracellular forms and two alternative transmembrane domains: exon 4 has 12 possible forms, exon 6 has 48 possible forms, and exon 9 has 33 possible forms<sup>12</sup>. The variable exons are quite divergent so that the encoded immunoglobulin domains are predicted to have very different three-dimensional structures, and recent studies suggest that homophilic interactions are stronger than heterophilic interactions<sup>13</sup>.

Single-cell analyses of *Dscam* diversity led to the unanticipated finding that *Dscam* alternative splicing introduces a stochastic aspect to the generation of diversity of neurons<sup>14</sup>. Although different populations of neurons have few limitations on choice and are therefore probably not distinguished by *Dscam*, each individual neuron seems to be different from its neighboring cells by virtue of expressing different *Dscam* isoforms. This allows each neuron to know when a portion of its membrane is contacting itself as opposed to an adjacent neuron (that could be identical except for the *Dscam* difference). This self versus nonself awareness at the single-cell level allows two adjacent cells, with theoretically identical levels of transcription of

every gene in the genome, to be distinct from each other<sup>11,14</sup>.

It is noteworthy that the vertebrate *Dscam* genes studied to date do not show extensive alternative splicing. Similarly, although the *Pcdh* genes in vertebrates including fish all have similar genomic structures<sup>15</sup>, the homologs in *Drosophila* species have a simple organization. Thus, a new general principle emerging from a variety of systems suggests that going beyond the diversity of cell types afforded by normal modes of gene regulation is an important aspect of forming complex neural systems. In different systems, different genes are involved and the mechanisms are quite diverse. But in each case, a stochastic process leads to important differences among adjacent (and otherwise identical) neurons. As humans, we like to think of ourselves as individuals. It seems as though our neurons strive for individuality as well.

1. Singh, N. *et al. Nat. Genet.* **33**, 339–341 (2003).
2. Esumi, S. *et al. Nat. Genet.* **37**, 171–176 (2005).
3. Kohmura, N. *et al. Neuron* **20**, 1137–1151 (1998).
4. Sugino, H. *et al. Genomics* **63**, 75–87 (2000).
5. Wu, Q. & Maniatis, T. *Cell* **97**, 779–790 (1999).
6. Lewis, C.M. *et al. Am. J. Hum. Genet.* **73**, 34–48 (2003).
7. Tasic, B. *et al. Mol. Cell* **10**, 21–33 (2002).
8. Ngai, J. *et al. Cell* **72**, 667–680 (1993).
9. Malnic, B. *et al. Cell* **96**, 713–723 (1999).
10. Mombaerts, P. *et al. Cell* **87**, 675–686 (1996).
11. Neves, G. & Chess, A. *Cold Spring Harb. Symp. Quant. Biol.* (in the press).
12. Schmucker, D. *et al. Cell* **101**, 671–684 (2000).
13. Wotjowicz, W.M. *et al. Cell* **118**, 619–633 (2004).
14. Neves, G. *et al. Nat. Genet.* **36**, 240–246 (2004).
15. Tada, M.N. *et al. Gene* **340**, 197–211 (2004).

## Repeats hasten evolution?

We normally think of evolutionary change, such as that underlying differences between species of mammals, occurring over many tens of millions of years. Even 'rapid evolution', exemplified by the diversity that defines us as human individuals, occurs over tens or hundreds of thousands of years. But the extreme changes in dog morphology that have taken place over the last 150 years are exceptionally fast by normal evolutionary standards. What mechanism would allow for such accelerated change? One school of thought holds that adaptation occurs primarily through selection for single-base pair changes. Because these mutations tend to occur relatively infrequently, however, it seems unlikely that they could underlie the rapid changes in the dog. Another school of thought favors mutations in regulatory regions, leading to changes in levels and tissue specificity of gene expression, as means for rapid shifts in evolution.

Now, John Fondon and Harold Garner of the University of Texas at Dallas report evidence for selection acting on tandem repeats (*Proc. Natl. Acad. Sci. USA* **101**, 18058–18063; 2004). Tandem repeats are plentiful, contributing to 3% of the human genome, and occur more frequently, 10<sup>5</sup> times as often as single-point mutations. Selection acting on this class of mutations may, therefore, explain a faster rate of evolution. Fondon's study examined variation in tandem-repeat length in 142 dogs, including 92 domesticated breeds. The authors sequenced 37 repeat regions of 17 genes homologous to human and mouse genes involved in development and found significant variation in the number of repeats, correlated to several morphological phenotypes. This may not reflect typical evolution in the wild, particularly because domesticated dogs have undergone intensive artificial breeding selection for traits desirable to their owners, including morphological changes.

Most of the tandem repeats analyzed were located in coding regions. Although the function of these changes in repeat length remains to be explored, previous studies have shown that repeat expansions or contractions can cause either loss of function or hypomorphic alleles. In the human genome, expansion of triplet repeats has been associated with several hereditary neurological diseases, including fragile X syndrome, Huntington disease, myotonic dystrophy and spinocerebellar ataxia. The function of repeat expansion or contraction in disease, as well as the relative prevalence and importance of tandem repeats compared with other types of mutations, are important questions to address in future studies in this emerging field.

Orli Bahcall

