

6-5-2012

# CTCF and cohesin help neurons raise their self-awareness

Job Dekker

*University of Massachusetts Medical School*

Follow this and additional works at: [https://escholarship.umassmed.edu/pgfe\\_pp](https://escholarship.umassmed.edu/pgfe_pp)



Part of the [Genetics and Genomics Commons](#)

---

## Repository Citation

Dekker, Job, "CTCF and cohesin help neurons raise their self-awareness" (2012). *Program in Gene Function and Expression Publications and Presentations*. 192.

[https://escholarship.umassmed.edu/pgfe\\_pp/192](https://escholarship.umassmed.edu/pgfe_pp/192)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Program in Gene Function and Expression Publications and Presentations by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

# CTCF and cohesin help neurons raise their self-awareness

Job Dekker<sup>1</sup>

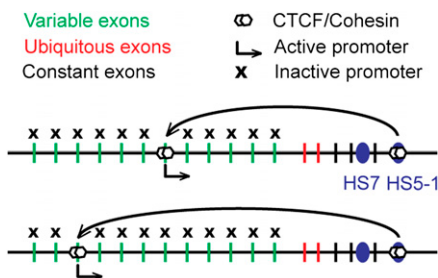
Programs in Systems Biology and Gene Function and Expression, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605-0103

In the brain neurons form extensively intermingled projections and synaptic interactions between them. To achieve proper neural connectivity, interactions between sections belonging to the same neuron are avoided. However, how can a cell tell, when it encounters another neuronal projection, whether it is part of the same cell? In other words, how do neurons become self-aware?

Work of the last several years has revealed a complex system of self-recognition that involves expression of variable and cell-specific sets of cell-surface proteins (1). These sets constitute unique “barcodes” for each neuron and serve as molecular readouts of cell identity so that cells can distinguish between “self” and “non-self.” In PNAS, Monahan et al. (2) present data that provide insights into the molecular process by which otherwise identical cells can express unique “barcodes” of membrane proteins. Their results point to a role of large-scale chromatin architecture in setting and maintaining stochastic and cell-specific expression patterns of cell-surface proteins.

Neurons express protocadherins, which are part of the larger family of calcium-dependent cell adhesion molecules (3). Cells express variable sets of protocadherin proteins, and each set defines an individual neuron by serving as a molecular recognition barcode that is used to differentiate between itself and other cells (4). Given that any part of the brain contains large numbers of neurons, the number of unique combinations of protocadherins a cell can potentially express must be similarly large. In mouse more than 50 protocadherin isoforms are expressed from three complex gene clusters, the protocadherin  $\alpha$ ,  $\beta$ , and  $\gamma$  clusters (5). Each of the clusters expresses a number of variants. When this is combined with the fact that each variant is monoallelically expressed, the number of possible unique combinations of expressed protocadherin protein sets is estimated to be up to 3 million (2).

Expression of variants from the protocadherin  $\alpha$  complex is being studied in detail. The cluster is composed of a set of 14 variable first exons, each of which contains its own promoter (Fig. 1). The variable exons are spliced to a set of three constant exons to form a unique variant (5). Expression of a given variant



**Fig. 1.** Stochastic and monoallelic expression of different protocadherin  $\alpha$  variants. The paternal and maternal alleles of the protocadherin  $\alpha$  cluster are shown. On each allele the HS5-1 enhancer touches only one stochastically chosen variant exon promoter. The expressed promoters and HS5-1 bind CTCF and cohesin, which may mediate long-range promoter–enhancer interactions.

is therefore determined by the activity of its promoter and specific splicing of that exon to the downstream constant exons with removal of all intervening variable exons (6). Further, two enhancers, located near the 3' end and downstream of the cluster (HS7 and HS5-1, respectively), drive expression of the cluster in neurons (7). Previous work has shown that the HS5-1 enhancer and each of the promoters of the variable exons bind the CCCTC-binding factor (CTCF) protein (8, 9). CTCF is involved in a number of processes related to gene regulation and that all seem to involve long-range looping interactions between regulatory elements such as enhancers and promoters (10). When analyzed by chromatin immunoprecipitation it seems that in whole-brain samples all promoters of the variable exons bind CTCF. Further, CTCF binding to the variable exon promoters depends at least in part on transcription of the exons, because binding is reduced in nonexpressing liver cells. Interestingly, the distal HS5-1 enhancer is required for strong binding of CTCF to variable exons 6–12, pointing to long-range communication between the downstream enhancers and the variable exon promoters.

The previous work hinted at roles for CTCF and long-range phenomena in regulating expression of the protocadherin  $\alpha$  cluster but did not provide clues for how cells decide to pick any variable exon for expression from a given allele. The work by Monahan et al. (2) starts to shed light on this stochastic process. The authors present three important results that suggest that CTCF and the cohesin

complex are involved in selective expression of variable exons. First, analysis of two mouse neuroblastoma cell lines (CAD and N2A) shows that CTCF is only bound to the subset of variable exons that are expressed in these cell lines and also to HS5-1. Thus, the original results obtained with whole-brain tissue that showed that CTCF associates with all variable exons reflect the average of a highly diverse population, in which each cell expresses different variable exons, and in those cells only the corresponding promoters associate with CTCF (Fig. 1). The present results obtained in the clonal cell lines suggest that once a cell has picked a particular variable exon for expression and CTCF binding, it can somehow maintain this choice even through mitosis.

Second, the authors find that the cohesin complex colocalizes with CTCF at the variable exon promoters and the HS5-1 enhancer and is also involved in protocadherin  $\alpha$  regulation. Cohesin was originally identified as a complex mediating interactions between sister chromatids, but more recently it has become clear that cohesin also has important other roles in long-range gene regulation (10). Previous work has shown that cohesin and CTCF colocalize on chromatin and act together, for example, as insulator-like elements in reporter assays (11). Interestingly, Monahan et al. (2) find that although all CTCF-bound sites also bind cohesin, the reverse is not always true: cohesin associates with HS7 and *Pcdhac2* (a ubiquitously expressed variable exon) in the absence of CTCF. Cohesin has been found to act independently of CTCF before [e.g., with the Mediator complex (12)]. Further, Monahan et al. provide evidence that cohesin and CTCF have overlapping but distinct roles in regulation of protocadherin  $\alpha$  expression. Knocking down CTCF resulted in reduced expression of most variable exons, whereas knocking down Rad21 (a subunit of cohesin) led to reduced expression of only the most 3' part of the locus containing the last variable exon (*Pcdh $\alpha$ 12*) and the two ubiquitously expressed exons, *Pcdhac1* and *Pcdhac2*. Thus, regulation of the

Author contributions: J.D. wrote the paper.

The author declares no conflict of interest.

See companion article on page 9125.

<sup>1</sup>E-mail: Job.Dekker@umassmed.edu.

cluster involves CTCF, cohesin, two enhancers, and complex interplay between them.

The third intriguing observation is that both CTCF and cohesin cobind pairs of sites at each of the variable exon promoters and at HS5-1. The complexes bind the previously identified conserved sequence element located approximately 200 bp upstream of the translation start site, as well as a second site located within the exon. The two sites are separated by a very similar distance, approximately 500 bp, at all variable exons and HS5-1. The functional significance of the occurrence of paired sites at all these elements is currently unknown. Recently, it was shown that intragenic CTCF-bound sites can affect splicing (13). Thus, it is tempting to speculate that the paired sites of CTCF and cohesin are somehow related to the two processes that are involved in expression of the variable exons: stochastic long-range activation of the promoter and alternative splicing to the constant exon. However, paired sites are not observed in the other protocadherin clusters (Pcdh $\beta$  and Pcdh $\gamma$ ).

Together, the present observations point to a role for CTCF and cohesin-mediated long-range interactions between the two enhancers (HS5-1 and HS7) and the variable exon promoters

## Monahan et al. provide evidence that cohesin and CTCF have overlapping but distinct roles in regulation of protocadherin $\alpha$ expression.

in stochastic expression of the protocadherin  $\alpha$  cluster (Fig. 1). In one highly speculative model, the enhancers may pick a single variable exon by a random looping event. Subsequently, this promoter will acquire or maintain stable binding by CTCF and cohesin, whereas the other promoters will not. Alternatively, another yet to be defined stochastic event could

determine CTCF and cohesin binding to any of the variable exons. As a result only the CTCF/cohesin-bound promoter can be activated through communication with the distal CTCF/cohesin-bound enhancers. In both models DNA methylation may be involved in inhibiting CTCF binding to the remaining promoters. Future studies with clonal cell populations aimed at DNA methylation mapping, as well as assessment of chromatin looping [e.g., by 3C (14)] will be important to more directly relate promoter selection to local chromatin structure and long-range 3D chromosome architecture.

Regulation of the expression of the protocadherin clusters provides an important model for the study of the mechanisms that cells use to generate cell-to-cell variability in gene expression. Other examples involve the stochastic selection and expression of single odorant receptors in individual olfactory neurons (15) and the complex cell-specific alternative splicing of the DSCAM gene in *Drosophila* (16). Could variable and cell-specific 3D arrangement of chromatin be a common determinant in stochastic and monoallelic gene expression throughout the genome?

- Zipursky SL, Sanes JR (2010) Chemoaffinity revisited: Dscams, protocadherins, and neural circuit assembly. *Cell* 143:343–353.
- Monahan K, et al. (2012) Role of CCCTC binding factor (CTCF) and cohesin in the generation of single-cell diversity of Protocadherin- $\alpha$  gene expression. *Proc Natl Acad Sci USA* 109:9125–9130.
- Frank M, Kemler R (2002) Protocadherins. *Curr Opin Cell Biol* 14:557–562.
- Esumi S, et al. (2005) Monoallelic yet combinatorial expression of variable exons of the protocadherin- $\alpha$  gene cluster in single neurons. *Nat Genet* 37:171–176.
- Wu Q, Maniatis T (1999) A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97:779–790.
- Tasic B, et al. (2002) Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Mol Cell* 10:21–33.
- Ribich S, Tasic B, Maniatis T (2006) Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proc Natl Acad Sci USA* 103:19719–19724.
- Kehayova P, Monahan K, Chen W, Maniatis T (2011) Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proc Natl Acad Sci USA* 108:17195–17200.
- Golan-Mashiach M, et al. (2012) Identification of CTCF as a master regulator of the clustered protocadherin genes. *Nucleic Acids Res* 40:3378–3391.
- Phillips JE, Corces VG (2009) CTCF: Master weaver of the genome. *Cell* 137:1194–1211.
- Wendt KS, et al. (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451:796–801.
- Kagey MH, et al. (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467:430–435.
- Shukla S, et al. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479:74–79.
- Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* 295:1306–1311.
- Chess A, Simon I, Cedar H, Axel R (1994) Allelic inactivation regulates olfactory receptor gene expression. *Cell* 78:823–834.
- Neves G, Zucker J, Daly M, Chess A (2004) Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nat Genet* 36:240–246.