

Cerebellar Development Transcriptome Database (CDT-DB)

The "cerebellar development transcriptome" (CDT) is a set of all transcription events responsible for the formation of cerebellar circuit. The "Cerebellar Development Transcriptome Database (CDT-DB)" project systematizes the spatiotemporal gene expression data obtained by genome-wide analysis approaches including DNA microarray and *in situ* hybridization and creates an analyzable database from a neuroinformatics perspective. The committee for the CDT-DB (currently called as BrainTx) project belongs to the Neuroinformatics Japan Center (NIJC), which is the Japan Node of the International Neuroinformatics Coordinating Facility (INCF). The database is integrated into the RIKEN integrated database of mammals of the RIKEN Scientists' Networking System (SciNetS).

Contents

1. Cerebellar Development Transcriptome Database (CDT-DB) Project

- 1-1. Objectives
- 1-2. Cerebellar postnatal development
- 1-3. Cerebellar development transcriptome (CDT)
- 1-4. Privacy Policy
- 1-5. Copyright
- 1-6. How to create URL links to the CDT-DB
- 1-7. Acknowledgments
- 1-8. CDT-DB Committee
- 1-9. Contact information

2. CDT-DB

- 2-0. Notes
- 2-1. About the CDT-DB
- 2-2. Information about cerebellar development (CD) genes
- 2-3. Search for CD genes and spatiotemporal gene expression patterns

3. Publications regarding the CDT-DB

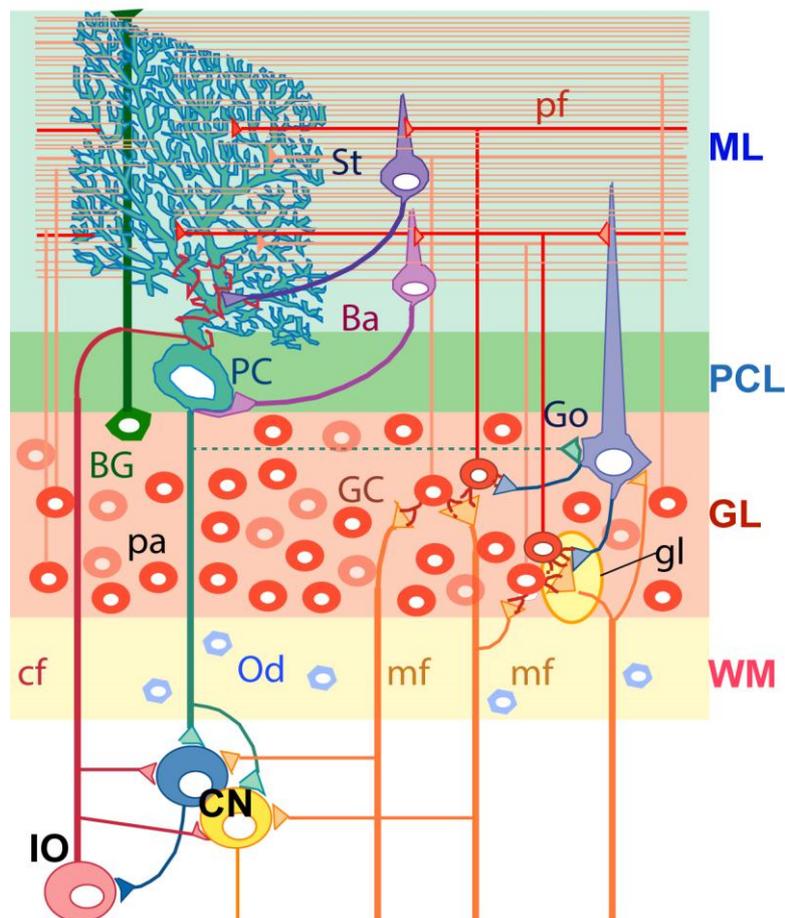
1. Cerebellar Development Transcriptome Database (CDT-DB) Project

1-1. Objectives

The basic design for the brain, which is a very complex structure, must be encoded in the genome, and is attributable to the controlled expression of thousands of specific genes in time and space. Toward deciphering the genetic blueprint for mouse cerebellar development, the CDT-DB project systematizes spatiotemporal gene expression profiles during postnatal development of mouse cerebellum in a database. The project to collect experimental data and relevant information was launched in the Laboratory for Molecular Neurogenesis, RIKEN Brain Science Institute (BSI) in 1999. The database (opened since 2005) is now organized by the BrainTx (formerly, CDT-DB) committee that belongs to the NIJC, RIKEN-BSI.

1-2. Cerebellar postnatal development

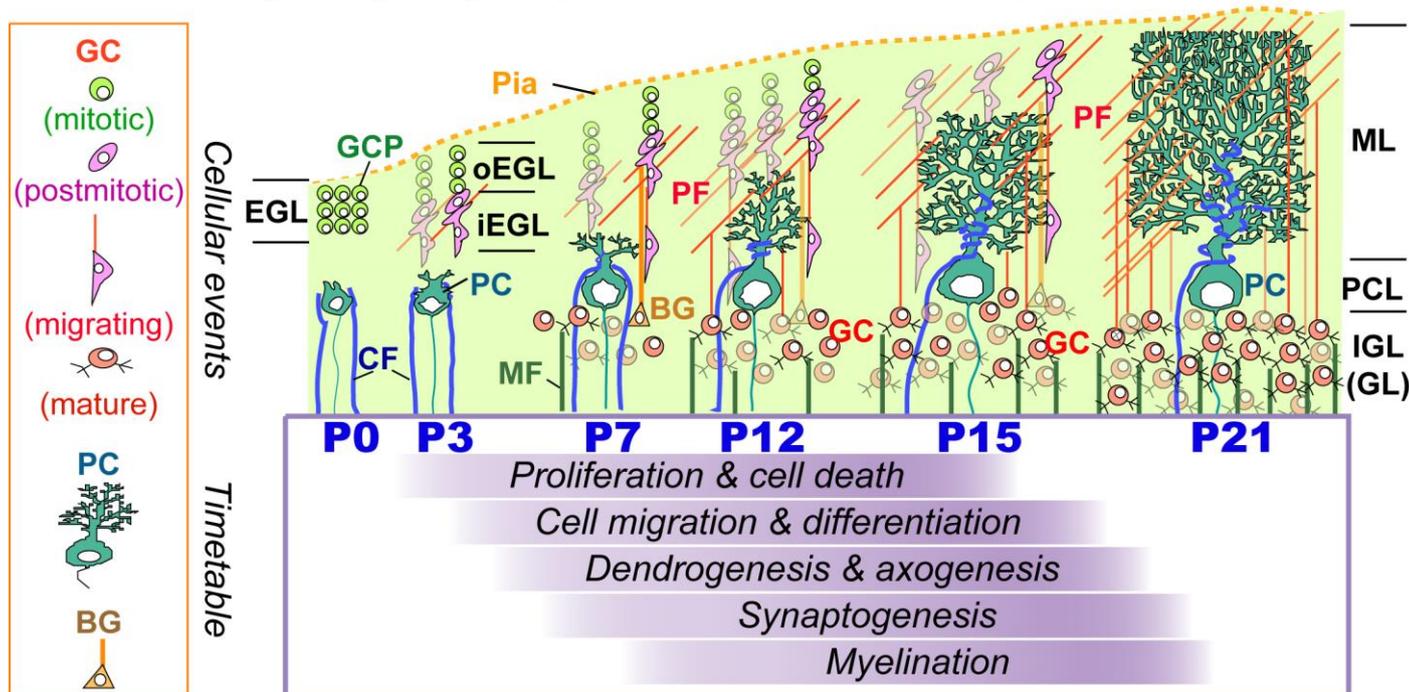
Fig. 1. The basic cellular organization and layer structure of the cerebellar cortex circuit



The cerebellum is a hindbrain region that coordinates gait and voluntary movement, is responsible for balance and posture, and is important in speech and control of gaze. We focus on the mouse cerebellum, which develops

into a functional circuit and architecture (Fig. 1) during the postnatal three-week period through a series of magnificent cellular developmental events (Fig. 2).

Fig. 2. Major stages of posnatal cerebellar cortex development in mice



Granule cells (GCs), the sole excitatory neurons in the cerebellar circuit, are generated by the vigorous proliferation of granule cell progenitors (GCPs) in the external germinal (or granular) layer (EGL) during the first two weeks after birth. This proliferation leads to an immense number of cells, accounting for approximately half of the neurons in the mammalian brain. Post-mitotic GCs then bilaterally extend their parallel fiber (PF) axons, and their cell bodies start to migrate downward through the developing molecular layer (ML), eventually settling in the internal granular layer (IGL) underneath the Purkinje cell layer (PCL). The EGL is divided into two sub-layers: the outer EGL (oEGL) containing proliferating GPCs and the inner EGL (iEGL) containing nascent post-mitotic GCs.

During these first three weeks, cells in the pia matter (PM) play a role in GC proliferation and migration, whereas Bergmann glia (BG) processes extending into the ML appear to guide GCs. GCs that reach the IGL further differentiate, and extend their dendrites into glomeruli in which GCs connect excitatory afferent mossy fibers (MFs) and inhibitory Golgi cell (Go) axons. A portion of the GC population in the EGL undergoes cell death (apoptosis), which is thought to occur to fine-tune the proper cell numbers and connectivity. During these GC events, Purkinje cells (PCs), the principle neurons of the cerebellar circuit, undergo a robust outgrowth of their dendrites and form elaborate arborizations on which PCs receive two excitatory inputs from PFs and climbing fibers (CFs). Distal spiny dendrites of PCs have numerous spines, and each spine forms a synaptic connection with an extending PF. Large proximal dendrites of PCs form synapses with CFs that are pruned away from multiple (at an early stage) to mono (at late stage) innervation in an activity-dependent manner. PC axons are the sole output from the cerebellar cortex to the cerebellar nuclei (CN) (or deep cerebellar nuclei, DCN). These developmental events involving GCs and PCs are mutually regulated by synergistic actions between GCs and PCs.

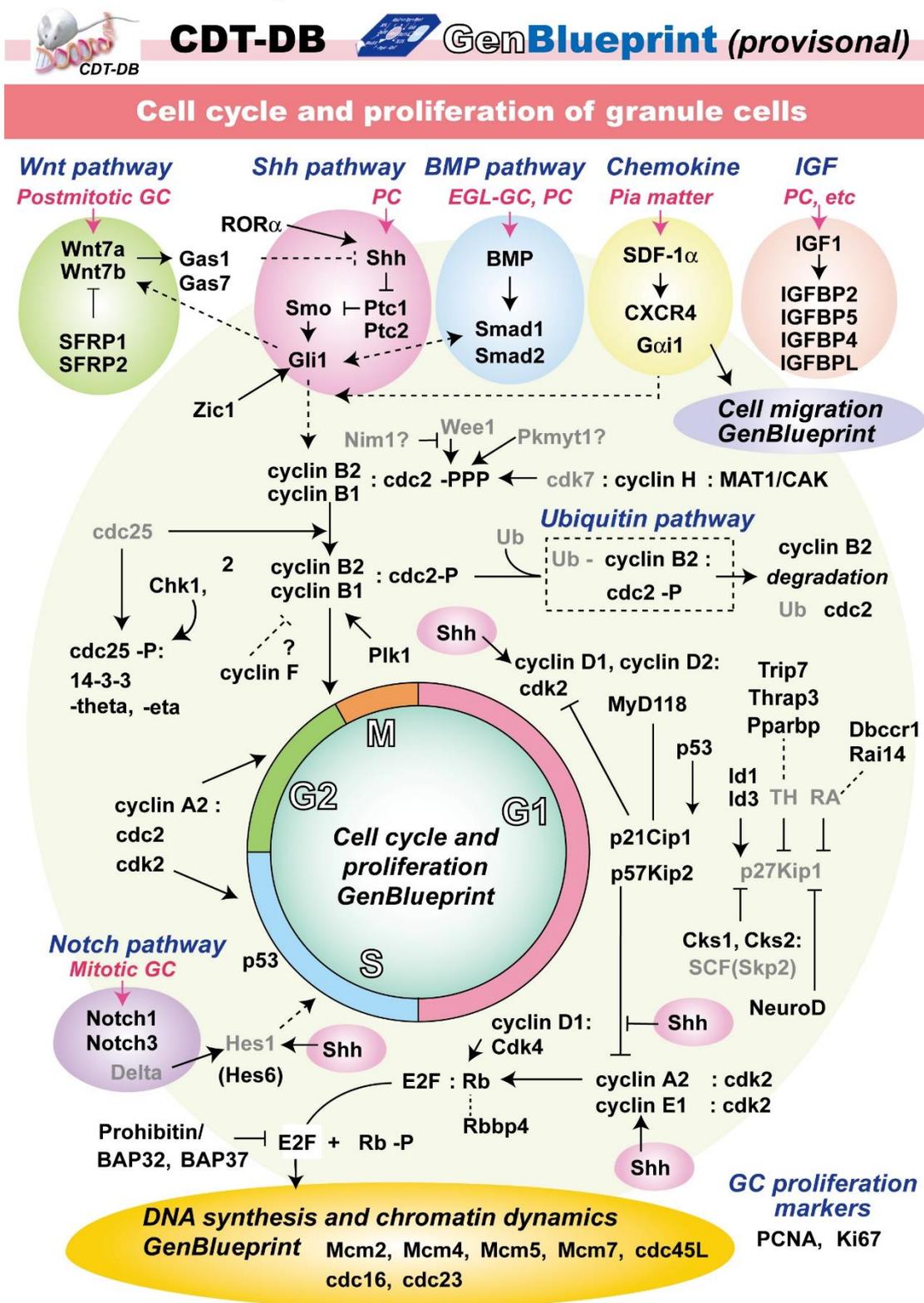
Three types of inhibitory interneurons, stellate cells (St), basket cells (Ba) and Golgi cells (Go), proliferate and then migrate during the first or second postnatal week to their proper positions-the outer two-thirds of the ML, the inner third of the ML, and the upper IGL, respectively-where they each form specific local connections (St and Ba, feed-forward pathways from PFs to PCs; Go, a feedback pathway from PFs to GCs) by the third postnatal week. Other minor interneurons, such as unipolar brush cells and Lugaro cells, are not indicated in Fig. 1 (see Fig. 1 on the "ISH Atlas" page). Oligodendrocytes (Od) proliferate and differentiate in the white matter (WM). Myelination starts to appear from P3 in the central deep WM and progresses in the WM of the folia in the first and second postnatal weeks, whereas that of PC axons in the IGL starts to appear by P10 and progresses by the third postnatal week. Astrocytes (As) and other cells are thought to play some roles in these cerebellar developmental events.

1-3. Cerebellar development transcriptome (CDT)

The genetic blueprint for cerebellar development should be mirrored in the cerebellar development transcriptome (CDT), a set of all transcription events during cerebellar development. In order for this series of postnatal development events to proceed smoothly, expression of specific genes and gene groups must be controlled in a timely way at each developmental stage. Assuming these genes are expressed in an orderly fashion on a developmental timetable, we formulated the idea that genes crucial to a series of these events could be efficiently identified by genome-wide analysis of gene expression at each developmental stage using fluorescent differential display (FDD) and DNA microarray (GeneChip and CDT array) techniques. In addition to this developmental time series, providing temporal gene expression profiling data, we spatially map gene expression on the developing cerebellar cortex circuit by *in situ* hybridization (ISH) brain histochemistry to profile spatial cellular gene expression. These spatiotemporal gene expression profile data are annotated by citing pre-defined annotation terms (see 2-2-2) and by classifying data based on the developmental and anatomical context (see 2-2-4) (also see Publication-1, Sato et al., 2008).

At present, the CDT-DB includes a list of cerebellar development (CD) genes and their integrated spatiotemporal expression profiles during postnatal cerebellar development. To decipher the genetic blueprint for cerebellar development, it will be necessary to further elucidate the gene cascades involved in each event. Figure 3 illustrates a functional network of CD gene products involved in granule cell proliferation. Most of the CD genes listed here were identified as part of the CDT-DB project (genes written in black), and thus can be found by searching the CDT-DB. Besides the listed genes, many others are involved in the cell cycle and proliferation of granule cell progenitors, and will be compiled in the CDT-DB in the future. After that, transcriptional regulation cascades for each gene or gene group will be elucidated so that the complete genetic blueprint (GenBlueprint) for cerebellar development can emerge.

Fig. 3. A functional network of the CD genes involved in the cell cycle and proliferation of cerebellar granule cells



2. CDT-DB

2-0. Notes

Note 1: Although some digital image data may not be of the highest quality, all data registered here are representative and are the most readily reproducible images we have obtained thus far. Digital images will be

updated as they become available (2-1).

Note 2: Because the FDD-derived, genome-hit CD clones marked with "Int" (intron sequence), "5'End" (5'-upstream sequence), or "3'End" (3'-downstream sequence) are provisional, we leave their gene descriptions up to the judgment of users (2-2-2). These FDD clones might be derived from non-coding RNAs and/or undefined exon sequences that are produced by alternative splicing events during cerebellar development.

Note 3: Some CD genes that show only slight developmental changes in their expression levels by our conventional semi-quantitative RT-PCR or GeneChip analyses are categorized for now into the "almost constant" regulation type, since more accurate methods are needed to determine the precise degree of change. The temporal patterns of some genes differ slightly depending on the method used, possibly due to differences in sensitivity between methods and other technical variations (2-2-4).

Note 4: ISH images of some CD genes, including even known genes, show nuclear staining patterns (mostly nucleoli), which may be due to either hybridization with unspliced RNAs or unknown causes (2-2-4).

2-1. About the CDT-DB

To elucidate the cerebellar development transcriptome (CDT), we are exploring differential expression of cerebellar development genes (CD genes) during the postnatal development of the mouse cerebellum by utilizing genome-wide gene expression analysis approaches, such as fluorescence differential display (FDD), cDNA microarray (CDT array) and GeneChip analyses. The temporal expression patterns of some CD genes are further analyzed by semi-quantitative reverse transcription-polymerase chain reaction (RT-PCR). The spatial expression patterns of CD genes are analyzed by *in situ* hybridization (ISH) brain histochemistry. The brain specificity of gene expression is being determined by RT-PCR or GeneChip analyses. We have systematized the spatiotemporal CD gene expression profile information and generated an integrative CDT-DB that can be used with Web browsers (see "**Help**" for how to use the CDT-DB). The CDT-DB features various search functions for CD genes and their spatiotemporal expression profiles, and provides easy accessibility to relevant public bioinformatics database websites.

In the CDT-DB, we intend to provide researchers with nearly raw data of our experiments (RT-PCR gel images, ISH brain images, and microarray diagrams) in formats that are familiar to most neuroscience researchers, so that users will be able to benefit from the CDT-DB immediately.

Note: Although some digital image data may not be of the highest quality, all data registered here are representative and reproducible. Digital images will be updated when improved data are obtained.

We hope that the CDT-DB will be of some help to researchers who are concerned with the molecular basis of brain development and disorders.

2-2. Information about cerebellar development (CD) genes



1. Cerebellar development (CD) genes and their identification number (CD ID)

CD genes are identified from mouse cerebellum tissue (ICR or C57B/6J) using FDD and GeneChip analyses, as described above.

The **CD ID** (CD plus five numerals: example CD98765) is the identification number assigned to each CD gene.

Gene expression data ID (= Parental ID (CD ID) plus sub ID): Multiple different transcripts of the same genes have been identified by employing FDD analysis with different primer sets and microarray analysis with different probes. In addition, multiple different probes, which are derived from different positions of the same gene, can react with either the same transcript or different ones such as alternative splice variants. The CDT-DB assigns the parent ID number (= CD ID) to each gene as described above and classifies different expression data for each gene, which are obtained from different transcripts of a gene or obtained by different probes of a gene, as sub ID numbers (gene expression data IDs) that follow the parent ID number.

Example: CD00001.1 and CD00001.2 mean IDs for two different expression data obtained from the same gene (parent ID = CD00001) and are distinguished from each other by sub IDs ".1" and ".2".

2. Gene name, gene description, alternative name

We have assigned **gene symbols**, **gene descriptions** and **alternative names** to CD genes by referring to the *NCBI(Entrez)-Gene*, *UniGene*, *MGI*, and *Ensembl* databases.

Although a fraction of the ESTs (expressed sequence tags) identified by the FDD analysis do not match any cDNA or EST sequences, but do match known genomic sequences, many of these FDD clones correspond to sequences within introns, 3'-flanking regions, and 5'-flanking regions of known genes. We thus assume that some of these are derived from alternatively spliced mRNAs, differentially terminated or initiated mRNAs, nuclear pre-mRNAs, or noncoding small RNAs (see also 2-0 Note 2). In the CDT-DB, such ESTs are *provisionally* annotated as follows:

Int: within the intron of a corresponding known gene

3'End (predicted): within the 3'-flanking region of a corresponding known gene

5'End (predicted): within the 5'-flanking region of a corresponding known gene

Here, we restrict the predicted flanking regions within, at most, 2.5 kb from either the 5' or the 3' ends of known genes. In any case, *it should be noted that because the FDD-derived, genome-hit CD clones marked "Int", "5'End", or "3'End" are provisional, we leave their gene description up to the judgment of users.*

3. Gene category



The CD genes are classified into **34 gene categories** according to the structural and functional properties of the gene products (encoded proteins), by referring to their annotations in the literature and/or to the terms used to describe them in the *MGI* and *Gene Ontology (GO)*.

4. Expression profiles



Timetabling of gene expression was carried out by developmental RT-PCR, GeneChip, and custom-made cDNA microarray (CDT array) analyses. Cellular mapping of gene expression was conducted by ISH analysis. Brain specificity (tissue distribution) of gene expression was estimated by tissue-specific RT-PCR and GeneChip analyses. The names of mouse strains analyzed (ICR or C57B/6J) are indicated on the expression information pages.

Temporal expression indicates developmental gene expression patterns determined by RT-PCR, GeneChip and CDT array analyses using RNA sources prepared from developing cerebella (RT-PCR and CDT array; E18, P0, P3, P7, P12, P15, P21, and P56) (GeneChip; E18, P7, P14, P21, and P56). **Note:** *Some CD genes that show only slight developmental changes in their expression levels by our conventional semi-quantitative RT-PCR are categorized for now into the "almost constant" expression pattern type, since more accurate methods are needed to determine the precise degree of change. The temporal patterns of some genes differ slightly according to the methods used, which may be due to differences in sensitivity between methods or other technical differences.*

Spatial expression and **Brain distribution** indicate cellular expression patterns in cerebella and regional distribution patterns in brains, respectively, evaluated by *in situ* hybridization (ISH) analysis of P7 and P21 mice. **Note:** *ISH images of some CD genes, including even known genes, show nuclear staining patterns (mostly nucleoli), which may be due to hybridization with unspliced RNAs or to unknown causes.*

Brain specificity indicates tissue distribution patterns determined by RT-PCR or GeneChip analyses using RNA sources from eight different tissues. For RT-PCR analysis, RNAs at either P7 or P21, depending on which stage shows a higher expression level, were used, whereas for GeneChip analysis RNAs from mice at both P7 and P21 were used.

5. EST sequences

The DNA sequences of all ESTs identified in this CDT-DB project were registered to the DNA Data Bank of Japan (DDBJ) and are available by accession number (BP426256-BP428449) from DDBJ/GenBank/EMBL.

6. Links

The CDT-DB includes easy links to relevant bioinformatics database sites. Thus, one can easily access additional information about most CD genes through these links.

Accession ID and EST sequences information: NCBI-Nucleotide. **General information:** MGI, Ensembl, UniGene, NCBI(Entrez)-Gene, OMIM, GEO, KEGG. **Genome information:** Harvester, UCSC Mouse Genome Browser Gateway. **Functional genomics:** GeneNetwork, Mouse Phenome Database, BioGPS, STRING. **RIKEN integrated database in mammals:** SciNetS, FANTOM, BioResource. **Nervous system:** SynDB, Mouse

Neuronal Expression Database, Molecular Brain Mouse Brain Transcriptome. **Reference information:** GoPubMed, iHOP, PubMed (NCBI). **Cellular expression information:** ABA, GenePaint, BGEM. **Microarray information:** Affymetrix. **INCF Japan-Node (NIJC) databases:** J-Node Cross-database Search, Cerebellar Platform, Mouse Phenotype Database.

7. References

To complement the gene annotation information, the CDT-DB lists papers selected by us and cited by relevant databases, and contains links to PubMed.

3. Publications regarding the CDT-DB

- 1) Sato, A., Sekine, Y., Saruta, C., Nishibe, H., Morita, N., Sato, Y., Sadakata, T., Shinoda, Y., Kojima, T., and Furuichi, T. (2008) Cerebellar development transcriptome (CDT-DB): profiling of spatio-temporal gene expression during the postnatal development of mouse cerebellum. *Neural Networks* 21:1056-1069. PubMed ID: [18603407](#)
- 2) Sato, A., Morita, N., Sadakata, T., Yoshikawa, F., Shiraishi-Yamaguchi, Y., Huang, JH., Shoji, S., Tomomura, M., Sato, Y., Suga, E., Sekine, Y., Kitamura, A., Shibata, Y., Furuichi, T. (2004) Deciphering the genetic blueprint of cerebellar development by the gene expression profiling informatics. *Neural Information Processing. Lecture Notes in Computer Science* 3316:880-884. Springer-Verlag, Berlin, Heidelberg, Germany. DOI: [10.1007/978-3-540-30499-9_135](#)
- 3) Kagami, Y., and Furuichi, T. (2001) Investigation of differentially expressed genes during the development of mouse cerebellum. *Brain Research Gene Expression Patterns* 1:39-59. PubMed ID: [15018818](#)
- 4) Furuichi, T., Shiraishi-Yamaguchi, Y., Sato, A., Sadakata, T., Huang, J., Shinoda, Y., Hayashi, K., Mishima, Y., Tomomura, M., Nishibe, H., and Yoshikawa, F. (2011) Systematizing and cloning of genes involved in the cerebellar cortex circuit development. *Neurochem Res.* 36:1241-1252. PubMed ID: [21243430](#)

The data obtained by developmental time series GeneChip analysis ([Publication-3](#), Kagami and Furuichi, 2001) are also available in the NCBI Gene Expression Omnibus (GEO) (Platform GPL8, Series GSE2, and Sample GSM50, GSM51, GSM52, GSM53, and GSM54).

- [GPL8](#)
- [GSE2](#)
- [GSM50](#)
- [GSM51](#)
- [GSM52](#)
- [GSM53](#)
- [GSM54](#)