Deciphering the Genetic Blueprint of Cerebellar Development by Gene Expression Profile Informatics

Teiichi Furuichi

Lab. for Molecular Neurogenesis, RIKEN Brain Science Institute, Wako, 351-0198 Japan

tfuruichi@brain.riken.jp

http://www.brain.riken.go.jp/labs/lmn/index.html

Abstract. The brain is the ultimate genetic system to which a large number of genes are devoted. In the post-sequencing era, it is now possible to elucidate how the brain develops and functions on a genetic basis. We are focusing on the postnatal development of mouse cerebellum as a model system, since it develops through a series of cellular developmental events on schedule within the fist three weeks of life. We explore the spatio-temporal gene expressions responsible for cerebellar development by utilizing the genome-wide approaches, and systematize them in an integrative database designated as the Cerebellar Development Transcriptome (CDT) database. The CDT database is an online neuroscience database with a number of search functions for gene expression profiles during cerebellar development and with a keyword search function (gene annotation, structures, cellular functions, etc.), and has multiple links to Web sites of public relevant bioinformatics databases to facilitate easy access to additional gene information. As a result, this CDT informatics provides us a foundation to extract new knowledge and to visualize an outline of the transcriptomic basis of cerebellar development.

1. Introduction

Recently, the genome sequencing projects suggest that the mouse and the human genomes each seem to contain in the neighborhood of 30,000 protein coding genes (1, 2). The brain with the structural and functional complexity is thought to express a large number of genes (3, 4, 5). In the post-sequencing era, it is time to address how many and what kinds of genes are involved in its development, wiring up, and function, by genome-wide and bio/neuroinformatic approaches.

The postnatal development of the mouse cerebellum is accomplished by a series of cellular developmental events (cell proliferation and migration, dendrogenesis and axogenesis, synaptogenesis, myelination, foliation and fissurization, etc.) that are genetically coded (Fig. 1). In the 1980's, there were some reports, by classical RNA-DNA hybridization kinetics analyses, showing

that complexity of poly(A)+ RNAs were increased during postnatal cerebellar development (4, 5, 6). Although many researchers have focused their attention on the hereditary plan, the genetic basis of cerebellar development remains incompletely understood. In order to decipher the genetic blueprint for cerebellar development, it would be necessary to conduct precise gene expression profiling in time and space, on a genome-wide basis, of all of the stages of development.

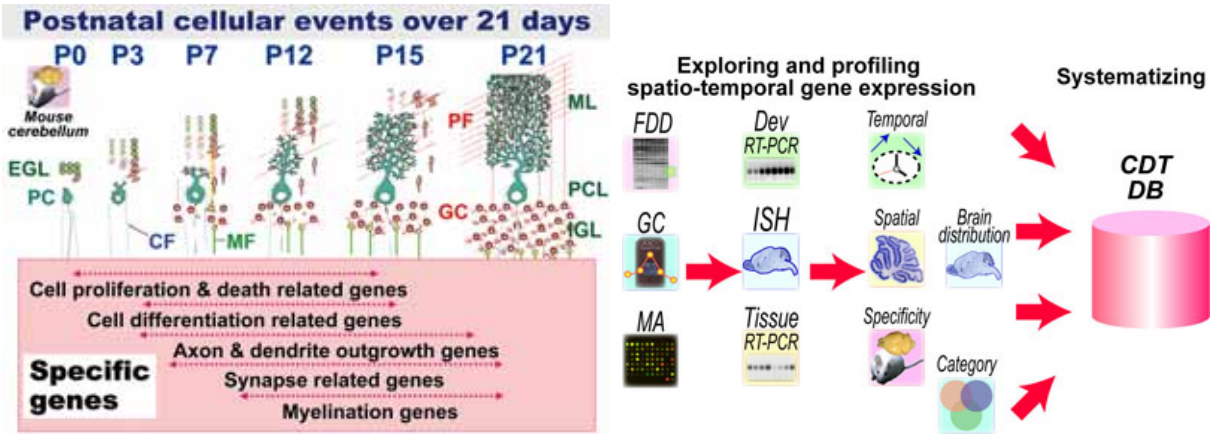

Figure 1. Cellular events and their responsible gene expressions during postnatal development of mouse cerebellum. *Left*: Differential expression of specific gene groups responsible for a series of magnificent developmental events of mouse cerebellum within the first three weeks of life. *EGL*, external germinal layer; *ML*, molecular layer; *PCL*, Purkinje cell layer; *IGL*, internal granular layer; *PC*, Purkinje cell; *GC*, granule cell; *CF*, climbing fiber; *MF*, mossy fiber; *PF*, parallel fiber. GCs are generated by vigorous cell proliferation in the EGL, extend their PF axons, and migrate downward to the IGL where the MF-GC synapses are formed. PCs undergo robust outgrowth of dendrites and form elaborate arborization with numerous synapses with the PFs and CFs. *Right:* Spatio-temporal gene expression profiles are analyzed by fluorescent differential display (*FDD*), GeneChip (*GC*), microarray (*MA*), developmental (*Dev*) and tissue-specific (*Tissue*) RT-PCR, and in situ hybridization (*ISH*). These expression data plus the data of brain distribution and gene category are systematized in an online Cerebellar Development Transcriptome database (*CDT-DB*).

From this perspective, we searched for genes that are differentially expressed at various developmental stages by employing a fluorescence differential display (FDD), a cDNA microarray, and a GeneChip (Affymetrix). Thereafter, the spatial (cellular) and temporal (developmental) specificities associated with the expression patterns were analyzed by performing *in situ* hybridization (ISH) brain histochemistry and reverse transcription-polymerase chain reaction (RT-PCR), respectively. We have successfully systematized the information collected regarding gene expression (e.g., gene annotation and clustering, spatio-temporal expression, tissue distribution, etc.) in an online CDT database. This CDT informatics will open a new window to understand the genetic blueprint for cerebellar development.

2. Results and Discussion

To understand the genetic blueprint for mouse cerebellar postnatal development, we are attempting to clarify the cerebellar development transcriptome (CDT) by exploring the gene expression responsible for the developmental stage by the genome-wide analysis approaches, and to systematize these lines of gene expression information in an integrative CDT database.

2-1 Exploration of Differential Gene Expression during Cerebellar Development

In this project, we extensively investigated spatio-temporal gene expression profiles during the postnatal development of the mouse cerebellum using genome-wide approaches. We then systematize these profiles in an integrative online CDT database with search functions for various gene expression patterns and with easy access to relevant public databases. As a result, our CDT database provides a foundation from which to extract and visualize an outline of the transcriptomic basis of cerebellar development.

Fluorescent Differential Display (FDD): First, by utilizing the FDD technique, we analyzed differential gene expression at eight stages of development of the mouse cerebella; these stages included embryonic day (E) 18, and postnatal days (P) 0, P3, P7, P12, P15, P21, and P56. From among total of approximately 12,000 bands analyzed, 83.1% were constantly expressed throughout the postnatal stages, whereas the remaining bands (16.9%, ca. 2,000 bands) showed differential patterns. Especially during the first and second postnatal weeks, when various cytogenetic and morphogenetic events are occurring simultaneously, a peak of molecular complexity (i.e., diversity in gene expression) was observed. By cloning and sequencing these differential bands, we succeeded in obtaining 2,194 non-redundant clones. These clones were categorized as a result of a homology search of DNA databases: 1,843 known gene clones (85.3% of the total), 57 homologous sequences (2.7%), 162 expressed sequence tags (ESTs) (7.5%), and 141 from unknown sequences (6.5%) (as of August, 2004). These genes were subjected to further detailed analyses of their spatio-temporal expression patterns, as described below.

GeneChip: Second, by utilizing the Affymetrix GeneChip system (Mu11K containing 12,654 genes), we analyzed the gene expression profiles observed during cerebellar development (7). Most of the genes represented on the GeneChip (total 10,321 genes, 81.6%) were found to be expressed in at least one of the five postnatal stages tested (E18, P7, P14, P21, and P56). Among these, 8.7% (897/10,321) showed apparent differential expression with more than two-fold changes. We already made the underlying data publicly available at our web site (http://www.brain.riken.go.jp/labs/lmn/GeneChipCblDev.html) and also through the NCBI-Gene Expression Omnibus database (http://www.ncbi.nih.gov/geo). From among the differentially-

expressed genes, 68.5% (614/897) showed the highest expression levels at early stages of development (i.e., either E18 or P7), suggesting that the high-level expression of various developmental genes is probably required for early-stage events such as cell proliferation and differentiation. The functional clustering of the genes identified by GeneChip analysis (34 gene clusters) indicates a transcriptomic feature of cerebellar postnatal development.

Microarray: Third, for the parallel monitoring of gene expression in brain development, function, and dysfunction, we amplified 2,404 cDNA sequences (2,137 FDD clones plus 267 cerebellar genes known to be indispensable for cerebellar development and function) by PCR and generated a custom-made "cerebellar development cDNA microarray" containing these amplified cDNA sequences. We applied this cDNA microarray to further investigate the gene expression profiling observed in cerebellar development. The data from the developmental analysis were then compiled in our CDT database.

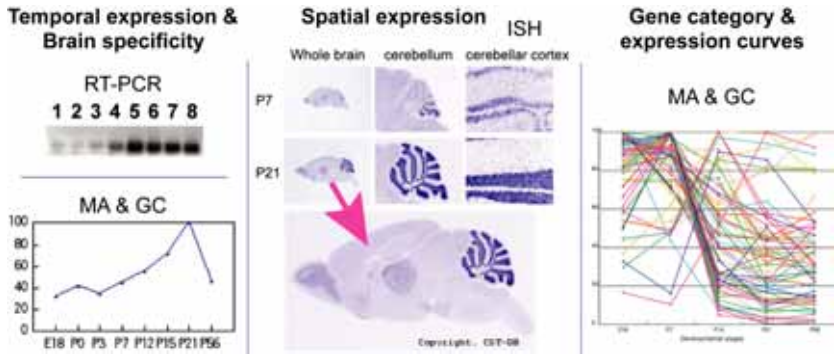

Figure 2. Spatio-temporal gene expression profiling data. Temporal patterns and brain specificities are defined by RT-PCR. GeneChip (GC) and microarray (MA) curves are also compiled for temporal expression profiling. Spatial patterns are defined by ISH. ISH images of whole brains, cerebella, and cerebellar cortices (including magnified views). 34 gene categories and expression curves (GC and MA data) within each category are compiled.

2-2. Spatio-Temporal Gene Expression Profiling

We intensively analyzed the spatio-temporal expression profiles of FDD and microarray clones (Fig. 2). The developmental expression patterns were analyzed by RT-PCR using cerebellar RNAs at E18, P0, P3, P7, P12, P21, and P56 (*temporal expression*). Then, the cellular expression patterns in sagittal sections of P7 and P21 mouse brains were analyzed using ISH histochemistry (*spatial expression*). The tissue distribution patterns were determined by RT-PCR analysis using RNAs prepared from 8 different tissues (brain, thymus, heart, lung, liver, kidney, spleen, and testis) *(brain specificity*).

2-3. Cerebellar Development Transcriptome (CDT) Database

We integrated into the CDT database the data obtained for approximately 2,600 genes with respect to the following features: spatio-temporal expression patterns (RT-PCR patterns of ca. 1,500 genes, ISH brain images of ca. 1,200 genes, expression plots and relative values of ca. 1,100 genes), annotation, functional clusters, etc. (Fig. 3). The CDT database is an online database (OS, Linux; database software, Oracle; operation software, Internet Explorer or Netscape) with a number of search functions (temporal patterns, spatial patterns, brain distribution patterns, brain specificity, keywords, ID number, etc.). The database has multiple links to public bioinformatics databases (Jackson Lab.-*MGI*, EBI-*Ensembl*, NICBI-*Entrez Gene*, *-GEO*, *-Nucleotide*, *-OMIM*, *-UniGene*, *-PubMed*) to facilitate easy access to additional gene information. By utilizing our CDT database, we were able to access many novel developmental gene candidates, *in silico*, which had not identified by conventional approaches (3, 8).

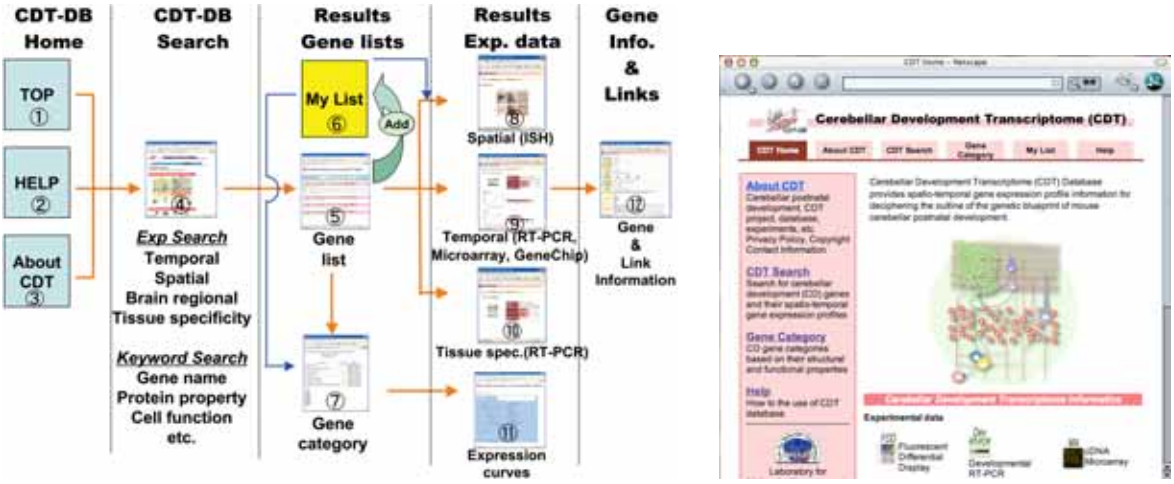

Figure 3. The outline of the Cerebellar Development Transcriptome (CDT) database (*left*) and the top page of online CDT database (*right*). *The CDT-DB will be online for the public on Feb. 2005.*

3. Conclusions

We have demonstrated that the postnatal development of the mouse cerebellum is genetically programmed and is accomplished by differential expression of thousands of genes that represent a variety of gene groups. Over 10,000 genes are expressed at various stages of mouse cerebellar development, and several thousand genes show differential expression patterns in time and space during this period. After integrating the spatio-temporal expression data, we developed a CDT database in which about the spatio-temporal expression information of 2,600 genes is systematized. In addition to regular updates and maintenance of the CDT database (all data sets, software, and

hardware), further comprehensive and bio/neuroinformatic studies in these areas will shed light on the genetic blueprint required for cerebellar development.

Acknowledgements

This study was supported by Grants-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the Japan Society for the Promotion of Science (JSPS) and by the Institute of Physical and Chemical Research (RIKEN).

References

1. International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. Nature (2001) 409: 860-921.
2. Mouse Genome Sequencing Consortium: Initial sequencing and comparative analysis of the mouse genome. Nature (2002) 420: 520-562.
3. Shiraishi, Y., Mizutani, A., Bito, H., Fujisawa, K., Narumiya, S., Mikoshiba, K., Furuichi, T.: Cupidin, an isoform of Homer/Vesl, interacts with the actin cytoskeleton and activated Rho family samall GTPases and is expressed in developing mouse cerebellar granule cells. J. Neuosci. (1999) 19: 8389-8400.
4. Kaplan, B.B.: Current Approaches to the Study of Gene Expression in the Adult and Developing Brain. In: Rassin, D., Harber, B, Brujan B. (eds.): Basic and Clinical Aspects of Nutrition and brain Development. Alan R. Liss, Inc., New York (1987) 131-155.
5. Kaplan, B.B.: RNA-DNA Hybridization: Analysis of Gene Expression. In: Lajtha, A. (ed.): The Handbook of Neurochemistry. Vol. 2. Plenum Press, New York (1982) 1-26.
6. Sutcliffe, J.G.: mRNA in the Mammalian Central Nervous System. In: Ann. Rev. Neurosci. Vol. 11. (1988) 157-198.
7. Kagami, Y., Furuichi, T.: Investigation of differentially expressed genes during the development of mouse cerebellum. Gene Expression Patterns (2001) 1: 39-59.
8. Sadakata, T., Mizoguchi, A., Sato, Y., Katoh-Semba, R., Fukuda, M., Mikoshiba, K., Furuichi, T.: Secretory granule-associated protein CAPS2 regulates neurotrophin release and cell survival. J. Neurosci. (2004) 24: 43-52.