Developmental Biology 397 (2015) 18–30

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

Resource

CbGRiTS: Cerebellar gene regulation in time and space

Thomas Ha [a], Douglas Swanson [a], Matt Larouche [a], Randy Glenn [a], Dave Weeden [a], Peter Zhang [a], Kristin Hamre [b], Michael Langston [d], Charles Phillips [d], Mingzhou Song [e], Zhengyu Ouyang [e], Elissa Chesler [f], Suman Duvvurru [f], Roumyana Yordanova [f], Yan Cui [c], Kate Campbell [a], Greg Ricker [g], Carey Phillips [g], Ramin Homayouni [h], Dan Goldowitz [a,*]

[a] Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Department of Medical Genetics, University of British Columbia, 950 West 28th Avenue, Vancouver, BC, Canada V5Z 4H4
[b] Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, TN, USA
[c] Department of Molecular Science, University of Tennessee Health Science Center, Memphis, TN, USA
[d] Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA
[e] Department of Computer Science, New Mexico State University, Las Cruces, NM, USA
[f] The Jackson Laboratory, Bar Harbor, ME, USA
[g] Department of Biology, Bowdoin College, Brunswick, ME, USA
[h] Bioinformatics Program, Department of Biology, University of Memphis, Memphis, TN, USA

ARTICLE INFO

Article history:
Received 30 January 2014
Received in revised form
23 August 2014
Accepted 27 September 2014
Available online 23 October 2014

Keywords:
Cerebellum
Development
Transcriptome
Granule cell
Mouse

ABSTRACT

The mammalian CNS is one of the most complex biological systems to understand at the molecular level. The temporal information from time series transcriptome analysis can serve as a potent source of associative information between developmental processes and regulatory genes. Here, we introduce a new transcriptome database called, Cerebellar Gene Regulation in Time and Space (CbGRiTS). This dataset is populated with transcriptome data across embryonic and postnatal development from two standard mouse strains, C57BL/6J and DBA/2J, several recombinant inbred lines and cerebellar mutant strains. Users can evaluate expression profiles across cerebellar development in a deep time series with graphical interfaces for data exploration and link-out to anatomical expression databases. We present three analytical approaches that take advantage of specific aspects of the time series for transcriptome analysis. We demonstrate the use of CbGRiTS dataset as a community resource to explore patterns of gene expression and develop hypotheses concerning gene regulatory networks in brain development.

Introduction

One of the most daunting biological research challenges is understanding how a genome consisting of $\sim 3 \times 10^4$ genes guides the development of the most complex of organ systems—the mammalian brain. In the case of humans, this complex system is composed of over 10^{11} cells with at least 10^3 different cell types, each with special molecular signatures and relations with other cells and brain regions. Although these cells begin as a histologically homogeneous sheet of neuroepithelial cells, they progressively divide into numerous cellular cohorts that are highly differentiated from each other in terms of morphology and functional potential. As these events occur according to a developmental timeline, time series transcriptome analyses can provide rich information regarding observed developmental events and underlying regulatory genes. The pioneering work of Davidson and colleagues, to lay out the gene regulatory networks of the developing endomesoderm of the sea urchin (Davidson and Erwin, 2006), emboldened us to start along that path for the mammalian brain. We have now finished the first major step with the creation of a developmental time series transcriptome analysis for the cerebellum, a brain region that may be the most tractable due to its simple laminar structure and limited number of neuronal types. The cellular composition, key molecular players, and structural and functional aspects of the developing cerebellum has been extensively studied (Hatten et al., 1997; Goldowitz and Hamre, 1998; Wang and Zoghbi, 2001). There are two main cell types that receive input to the cerebellar cortex: the granule cell, which is the most numerous neuron in the brain, and the Purkinje cell, which is the sole output of the cerebellar cortex. Although relatively simple in structure, the cerebellum has been shown to be important for higher cognitive function (Pernet et al., 2009) and has been implicated in several neurocognitive

* Corresponding author. Fax: +1 604 875 3840.
E-mail address: dang@cmmt.ubc.ca (D. Goldowitz).

http://dx.doi.org/10.1016/j.ydbio.2014.09.032
0012-1606/

disorders such as autism (Nayate et al., 2005; Tsai et al., 2012), schizophrenia (Andreasen and Pierson, 2008), and Fragile X (Koekkoek et al., 2005; Huber, 2006).

The number of identified genes known to be critically involved in cerebellar development has grown by more than an order of magnitude over the last 15 years, from about 15 genes (Goldowitz and Hamre, 1998) to over 200, and the number continues to grow (MGI and PubMed search terms: cerebellum, development, gene and knockout). But even this number of genes is only 1% of the genome. Microarray technology permits the analysis of whole transcriptomes, but typically only provides single snapshots of gene expression (Storey et al., 2005). Understanding the interplay of genes in development requires a more dynamic approach such as time series analysis (Bar-Joseph, 2004; Klevecz et al., 2007). Previously, time series transcriptome analyses have been published for mouse postnatal cerebellum and human prefrontal cortex (Sato et al., 2008; Colantuoni et al., 2011). Sato et al. (2008) generated a mouse cerebellum transcriptome database from 5 largely postnatal time points by sampling E18, P7, P14, P21 and P56 and Colantuoni et al. (2011) explored the temporal dynamics and genetic control of transcription in human prefrontal cortex samples that covers much of human developmental time. While these studies depict dynamic transcriptomes for the developing mammalian brain, they either do not sample from time periods where key early developmental events occur or the temporal resolution and sample size are limited compared to the time window covered. Here we report a novel time series transcriptome database that spans critical embryonic as well as postnatal cerebellar developmental times. We performed microarray analysis on whole cerebellar tissues from two inbred strains of mouse [C57BL/6J (B6) and DBA/2J (D2)] at 24-hour intervals in embryonic development from embryonic day 12 (E12) to postnatal day 0 (P0), and at 3-day intervals from P0 to P9 in the postnatal period, and recombinant inbred mice between these two genotypes (BXD RI lines) at 3–7 representative time points (E12.5, E15.5, E18.5, P0, P3, P6 and P9), and three mutant lines of mice whose mutant genes are known to target a single cell type of cerebellum, the cerebellar granule cell at E15.5 (*Math1* KO and *meander tail* mutant) or at E13.5, E15.5 and E18.5 (*Pax6* KO). We illustrate the use of this developmental time series transcriptome data with three bioinformatics analyses: differential equation modeling to predict transcript on and off time, paraclique analysis to identify genes with similar dynamics, and dynamical system modeling to infer transcriptional causal relationship from the time series data. We also demonstrate the utility of a new web-based toolkit called Cerebellar Gene Regulation in Time and Space (CbGRiTS) as a data exploration and analysis platform for studying gene regulatory networks in cerebellar development.

Results

In the "Results" section, we present three main sections. First, we will describe the CbGRiTS time course microarray datasets. Second, we will present the three informatic and statistical analyses that take advantage of the time series dataset and provide illustrations about how the dataset can be utilized. Finally, we describe the CbGRiTS web-based tools as a community resource.

Time series microarray analysis of the developing cerebellum

A genome-scale microarray analysis was conducted using the Illumina MouseWG-6 v1 Expression BeadChip platform on pooled RNA samples from microdissected whole cerebellar tissues (3–10) from B6 and D2 strains at 24-hour intervals from E12 to P0 and then every 3 days until P9 (Fig. 1A). We collected our samples from timed-matings to minimize developmental noise. The reliability among biological replicates (typically $N=3$) was examined by calculating the Pearson correlation coefficient between replicates in each group. The median correlation coefficient was 0.99 and the chips with correlation coefficient of lower than 0.97 were discarded (Fig. 1B). After eliminating individual chips with low fidelity, each time point contained 3 biological replicates except the E14 time point of B6 ($N=2$) and P9 of D2 ($N=1$). In the time series data, of the 46,632 probe sets in the Illumina platform, 24,257 probes were significantly different (factorial ANOVA, false discovery rate 5%) from the median signal intensity at one or more time points, which indicates that more than half of the transcripts (52%) queried with the microarray platform were dynamically expressed during cerebellar development. In contrast, only 7141 probes showed different temporal expression pattern between the two strains (false discovery rate 5%). These differentially regulated genes are the candidate genes that may underlie strain differences in cerebellar development and function.

Hierarchical clustering analysis was performed on the time series data and clearly divided the data into three groups (Fig. 1C): early embryonic stage (E12–E14), late embryonic (E15–E19) and postnatal stage (P0–P9). These groupings likely reflect major changes in the transcriptome landscape of the developing cerebellum that are reflected in significant morphogenetic changes associated with development. To explore this further, we performed a Gene Ontology enrichment analysis with the top 150 genes that showed upregulation in each time window compared to the rest of time points (Table 1). At E12–E14, the ventricular and rhombic lip neuroepithelia produce the Purkinje and future granule cells, respectively, and the cerebellar nuclear neurons leave the rhombic lip region to colonize the nuclear transitory zone. The enrichment analysis showed that the GO terms such as transmission of nerve impulse, cell–cell signaling and cellular ion homeostasis were highly enriched during this time window (Table 1). Cell–cell signaling would be required for appropriate neuronal specification and migration (Goldowitz and Hamre, 1998; Wang and Zoghbi, 2001), and neural transmission and ion homeostasis is critical for neuronogenesis (Wang and Kriegstein, 2009); two key developmental events occurring in this time frame. At E15–E19, the future granule cells colonize the external germinal layer (EGL) via tangential migration and cerebellar nuclear neurons and Purkinje cells have migrated to form the cerebellar nuclei and Purkinje cell plate, respectively. Interestingly, the number of differentially regulated genes in this period ($n=1248$) was smaller compared to the early embryonic ($n=1962$) and postnatal periods ($n=1817$) and many of the GO terms enriched during this time window were related to chromatin/chromosome organization (Table 1). The development of the large cerebellar nuclear neurons and Purkinje cells may need upregulation of nucleosome assembly or chromosome organization-related genes (Martou and De Boni, 2000; Vadakkan et al., 2006). Finally, the P0–P9 period is marked by dendritic expansion of Purkinje cells and the elaboration of parallel fiber axons by the numerous granule cells that start to make synapses with Purkinje cells. In support of these processes, the GO terms enriched during postnatal time points were neuron differentiation, neuron development and cell migration (Hatten et al., 1997; Goldowitz and Hamre, 1998; Wang and Zoghbi, 2001) (Table 1). Thus, the hierarchical clustering and GO term enrichment analyses highlight the major temporal divisions of cerebellar development and ongoing biological processes in each time window. To facilitate resource sharing, the CbGRiTS microarray data resource is publicly available at www.cbgrits.org (GEO accession number GSE60437).

Temporal expression pattern of cerebellar neuronal type markers

As our microarray time series data demonstrated, more than half of the genes tested were dynamically regulated during cerebellar development. The developmental time and time window of gene expression would presumably be an indication of a

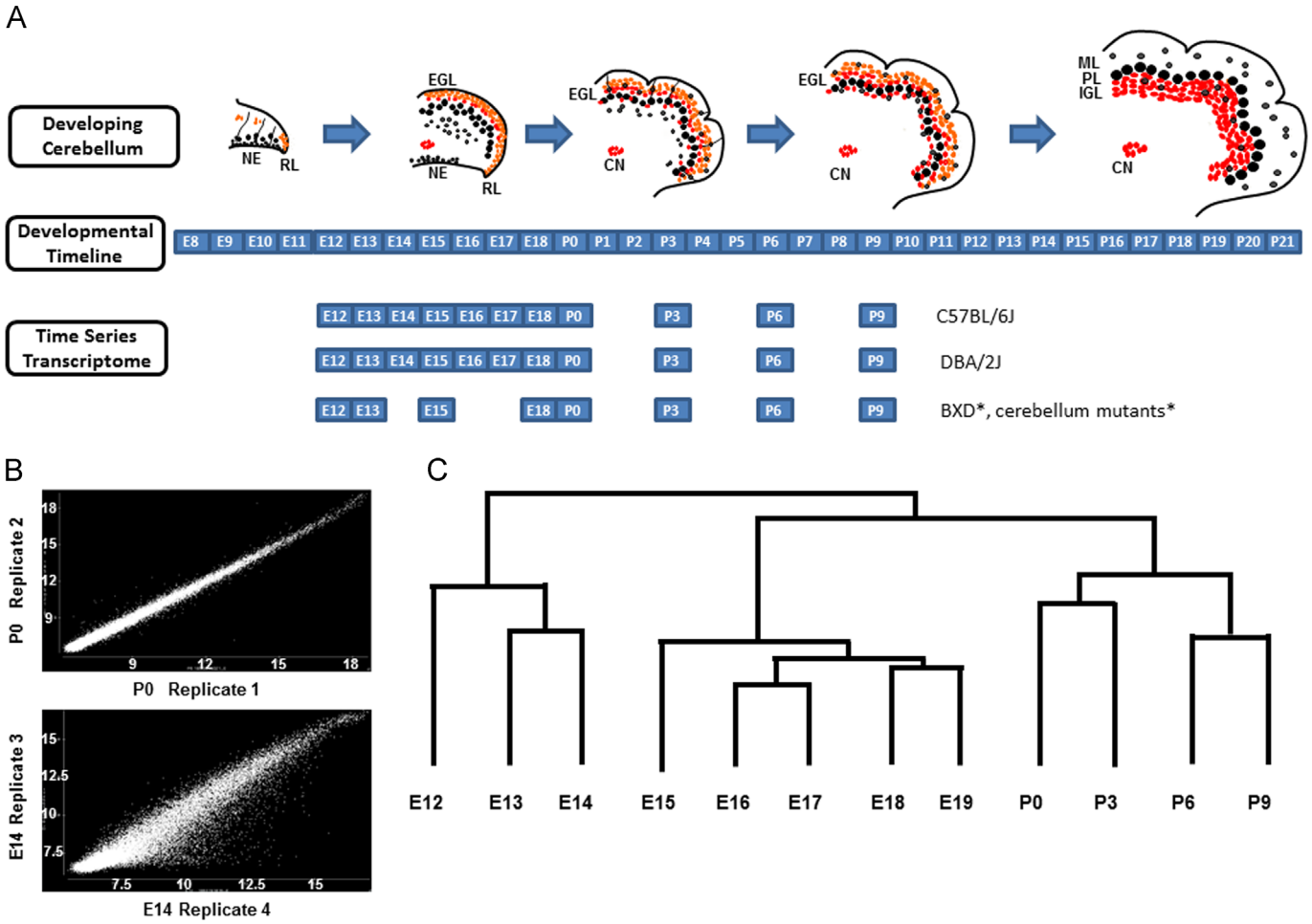


Fig. 1. Overview of Cerebellar Gene Regulation in Time and Space (CbGRiTS) time series transcriptome dataset. (A) Timeline and portrayal of cerebellar development over embryonic and postnatal development and the time points selected for time series microarray data collection. *The transcriptome data of BXD lines and cerebellar mutants are limited to a few representative time points (BXD lines and Pax6 mutant) or a single time point (E15.5 for Math1 and Mea mutant). Large black dots: Purkinje cells, small black dots: molecular layer interneurons, red dots: cerebellar granule cells and cerebellar nuclear neurons, orange dots: granule cell precursors. CN – cerebellar nuclei, EGL – external germinal layer, IGL – internal granular layer, ML – molecular layer, NE – neuroepithelium, PL – Purkinje cell layer, RL – rhombic lip. (B) As a quality control step, biological replicate chips were compared and correlation coefficients were calculated. The correlation coefficient ranged from 0.92 to 0.99. We removed chips with correlation coefficient value lower than 0.97 from our analysis, and this resulted in discarding 15% of total chips. An example of a high quality array chip ($r=0.99$, upper graph) vs. a low quality chip ($r=0.96$, lower graph, discarded) is shown. *X* and *Y* axes are $\log_2$-transformed microarray intensity values. (C) Hierarchical clustering analysis was performed on the C57BL/6J (B6) and DBA/2J (D2) time series cerebellum microarray data. The analysis produced three subgroups of early embryonic (E12.5–E14.5), late embryonic (E15.5–E19.5) and postnatal stage (P0–P9). Both strains produced the same clustering outcome and only the clustering result of B6 strain is shown.

Table 1
Gene ontology analysis for the enriched genes in three time windows of cerebellar development.

Time window	GO term	# of genes	Fold enrichment	*P*-value
Early embryonic E12–E14	Transmission of nerve impulse	15	8.64	1.55E−09
	Cellular ion homeostasis	14	7.07	7.14E−08
	Cell–cell signaling	13	5.83	1.96E−06
	Neuron differentiation	10	3.30	0.003
	Regulation of apoptosis	10	2.37	0.02
Late embryonic E15–E19	Regulation of transcription, DNA-dependent	23	2.86	5.84E−06
	Chromosome organization	14	6.46	1.98E−07
	Cell cycle	10	2.29	0.02
	Nucleosome assembly	9	20.29	1.14E−08
	RNA splicing	6	4.17	0.01
Postnatal P0–P9	Neuron differentiation	12	5.29	1.28E−05
	Neuron development	10	6.00	3.62E−05
	Neuron projection development	9	7.28	2.85E−05
	Cell migration	8	5.86	3.89E−04
	Cell morphogenesis involved in neuron differentiation	7	6.71	5.66E−04

gene's potential role during development, for example, when their expression pattern is overlaid with a Gantt chart of the developing cerebellum (Fig. 2A; see modeling temporal expression pattern). The availability of many online in situ databases such as Allen Brain Atlas (Lein et al., 2007), GenePaint (Visel et al., 2004) and GENSAT (Heintz, 2004) greatly increases the utility of the time series dataset and both databases complement each other nicely for the purpose of exploring gene expression pattern over time and tissue.

To demonstrate the utility of time series approach and validate microarray dataset, we used qRT-PCR to examine the temporal expression pattern of cerebellar neuronal type markers and transcription factors; Pax6 for granule cell, Pvalb and Rora for Purkinje cell and Tbr1 for cerebellar nuclear neurons (Fig. 2B). The expression pattern of Pax6 during cerebellar development steadily increased over time with a plateau at E18, which correlated well with the proliferative nature of granule cells in cerebellum (Sato et al., 2008; Ha et al., 2012). In contrast, the Purkinje cell marker, Rora, showed rapid increase in expression level from E12 to E14 and remained at a high level through postnatal time points. Interestingly, the expression of Pvalb, thought to be a Purkinje cell marker in mature cerebellum, had a bimodal pattern of expression (Fig. 2B). The first wave of expression peaked at E15.5 in the EGL and slowly waned until birth while a second wave of expression occurred postnatally and the expression was localized in Purkinje cells (Allen Brain Atlas). The expression pattern of Tbr1, a cerebellar nuclear neuron marker, shows a rapid increase and peak at E14.5 and then a gradual decrease over time (Fig. 2B). These examples of dynamic temporal expression pattern of important neuronal markers and transcription factors are difficult to discover and demonstrate high temporal resolution and high quality expression data as the strength of CbGRiTS time series dataset for a community resource.

Application of informatic analyses with CbGRiTS data

Here we describe three informatics analyses applied to the CbGRiTS data: differential equation modeling, paraclique analysis, and comparative dynamical system modeling. All three take advantage of time series aspect of the CbGRiTS data resource.

Modeling temporal expression pattern during cerebellar development

Three analytical approaches were employed to illustrate the use of the temporal aspect of CbGRiTS time series data. First, we employed differential equation modeling (DEM) to examine expression patterns in the B6 and D2 time series data and predict turn-on and -off times (inflection points) for dynamically expressed genes (Sasik et al., 2002) (Fig. 3A). The application of

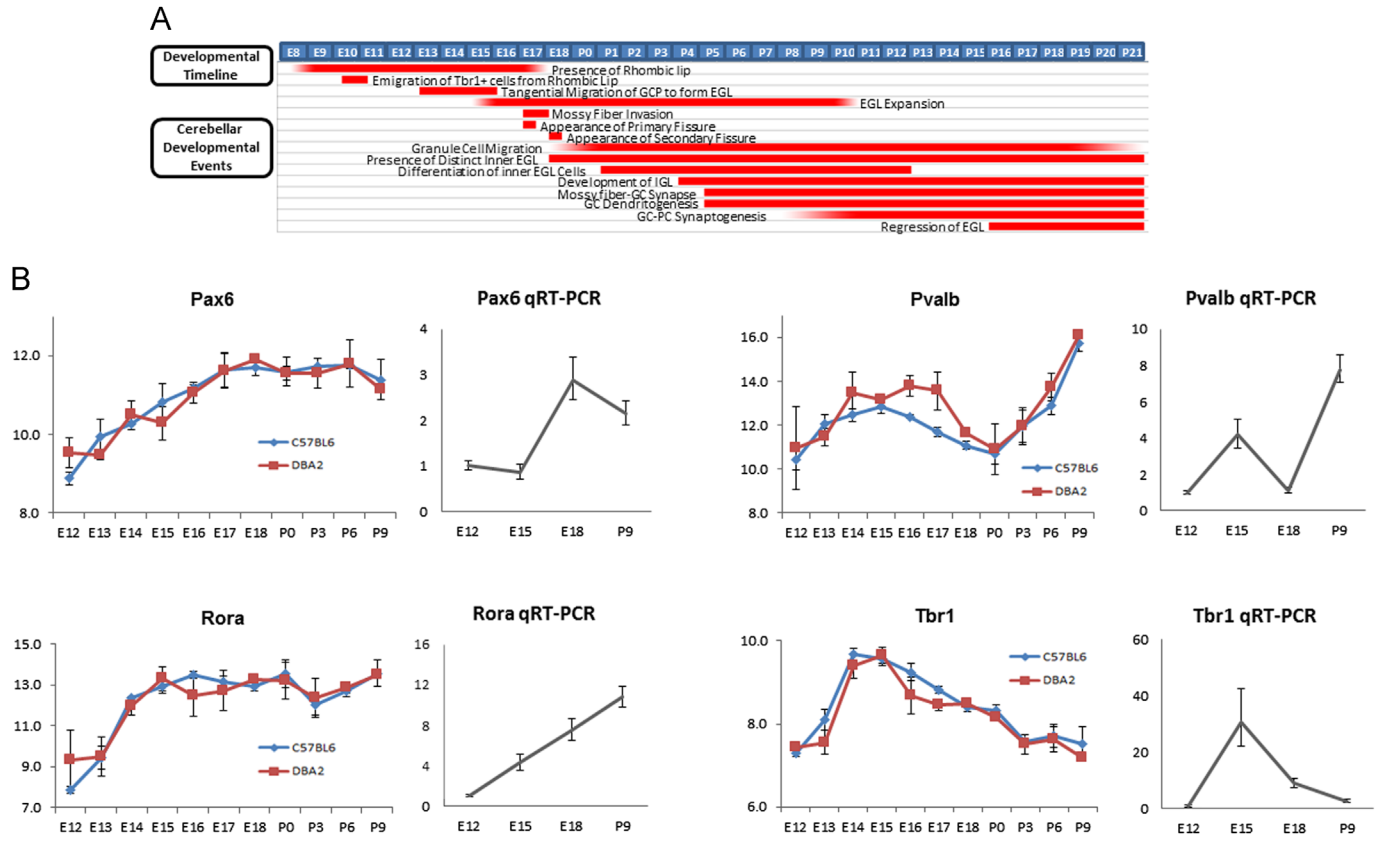


Fig. 2. The time series transcriptome dataset links developmental time with events in cerebellar development with gene expression. (A) Gantt chart of developmental processes in cerebellum. The cerebellum is an ideal system to discover novel developmental regulatory genes by using the temporal correlation between gene expression and developmental events due to the extended developmental time window and extensive knowledge on developmental processes and events. The developmental events and processes are by no means exhaustive but are listed for illustrative purposes. (B) Temporal expression patterns of neuron-specific genes in cerebellar development and qRT-PCR validation. To the left of each gene is the temporal expression profile generated from CbGRiTS and to the right is the qRT-PCR expression pattern at specific time points (E12.5, E15.5, E18.5, and P9) as a validation of microarray dataset. The selected genes have cell-type specific expression and exhibit dynamic expression pattern over cerebellar development (Pax6 – granule cell precursors, Rora – Purkinje cells, Pvalb – EGL during embryonic time points and Purkinje cells in postnatal time points, Tbr1 – cerebellar nuclear neurons). The microarray data and qRT-PCR showed the similar temporal expression pattern over time. Values on *X*-axis are time points and *Y*-axis $\log_2$ transformed, followed by a $2Z+8$ *Z*-score stabilized intensity values for microarray dataset (left panel) and relative quantity values with E12.5 expression level as 1 for qRT-PCR data (right panel).

DEM on the B6 and D2 cerebellum data produced turn-on and -off times for 564 genes with a $z < 0.1$.

To demonstrate the utility of the DEM approach, we tested the hypothesis that the expression of genes involved in the specification of early rhombic lip derivatives would coincide with the developmental time window for genesis of cerebellar nuclear neurons and granule cell precursors that arise from the rhombic lip. For this, we queried the database for genes with significant on/off inflection points of E10.5–12.5 and E12.5–14.5, respectively (Miale and Sidman, 1961; Machold and Fishell, 2005). This analysis yielded a list of 150 potential candidates, whose expression time window fit the designated on/off inflection points (Fig. 3B). We further reduced our candidate list to 5 potential regulators based on congruence between the B6 and D2 temporal expression patterns, more stringent *Z*-score and overall higher expression level. Subsequent qRT-PCR analyses demonstrated concordance of temporal expression pattern of CbGRiTS microarray data for the selected genes (Fig. 3C) We used in situ hybridization against these five genes to test whether they were expressed in the rhombic lip and/or EGL – two appropriate regions of the cerebellum for cerebellar nuclear neurons and granule cell precursors. Our analyses confirmed the localization of these transcripts to the rhombic lip and/or EGL (Fig. 3D).

Paraclique analysis to cluster genes with similar expression profiles during cerebellar development

Next, we explored our data to identify clusters of genes that shared similar expression patterns with the idea that they may be related to common developmental events in the cerebellum. To this end, we employed paraclique analysis (Chesler and Langston, 2006) to identify sets of genes that share the same temporal expression pattern. The paraclique algorithm clusters transcripts based on correlated temporal expression patterns using graph-based methods (Voy et al., 2006), with the underlying premise that common temporal expression patterns could be due to

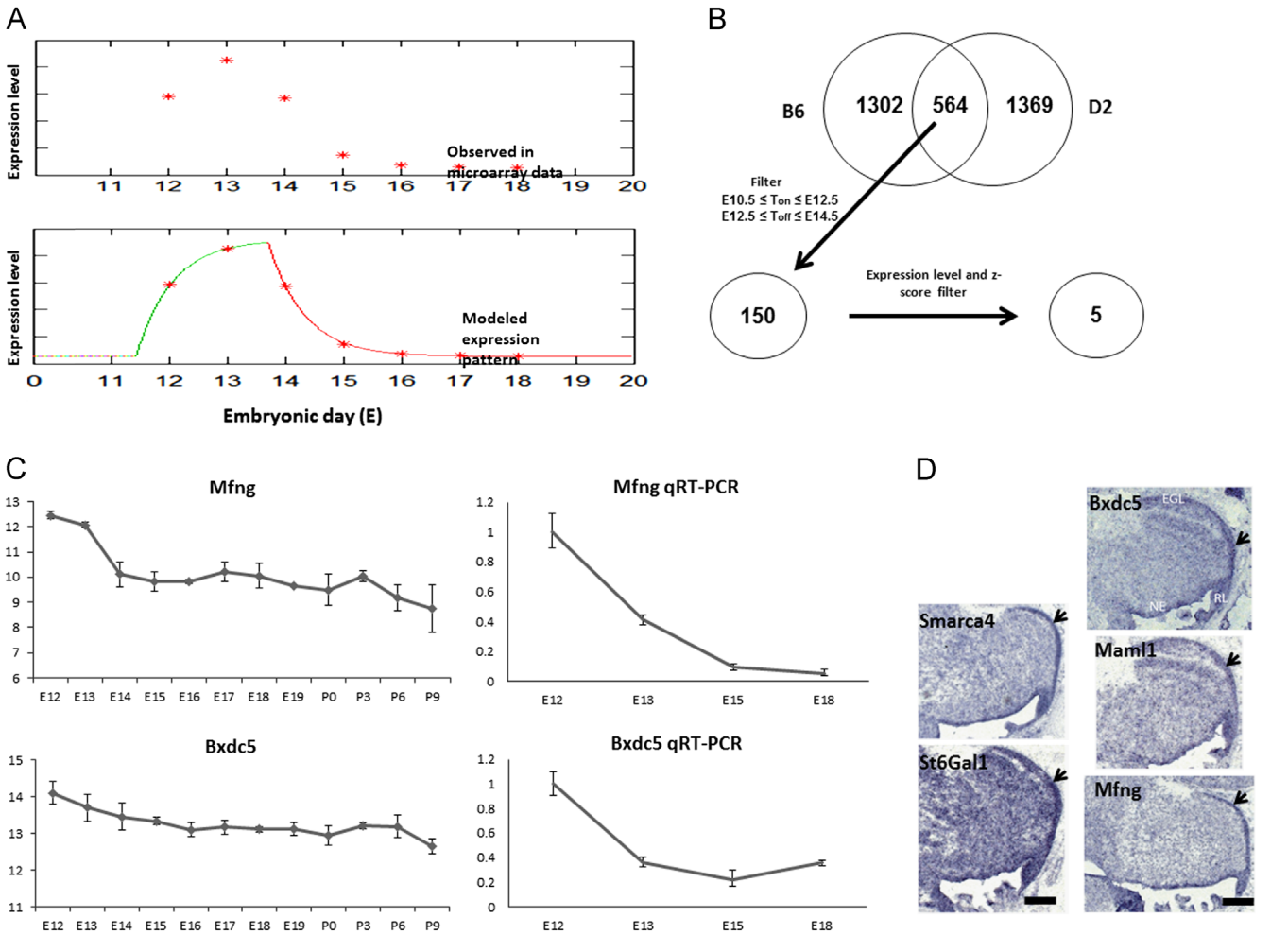


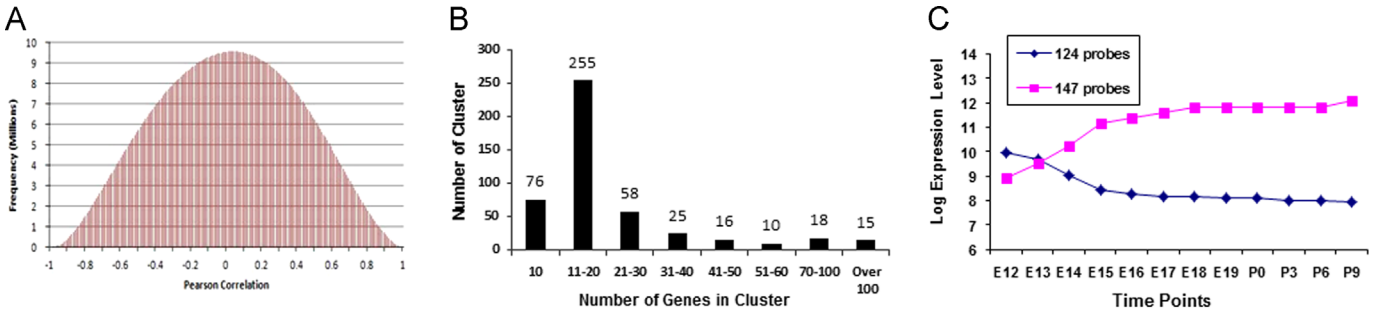
Fig. 3. Prediction of initiation and termination of gene expression using differential equation modeling (DEM) analysis. (A) The gene expression dynamics from the time series microarray data were modeled with a first-order differential equation and the analysis produced predictions of gene turn-on and -off times. The upper panel provides the observed temporal expression pattern. The lower panel is the DEM model fit of the dynamic expression pattern. (B) To test the utility of the DEM approach, we pursued genes that could be related to the specification of early rhombic lip derivatives, including cerebellar nuclear neurons and granule cell precursors that colonize the EGL. To this end, we input the on and off time window of E10.5–12.5 and E12.5–14.5 to Transcriptome Explorer to identify genes with significant on/off inflection at points that coincide with the genesis of cerebellar nuclear neurons and granule cell precursors. This analysis yielded a list of 150 candidate regulators of neuronal specification for both cell types. Expression level and more stringent *Z*-score filter narrowed our candidate list to 5 potential regulators of cell specification. (C) qRT-PCR analysis was performed on cDNAs derived from wildtype C57BL/6J null cerebella at E12.5, E13.5, E15.5, and E18.5. Left panel is the expression pattern from CbGRiTS microarray dataset and the right panel is the qRT-PCR expression pattern on selected time points. Values on *X*-axis are time points and *Y*-axis $\log_2$ transformed, followed by a $2Z+8$ *Z*-score stabilized intensity values (left panel) and relative quantity values with E12.5 expression level as 1 (right panel). The microarray data and qRT-PCR showed the same trend of down-regulation over time. (D) In situ analysis of candidate genes. We performed in situ hybridization analysis of the five candidate genes and found that all five genes showed positive expression in the EGL of the developing cerebellum at E15.5. Arrows indicate positive in situ staining in the EGL. Scale bar: 100 μm.

common developmental processes and/or transcriptional control mechanisms. Paraclique, *k*-Clique Communities, and other clique-based clustering algorithms generally tend to produce superior results (Eblen et al., 2009; Jay et al., 2012) and are highly resistant to false positives, as compared to commonly used, and less computationally demanding, clustering methods.

Paraclique analysis produced 473 clusters with 12,074 transcripts for B6 data, and 469 clusters with 10,095 transcripts for D2 data, using a correlation coefficient threshold of 0.9 (Fig. 4A). The cluster size ranged from 10 to 381 transcripts (Fig. 4B). Each paraclique (cluster) represents a group of transcripts whose expression pattern is highly correlated, either positively or negatively, over time. To illustrate this pattern, we generated a "signature" for each paraclique by averaging the expression value of each paraclique element at each time point. Such a signature shows the expression profile across time for the paraclique as a whole (Fig. 4C). In order to examine the hypothesis that high correlation of temporal expression patterns among paraclique members is due to common developmental processes, we utilized the DAVID Bioinformatics Database (Huang et al., 2009) to perform Gene Ontology (GO) enrichment analysis. Many of the paracliques showed enrichment for brain and development related categories, indicative of common developmental processes and/or transcriptional control mechanisms (Fig. 4D). Further, detailed anatomical analysis with ISH and ISH databases revealed that many of the genes have expression in the same cell type or in cells derived from the same progenitor pool (Fig. 4E).

A network of differential and conserved gene interactions in cerebellar development

Finally, we explored the predictive interactions of genes by using a recently developed comparative dynamical system modeling (CDSM) approach (Ouyang et al., 2011) which identifies causal relationships or "interactions" by an ordinary differential equation that calculates the expression rate of a target gene (effect) as a function of expression level of other regulatory genes (cause). CDSM provides a statistical testing framework to determine if interactions between genes are "conserved" over time or "differential" among observed time windows. We focused our analysis on



D

GO Term	Pvalue
Neuron Differentiation	1.83E-09
Forebrain Development	1.10E-08
Pattern Specification Process	4.43E-08
Neuron Fate Commitment	6.93E-08
Transmission of Nerve Impulse	7.48E-08
Cell Fate Commitment	1.71E-07
Synapse	1.09E-06
Synaptic Transmission	1.35E-06
Regionalization	1.57E-06
developmental protein	2.48E-06

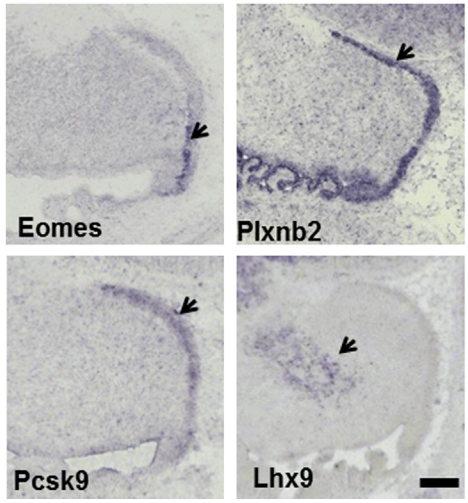


Fig. 4. The paraclique-based clustering approach identified groups of genes with similar temporal expression patterns over cerebellum development. (A) The distribution of Pearson correlation coefficients across all pairwise comparisons between microarray probes in the B6 strain. For paraclique-based clustering, we calculated all pairwise Pearson correlations between probes in the microarray data across developmental time points. As expected, both the B6 and D2 data showed a reasonably normal probability distribution. Based on the distribution, we used a threshold of correlation coefficient of 0.9 and a proportional glom factor of 0.9 to generate paracliques. (B) The distribution of number of genes in paracliques. 70% of paracliques had 20 genes or fewer and only 15 paracliques contained more than 100 genes. (C) Signature expression pattern of paraclique 3. In a paraclique, all genes have high correlation with each other across time. These correlations are both positive and negative. A paraclique can thus be divided into two groups of correlates. Genes in one group have high positive correlation with each other and high negative correlation with genes in the other group. A signature expression pattern is the average expression value of all the members of each group at each time point, and represents the expression profile across time for the cluster as a whole. Because of the negative correlations, the behavior of two groups of genes in a paraclique tends to appear as mirror images of each other. Each paraclique exhibited a distinct expression profile, with different paracliques exhibiting widely varying signatures. (D) Gene ontology-based enrichment analysis of paraclique 3. The top 10 most enriched GO terms with lowest p-values are shown. Paraclique 3 from the B6 strain contains 271 members including Eomes, Plxnb2, Pcsk9 and Lhx9 and the GO analysis of the cluster showed enrichment for brain- and development-related categories. Although certain portions of the correlation in the temporal domain could be due to general developmental processes such as cell/tissue growth and proliferation, the highly significant p-values for brain-specific categories indicate that a large portion of paraclique 3 members may have common regulatory mechanisms. (E) In situ analysis of paraclique 3 member genes: Eomes, Plxnb2, Pcsk9, and Lhx9. These genes in paraclique 3 show high correlation of expression pattern over time. Though the spatial expression patterns of these genes seemed different at E15.5, they are all expressed in the cell types that emerge from the rhombic lip. Scale bar: 100 μm.

1435 known transcription factors (TFs) in the mouse genome (Fulton et al., 2009; Vaquerizas et al., 2009). First, the entire time series data were grouped into three developmental stages – early embryonic (EE, E12.5–E14.5), late embryonic (LE, E15.5–E18.5) and postnatal (PN, P0–P9) (see Fig. 1C) and the full set of TFs were clustered based on their expression patterns. By applying CDSM on time series data, we identified conserved and differential interactions among gene clusters across three developmental stages. For the B6 strain, 669 significant differential and 83 conserved interactions were detected, while for the DBA strain, the CDSM analysis revealed 733 significant differential and 140 conserved interactions among the gene clusters (p-value ≤ 0.05). These identified interactions constituted a predicted transcriptional regulatory network underlying cerebellar development (Fig. 5A). To test the validity of these gene interactions, we utilized the Pax6^{SEY} mutant microarray data in CbGRiTS. We considered each regulatory relationship (not necessarily direct) from Pax6 to its target genes identified from the mutant microarray data and examined whether such interactions were also present in the CDSM analysis result. For the majority of differentially expressed TFs in the Pax6^{SEY} cerebellum, at least one transcript of each target gene was indeed included in a cluster downstream of the Pax6-containing cluster in the CDSM predicted networks (Fig. 5B). Specifically, we found that Pax6 positively regulates Tbr1 and Eomes, key genes in cerebellar development (Fink et al., 2006; Ha et al., 2012). The in situ analysis of Tbr1 and Eomes in Pax6^{SEY} cerebellum revealed marked downregulation in mutant compared to wild-type cerebellar tissue during embryogenesis (Fig. 5C).

Cerebellar gene regulation in time and space (CbGRiTS) web-based resource

In order to increase the utility of our time series transcriptome as a resource for researchers, we created an online tool to utilize our transcriptome data and informatics analyses. The CbGRiTS database contains the main cerebellar developmental time series microarray data and statistical analyses for B6, D2 and 14 complementary BXD RI mouse strains at 3–7 representative time points (E12.5, E15.5, E18.5, P0, P3, P6 and P9) as well as the transcriptome data and analyses for three mutants that target the granule cell population [the *Math1 KO* (E15.5), *Pax6 KO* (E13.5, E15.5 and E18.5), and *meander tail* (E15.5)]. To allow for the exploration of these transcriptome data and analyses, we built several features into the database (Fig. 6). First, we incorporated three bioinformatics analyses into the CbGRiTS database (see above) that allow researchers to explore the time series expression data and construct or test hypotheses regarding a transcript's involvement in specific processes of cerebellar development. This can be achieved by employing either an interactive, developmental event-centric approach to go from a specific developmental event or process to a refined list of candidate genes through an iterative process or a gene-centric approach where the user is provided with pre-defined lists of genes with common temporal expression patterns or predicted causal relationships (Fig. 6). To facilitate the usage of the DEM analysis, we developed a tool called Transcriptome Explorer (Fig. 6B) which can be used to identify genes involved in temporally specific developmental processes by superimposing the time frame of developmental processes and the one predicted by DEM (Fig. 6B). In addition, to facilitate the exploration of the paraclique clustering and CDSM analyses results, we developed an advanced search functionality which allows users to retrieve paraclique or CDSM cluster memberships of interested genes from the website (Fig. 6C). Second, we incorporated a graph function to plot expression levels of genes for convenient and prompt evaluation of temporal expression patterns (Fig. 6D). Third, we built gene-list functionality for import, export and save (Fig. 6E). This feature enhances the usability of the CbGRiTS database and allows users to compare and contrast between groups of genes. Fourth, the CbGRiTS database includes a collection of phenotypic data on the cerebellum for the B6 and D2 strains and the BXD lines that have been constructed from these parental strains (Fig. 6). Histological data from the embryonic cerebellum in B6 and D2 strains and adult data that include estimates of cerebellar volume and Purkinje cell number are also available online. The combination of phenotypic data and the set

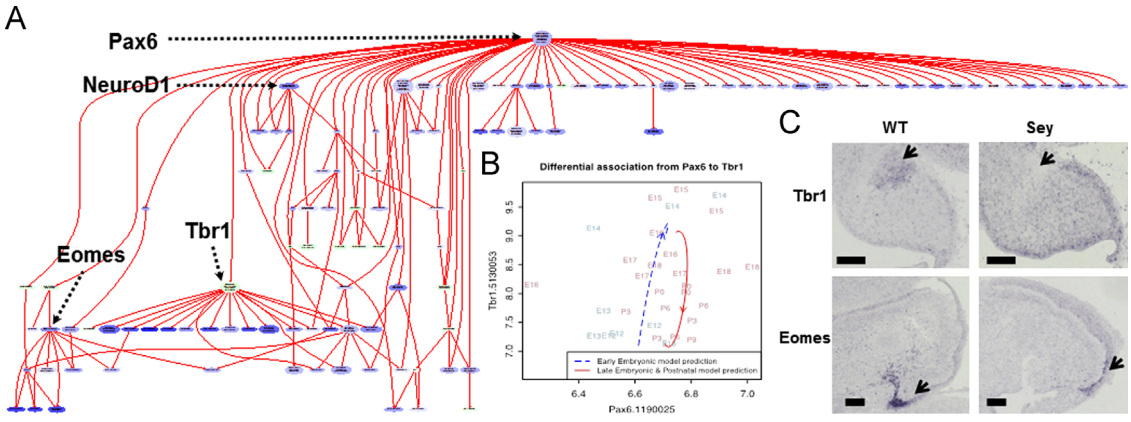


Fig. 5. Comparative dynamical system modeling (CDSM) revealed a network of conserved and differential gene interactions during cerebellar development. (A) CDSM predicted causal relationships among transcripts via their temporal expression patterns. A sub-network originating from a cluster with the transcription factor Pax6 is shown. Each circle represents a group of transcription factors (with their names and Illumina probe IDs) with similar expression patterns and each directed line represents paths detected by CDSM analysis. The network is composed of detected conserved (green) and differential (blue) interactions across (early and late) embryonic and postnatal development of the cerebellum. The network analysis with the differentially regulated genes from Pax6^{SEY} microarray from the CbGRiTS database showed that at least one path (not necessarily direct) linking the differentially regulated genes from the Pax6^{SEY} microarray data to the Pax6-containing cluster was identified in the CDSM result. The advanced search function can also be used to explore CDSM predicted gene interaction networks. (B) The predicted phase transition diagram of Pax6 and Tbr1 reveals their differential associations demonstrated by their dynamics during two developmental stages – early embryonic (blue) and late-embryonic-and-postnatal (red). Each point represents one replicate. For example, there are three points with label E17 indicating three replicates of measurements of the two genes at embryonic day 17. The numbers after the gene names are Illumina probe ID. Two different DSMs were obtained by CDSM for the entire system at each stage. Model simulation generated the fitted dynamics between this pair of genes during the two stages. The predicted early embryonic dynamics (blue dashed line) suggests strong up-regulation of Tbr1 associated with an increasing expression level of Pax6, while the predicted late-embryonic-and-postnatal dynamics (red solid line) indicates considerable down-regulation of Tbr1 associated with a high level of Pax6. (C) The decreased in situ gene expression of Tbr1 (E15.5) and Eomes (E18.5) in a Pax6 KO mutant mouse. These findings indicate that Tbr1 and Eomes are up-regulated by Pax6, confirming that the CDSM analysis is able to capture biologically relevant gene interactions. Scale bar: 100 μm.

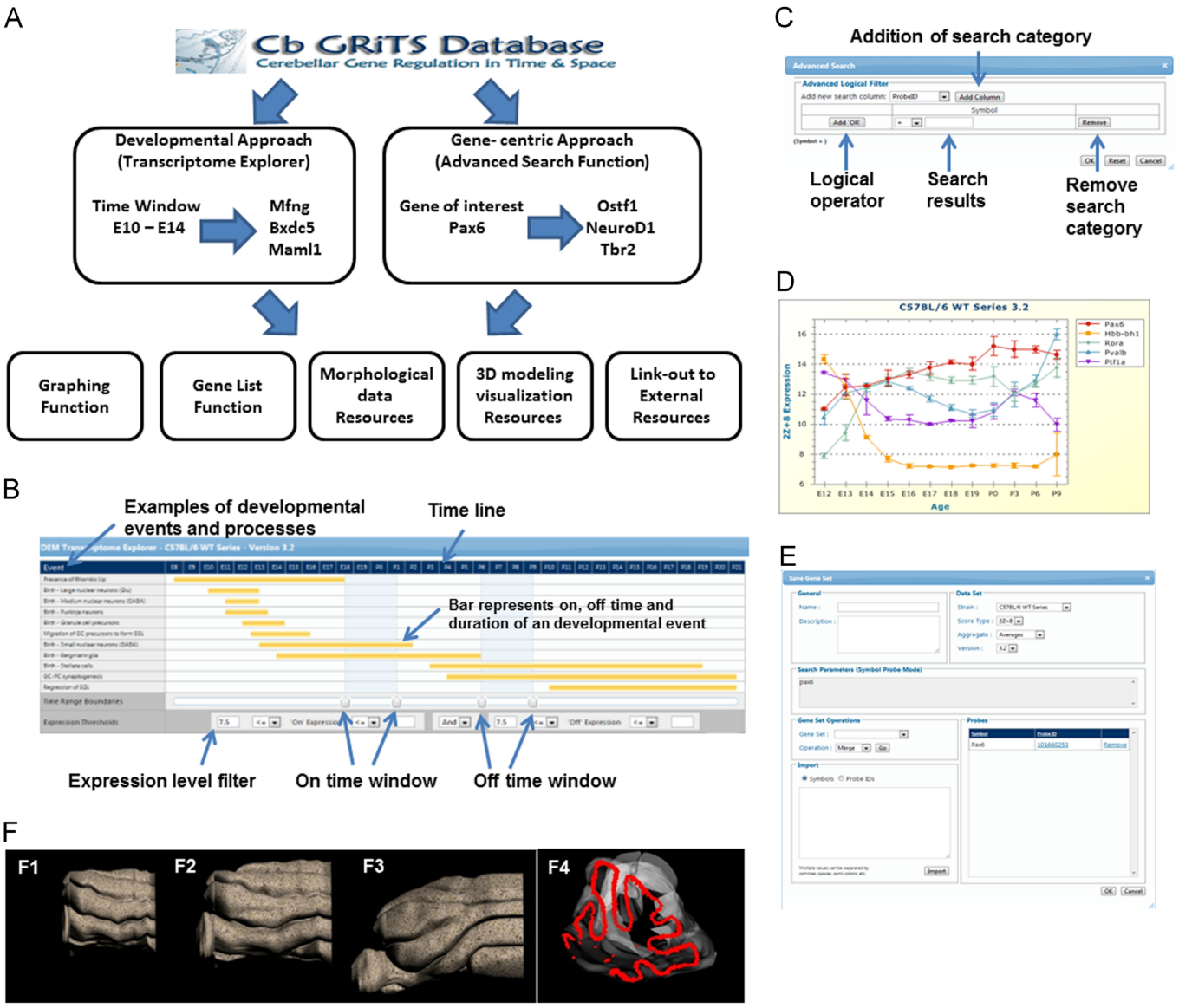


Fig. 6. The CbGRiTS web-based resources and tools. (A) The CbGRiTS database integrates the time series microarray resource, complex bioinformatics analyses, and various tools and offers two distinct ways to explore the database: a developmental event-centric approach through Transcriptome Explorer or a gene-centric approach through Advanced Search Function. Transcriptome Explorer, a web interface, allows users to start from a specific developmental event or process and obtain a refined list of candidate genes through an iterative process. Advanced Search Function enables users to take a gene-centric approach, where the user starts from a gene of interest and obtains lists of genes with common temporal expression pattern or predicted causal relationships in cerebellar development. Once a set of interesting genes is obtained, users can take advantage of various tools and functions in CbGRiTS such as graphing, gene list function, morphological data resources, 3D modeling and visualization resource, and link-out utility to external databases. Together these features and tools make CbGRiTS database a unique web-based resource that lets users examine gene expression in temporal as well as spatial domains and enables hypothesis testing and screening for novel genes and pathways in nervous system development. (B) The Transcriptome Explorer web interface enables exploration of the DEM analysis. The input is a time window of interest set by the user and the output is a list of genes whose expression pattern matches DEM-predicted turn-on and -off times. An expression level filter (e.g. one can filter out transcripts with $2Z+8$ expression level lower than 7.0) is implemented and exemplary developmental events of the cerebellum are also listed. Therefore, a researcher interested in specific developmental events or processes can designate the developmental time window and obtain possible candidate regulator genes for them. (C) Advanced search function of CbGRiTS database. The advanced search function allows users to explore paraclique clustering data and retrieve paraclique or CDSM memberships of interested genes. The expression level filter can be used in tandem to refine the search outcome. (D) An example of a plot generated by CbGRiTS. The graphing function enables users to plot the expression pattern of multiple genes online. (E) The web interface for gene list manipulation. The gene list function allows each user to create and save gene lists so that various gene lists can be compared and contrasted from different time windows and other searches and analyses. (F) CbGRiTS permits examination of spatial patterns of gene expression through development of the cerebellum with 3D model visualization. (F1–F3) Snapshots of the 3D model of cerebellar development. (F4) Snapshots of the 3D model viewer program showing a space-filled model of Calbindin expression.

of expression data in the BXD strains will serve as an entrée into a systems genetics analysis (Jansen and Nap, 2001; Brem et al., 2002; Bystrykh et al., 2005; Chesler et al., 2005) of cerebellar development. The combination and integration of these diverse biological data will create a new level of synergy in discovering novel genes and linking them to crucial developmental processes.

Fifth, we have developed an animation of cerebellar development and a prototype 3D modeling and visualization tool set that can store, visualize and compare spatial patterns of gene expression through development of the cerebellum (Fig. 6F; Visualization menu at cbgrits.org). For this, first we created a standardized set of space/time coordinates and developed the tools to fit spatial gene expression data from 3 cerebellar cell-type defining genes (*Calb* for Purkinje cells and *Pax6* and *Zic1* for granule cells) to the coordinate system. The gene expression information is stored as density values with space/time addresses, which allows the user to query

the database relative to both spatial and temporal relationships. The viewer displays the data as a 3D space-filled model that can be rotated and scaled interactively (Fig. 6F). The standardization of a space/time coordinate system of cerebellar development ensures compatibility between the existing data and future or user-submitted data entries. The 3D modeling and visualization resource provides an interactive, predictive modeling environment for exploration and visualization of multiple sets of gene expression patterns through developmental time. Lastly, we built in the link-out functionality to Allen Brain Atlas (Lein et al., 2007) and GenePaint (Visel et al., 2004) to allow exploration of in situ hybridization databases for the developing brain (Fig. 6). With these features, users can quickly examine the expression pattern of genes of interest.

Discussion

The present resource is one of the largest transcriptome databases of a single structure (over 15M gene expression measurements) in the mammalian brain over embryonic as well as postnatal time points. The public CbGRiTS database is the product of merging developmental biology with high throughput technologies and bioinformatics and offers a large resource for the neurodevelopmental biologist to explore and gain insights into the changing transcriptome over time as key cellular events shape the cerebellum. The findings from CbGRiTS can be applied to other brain region such as cortex and hippocampus as there is a conservation of genetic program between developing structures in the brain (Hevner et al., 2006). Furthermore, the anatomical ISH databases such as the Allen Brain Atlas (Lein et al., 2007), GenePaint (Visel et al., 2004) and Gensat (Heintz, 2004) make it possible to explore the spatial expression pattern of numerous genes *in silico*. Other gene-centric databases such as SynapseDB (Hadley et al., 2006), Ethanol-related gene resource (ERGR) (Guo et al., 2009), and Knowledgebase for Addiction-related Genes (KARG) (Li et al., 2008) provide lists of genes and resources for the specific area of interest. The CbGRiTS resource provides multiple datasets that are highlighted by deep time course gene expression information and enhanced with mutant and recombinant inbred strain transcriptome data as well as other phenotypic data and visualization tools. There are other existing databases on discrete brain regions such as the neocortex (Belgard et al., 2011; Bernard et al., 2012), retina (Siegert et al., 2009), hypothalamus (Shimogori et al., 2010) postnatal cerebellum (Sato et al., 2008), developing human brain (Johnson et al., 2009), and synapse-enriched samples from the mouse hippocampus (Cajigas et al., 2012). The integrative use of these databases should supply an invaluable set of resources that will allow for the analysis of brain region specificity and common motifs in the study of interacting genes or pathways.

An important feature of the CbGRiTS resource is the availability of two inbred mouse strains, C57BL/6J and DBA/2J, and recombinant inbred BXD strains. The two parental strains can be used to crosscheck for positive signals to filter out false positives. On the other hand, significant and reproducible expression profile differences between the two strains will be important validation to identifying genes responsible for strain differences. The availability of BXD lines at representative time points (E12.5, E15.5, E18.5, P0, P3, P6 and P9) opens a new door to the application of complex genetics of development using the vast adult microarray data from various regions of the brain, including the cerebellum, and the phenotypic data that largely focuses on behavior (Jansen and Nap, 2001; Brem et al., 2002; Bystrykh et al., 2005; Chesler et al., 2005). Furthermore, the CbGRiTS database is seen as a dynamic and growing web resource that accumulates quantitative anatomical data such as size of the cerebellum and number of cells in specific neuronal populations as well as other community cerebellum expression data such as the transcriptomes of Rora^{sg} mutant (Gold et al., 2003) and *Zic* KO mice (Blank et al., 2011). The intersection of genomic, transcriptomic and phenotypic datasets is a rich resource for the functional genomics of brain development (Jansen and Nap, 2001; Brem et al., 2002; Bystrykh et al., 2005; Chesler et al., 2005).

The three informatics analyses, DEM modeling, paraclique analysis and CDSM modeling, performed on the CbGRiTS resource provide representative analyses that take advantage of the time series aspect of the dataset and each has its own strengths and weaknesses. The DEM is a powerful tool for analyzing complex dynamical systems (Hirsch et al., 2013) such as cerebellar development that involves the coordinated regulation and expression of thousands of genes. It provides an explicit model for the inflection points in time series gene expression data (Zhang et al., 2012). The numerical solution of DEM for tens of thousands of genes is computationally intensive. However, accurate and robust estimation of the key parameters that characterize the dynamic gene regulation can be achieved with high-performance optimization algorithms. Clustering techniques based on graph theoretical algorithms have numerous advantages over traditional methods. Clique-centric strategies tend to be especially effective. They are resistant to false positives, and naturally accommodate pleiotropism through overlap for a systematic and comprehensive study (Jay et al., 2012). The effects of noise, which is inevitably present in experimental data, can generally be overcome by slightly relaxing the extreme stringency of cliques. We employed the paraclique algorithm (Chesler and Langston, 2006), a top-down and highly efficient noise abatement strategy roughly analogous to the well-known but considerably slower bottom-up k-clique communities method (Palla et al., 2005). Finally, comparative dynamical system modeling is sensitive to change in differential gene interactions indicative of transcription network rewiring. Through a heterogeneity *F*-statistic, it can effectively capture noisy nonlinear interactions by modeling the rate of change in gene interactions, while the few other related methods such as differential correlation (Fukushima, 2013) are not designed for dynamical systems and limited by their assumptions of linearity and simple noise characteristics, as demonstrated by Ouyang et al. (2011). Detected network association on the cerebellar gene expression time courses could thus provide valuable insight into changed transcription regulatory mechanisms during cerebellar development and across mouse strains.

We present the CbGRiTS Database as an accessible, multi-level, and interactive platform to explore and examine gene expression patterns of developmentally regulated genes in time and space and an outstanding resource for functional investigation of gene regulatory networks in cerebellar development. We integrate the time series microarray resource, bioinformatics analyses, morphological data, 3D modeling and other tools such as graphing, gene list function, and link-out utility to other anatomical ISH databases into a single web-based database. Together, these features and tools implemented in the CbGRiTS database are a powerful resource for neurodevelopmental biologists to explore data, generate hypotheses and test these hypotheses relative to the genetic underpinnings of cerebellar development.

Materials and methods

Mice and tissue collection

Females were exposed to males of the same genotype for 4 hours, midway through the light-phase, for timed-pregnant matings of C57BL/6J, DBA/2J and BXD recombinant inbred mice.

Detection of a vaginal plug defined embryonic day 0 (E0) and all subsequent gestational ages were determined from this time point. Mice were housed in a room with 12/12 h light/dark controlled environment. The use of mice was in accordance with NIH guidelines and was based on IACUC and CCAC approved protocols at the University of Tennessee HSC and the University of British Columbia.

We prepared whole cerebellar tissues for RNA expression analysis from C57BL/6J and DBA/2J strains across the full developmental time series at daily intervals from E12.5 to E18.5 and postnatally at 3-day intervals from P0 to P9. For 14 BXD recombinant inbred strains, whole cerebellar tissues were collected at 3-day intervals from E12.5 to E18.5 and P0 to P9. The cerebellum was isolated from each embryo, pooled with littermates, and snap-frozen in liquid nitrogen. Three-to-four replicate pools of 3–10 whole cerebellar samples were collected and processed for total RNA extraction. Care was taken to avoid thawing prior to homogenization of the tissue in RNASTAT60 (Tel-Test Inc., Friendswood, TX). The tissue was disrupted by passage through a 23G syringe needle and then passed over a Qiashredder column to ensure complete disruption. The homogenate was cleared by centrifugation and the RNA precipitated from the supernatant with isopropanol. RNA pellets were sent to Genome Explorations (Memphis, TN; URL: http//www.genome-explorations.com) for quality control (QC) and preparation for expression analysis. The RNA was resuspended, quantified, and the RNA integrity was verified using an Agilent Bioanalyzer 2100.

Microarray

Labeled cRNA was generated and hybridized using the Illumina Mouse 6 v1.1 and v2.0 Beadchip array. Each chip served as an independent hybridization experiment and three to four pooled samples were analyzed for each time point and strain. Once the microarray results were obtained, the Rank Invariant normalization algorithm was used in BeadStudio v3.0 to extract the expression values. These results were then $\log_2$ transformed, followed by a $2Z+8$ Z-score stabilization (Geisert et al., 2009). Scatter plots were then generated in DataDesk as an additional quality control measure to identify noisy microarrays (correlation coefficient of lower than 0.97 with other biological replicates) that were eliminated from further consideration. The quality controlled dataset was then used for all subsequent analytical procedures. The CbGRiTS dataset is also available from Gene Expression Omnibus (GSE60437). The microarray data for cerebellum mutant strains are described elsewhere (Ha et al., 2012).

In situ hybridization and qRT-PCR for analysis of candidate gene expression

For in situ hybridization (ISH), E15.5 and E18.5 embryos were decapitated and whole heads were immersion fixed for 1–2 hours in 4% paraformaldehyde made with DEPC treated water, cryopreserved in 20% sucrose, and quick-frozen in OCT (Fisher) for cryosectioning.

The cRNA probes were generated from a cDNA library generated from the E15.5 and P0 mouse brain using Invitrogen cDNA synthesis kit (Invitrogen). We used M13F and M13R primers to PCR amplify templates directly from plasmids (pGemT, Promega) that contained the target gene. The obtained PCR yield was directly used as a template to generate the Dig-labeled cRNA probe.

Embryonic tissues were sagittally cryosectioned, mounted on Superfrost™ slides (Fisher), air-dried and stored at -80 °C until use. The ISH was performed at 55 °C overnight in a hybridization buffer containing about 300 ng/ml of Dig-labeled cRNA probe following standard protocols (Ha et al., 2012). Slides were colorized with NTP/BCIP (Roche). Detailed methods of qRT-PCR are described elsewhere (Ha et al., 2012).

Analysis of cerebellar morphology

Morphologic analyses were conducted both during development and in adult animals in B6, D2 and BXD RI strains. The developmental analysis used tissue from embryos or neonates from the same litters as used for microarray analysis and thus, the same breeding schemes were used. Embryonic tissue was collected and immersion fixed in acetic acid:ethanol (1:3) while tissue from neonatal animals was fixed (in the same fixative) via cardiac perfusions. Tissue was embedded in paraffin using standard protocols and sectioned on a rotary microtome at 8 μm. Every 20th section was mounted on slides and stained with cresyl violet for analysis. The following measures were taken: (1) volume of the cerebellum as a whole, (2) volume of the EGL, (3) density of cells in the EGL and (4) density of cells in the cerebellum.

Adult animals were also intracardially perfused as described above. The cerebellum was dissected out and cryosectioned at 20 μm. One series of sections was immunolabeled with an anti-Calbindin antibody following standard avidin-biotin immunocytochemical procedures and used to determine numbers of Purkinje cells. A separate series of sections was stained with cresyl violet and used to measure the density of granule cells and molecular layer interneurons, as well as the volume of the (1) cerebellum as a whole, (2) molecular layer, (3) IGL and (4) cerebellar white matter.

All volumetric measures were taken using the Neurolucida Image Analysis software from Microbrightfield. All Purkinje cells were counted from selected sections and then used to calculate the total cell number using the Abercrombie correction factor. Density of other cell types was determined by counting all neurons within a grid and using the density and volume to estimate total number as done for Purkinje cells (Abercrombie, 1946).

Bioinformatic tools for gene ontology categorization and hierarchical clustering

For Gene Ontology category analysis of identified differentially regulated genes from microarray data analysis, DAVID bioinformatics Resources 6.7 (http://david.abcc.ncifcrf.gov/) was used (Dennis et al., 2003). The filtered gene list was entered via a web interface and the results were downloaded directly from the website. Hierarchical clustering was performed with the Hierarchical Clustering Explorer program (http://www.cs.umd.edu/hcil/hce/) with default settings. The detailed statistical approach of differential equation modeling, paraclique clustering analysis, and comparative dynamical system modeling are described in the supplemental methods.

Estimation of gene turn-on and off time with differential equation modeling

To determine the key parameters of turn-on (t_1^i) and turn-off (t_2^i) time of genes during the cerebellum development, we used differential equation modeling. Our transcriptome profiling of developing cerebellum produced a temporal sequence of expression levels $E_i(t)$ for each gene i in B6 and D2 strains. We adapted a kinetic model (Sasik et al., 2002) by allowing non-zero basal transcription and determined the critical parameters characterizing each gene expression profile.

A first order differential equation was used to model the abundance of gene transcript i at time t:

$$\frac{dA_i(t)}{dt} = R_i(t) - w_i A_i(t), \quad (1)$$

where $R_i(t)$ is a gene-specific transcription regulation term, and w_i is a gene-specific decay rate. It is reasonable to assume the transcription regulation is a sharp function of time

$$R_i(t) = \begin{cases} S_i & \text{if } t_1^i < t < t_2^i \\ B_i & \text{otherwise} \end{cases}, \quad (2)$$

where $S_i \gg B_i$, i.e. there are two levels of regulation – basal transcription B_i and stimulated transcription S_i that starts at t_1^i and ends at t_2^i.

The solution of (1) and (2) is

$$A_i(t) = \begin{cases} \frac{B_i}{w_i} & t \le t_1^i \\ \frac{S_i}{w_i}\left[1 - \frac{S_i - B_i}{S_i} e^{-w_i(t - t_1^i)}\right] & t_1^i < t \le t_i^2 \\ \frac{S_i}{w_i}\left[\frac{B_i}{S_i} + \frac{S_i - B_i}{S_i}(1 - e^{-w_i(t_2^i - t_1^i)}) e^{-w_i(t - t_2^i)}\right] & t_2^i < t \end{cases}$$

This solution is characterized by the five parameters ,t_1^i, t_2^i, S_i, B_i and w_i. To determine the parameters, we fitted the solution with the actual mRNA abundance levels $E_i(t)$ by minimizing the sum of square

$$\sum_t [A_i(t) - E_i(t)]^2$$

in the space of the five parameters. The quality of the fit was measured by

$$z = \frac{\sum_t [A_i(t) - E_i(t)]^2}{\sum_t [A_i(t) - \overline{E_i}]^2},$$

where $\overline{E_i} = \sum_t E_i(t)/N_t$ is the average expression level of gene i, N_t is the number of time point.

Clique-based clustering approach to dynamically expressed genes

To construct a data structure suitable for graph algorithm clustering, we first calculated all pairwise Pearson correlations between microarray probes across developmental time points. In most analyses, such correlations are expected to follow a reasonably normal probability distribution, and indeed, that is the case for both the B6 and D2 data (Fig. 1A). Such correlations constitute edge weights in a graph, the vertices being probes. To convert from a weighted to an unweighted graph, we selected a correlation threshold of 0.9 and retained those edges with magnitudes at or above this threshold, discarding edges with weights below the threshold. Selecting an ideal threshold for such graphs has been the subject of recent study (Zhang and Horvath, 2005; Borate et al., 2009; Perkins and Langston, 2009), but it remains for the most part analogous to the selection of p-value to determine significance. A threshold of 0.9 yields an exceptionally low correlation p-value over the 12 and 13 time points of the B6 and DBA data, which translates into the expectation that dense subgraphs will have very low rates of false positives. Once we obtained an undirected graph, we applied the paraclique algorithm(Chesler and Langston, 2006), which works by iteratively extracting first a maximum clique, then adding vertices connected to some predetermined percentage of the vertices in this clique. The percentage chosen is called the "proportional glom factor." We used a factor of 0.9 for both strains, meaning that new vertices must be connected to 90% or more of the vertices in the original maximum clique to become part of the resulting paraclique. Once such a paraclique was identified, it was removed from the graph and the process continued. Thus, in this fashion, the paraclique algorithm allowed us to obtain a collection of clusters, all the while preserving density and unraveling excessive clique overlap.

Comparative dynamical system modeling (CDSM) to identify differential and conserved interactions

CDSM is a recently developed computational method to identify conserved and differential gene interactions in time course transcriptome data (Ouyang et al., 2011), where CDSM demonstrated its effectiveness in nonlinear and noisy interactions over alternative approaches such as differential correlation. In a dynamical system model, the rate of change of a target gene is calculated by a nonlinear ordinary differential equation (ODE) of its regulators. CDSM studies two ODEs involving the same set of genes across two or more conditions. Conserved interactions share the same regulatory mechanisms represented by identical ODEs while differential ones differ in coefficients in the ODE describing interaction strength. CDSM uses a decomposition rule to derive homogeneity and heterogeneity of interactions from total interaction strength across conditions. Homogeneity and heterogeneity are two F-statistics measuring similarity and difference of an interaction across conditions based on observed time course data. The first F-statistic F_c measures homogeneity by indicating how well a single homogeneous interaction model can explain pooled data observed under differential conditions. Another F-statistic F_d measures heterogeneity by indicating how interactions under each condition deviate from the homogeneous interaction model. Another F-statistic F_t measures total activity of the interactions under both conditions. In the decomposition rule we have

$$\log(1 + r_t F_t) = \log(1 + r_c F_c) + \log(1 + r_d F_d)$$

where r_t, r_c, r_d are determined by the degrees of freedom of the F-statistics defined in Ouyang et al. (2011). We use F_t to select the most active gene interactions in a gene network. Then we label an interaction as differential if the p-value of F_d is significant ($p < 0.05$); we call an interaction conserved if it is not differential and the p-value of F_c is significant ($p < 0.05$). CDSM requires sufficiently sampled time course data for accurate estimation of the rate of gene expression.

To identify an active transcription regulatory network for cerebellar development, we included in our analysis all 1435 known mouse transcription factors (TF) that have been reported in the literature (Fulton et al., 2009). Transcripts from these TFs and their alternative splicing variants as well as duplicate copies were grouped into 291 clusters for the DBA mouse strain and 251 for the B6 strain by linear correlation of the time courses of each transcript. A representative is selected for each cluster such that it best characterized the time course of transcripts in that cluster. CDSM was then applied to the time courses of the cluster representatives.

We divided the cerebellar development time course into three stages: early embryonic (EE), late embryonic (LE) and postnatal (PN). This division enabled the comparison of interactions between EE and LE+PN, LE and EE+PN, PN and EE+LE. The same comparisons were applied to both the DBA and B6 mouse data. In CDSM analysis, we allowed each cluster to have at most two regulatory clusters and an interaction is detected at the p-value of less than 0.05. After applying CDSM on time course gene expression data during cerebellar development in the B6 and DBA mouse strains, we identified conserved and differential gene interactions. To validate the CDSM results, we compared the detected gene interactions against those obtained from an independent gene expression microarray experiment on Pax6 KO mutant mice.

Acknowledgments

Supported by NIH Grant R01 HD 52472, Genome BC, CMMT and NeuroDevNet.

References

Abercrombie, 1946. Estimation of nuclear population from microtome sections. Anat. Rec. 94, 239–247.

Andreasen, N.C., Pierson, R., 2008. The role of the cerebellum in schizophrenia. Biol. Psychiatry 64, 81–88.

Bar-Joseph, Z., 2004. Analyzing time series gene expression data. Bioinformatics 20, 2493–2503.

Belgard, T.G., Marques, A.C., Oliver, P.L., Abaan, H.O., Sirey, T.M., Hoerder-Suabedissen, A., Garcia-Moreno, F., Molnar, Z., Margulies, E.H., Ponting, C.P., 2011. A transcriptomic atlas of mouse neocortical layers. Neuron 71, 605–616.

Bernard, A., Lubbers, L.S., Tanis, K.Q., Luo, R., Podtelezhnikov, A.A., Finney, E.M., McWhorter, M.M., Serikawa, K., Lemon, T., Morgan, R., Copeland, C., Smith, K., Cullen, V., Davis-Turak, J., Lee, C.K., Sunkin, S.M., Loboda, A.P., Levine, D.M., Stone, D.J., Hawrylycz, M.J., Roberts, C.J., Jones, A.R., Geschwind, D.H., Lein, E.S., 2012. Transcriptional architecture of the primate neocortex. Neuron 73, 1083–1099.

Blank, M.C., Grinberg, I., Aryee, E., Laliberte, C., Chizhikov, V.V., Henkelman, R.M., Millen, K.J., 2011. Multiple developmental programs are altered by loss of Zic1 and Zic4 to cause Dandy–Walker malformation cerebellar pathogenesis. Development 138, 1207–1216.

Borate, B.R., Chesler, E.J., Langston, M.A., Saxton, A.M., Voy, B.H., 2009. Comparative analysis of thresholding approaches for microarray-derived gene co-expression matrices. BMC Research Notes, 2.

Brem, R.B., Yvert, G., Clinton, R., Kruglyak, L., 2002. Genetic dissection of transcriptional regulation in budding yeast. Science 296, 752–755.

Bystrykh, L., Weersing, E., Dontje, B., Sutton, S., Pletcher, M.T., Wiltshire, T., Su, A.I., Vellenga, E., Wang, J., Manly, K.F., Lu, L., Chesler, E.J., Alberts, R., Jansen, R.C., Williams, R.W., Cooke, M.P., de Haan, G., 2005. Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. Nat. Genet. 37, 225–232.

Cajigas, I.J., Tushev, G., Will, T.J., tom Dieck, S., Fuerst, N., Schuman, E.M., 2012. The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. Neuron 74, 453–466.

Chesler, E.J., Langston, M.A., 2006. Combinatorial genetic regulatory network analysis tools for high throughput transcriptomic data. In: Eskin, E. (Ed.), Systems Biology and Regulatory Genomics. Springer, New York, pp. 150–165.

Chesler, E.J., Lu, L., Shou, S., Qu, Y., Gu, J., Wang, J., Hsu, H.C., Mountz, J.D., Baldwin, N.E., Langston, M.A., Threadgill, D.W., Manly, K.F., Williams, R.W., 2005. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat. Genet. 37, 233–242.

Colantuoni, C., Lipska, B.K., Ye, T., Hyde, T.M., Tao, R., Leek, J.T., Colantuoni, E.A., Elkahloun, A.G., Herman, M.M., Weinberger, D.R., Kleinman, J.E., 2011. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature 478, 519–523.

Davidson, E.H., Erwin, D.H., 2006. Gene regulatory networks and the evolution of animal body plans. Science 311, 796–800.

Dennis Jr., G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol. 4, P3.

Eblen, J.D., Gerling, I.C., Saxton, A.M., Wu, J., Snoddy, J.R., Langston, M.A., 2009. Graph algorithms for integrated biological analysis, with applications to type 1 diabetes data. In: Chaovalitwongse, W.A. (Ed.), Clustering Challenges in Biological Networks. World Scientific, London, pp. 207–222.

Fink, A.J., Englund, C., Daza, R.A., Pham, D., Lau, C., Nivison, M., Kowalczyk, T., Hevner, R.F., 2006. Development of the deep cerebellar nuclei: transcription factors and cell migration from the rhombic lip. J. Neurosci. 26, 3066–3076.

Fukushima, A., 2013. DiffCorr: an R package to analyze and visualize differential correlations in biological networks. Gene 518, 209–214.

Fulton, D.L., Sundararajan, S., Badis, G., Hughes, T.R., Wasserman, W.W., Roach, J.C., Sladek, R., 2009. TFCat: the curated catalog of mouse and human transcription factors. Genome Biol. 10, R29.

Geisert, E.E., Lu, L., Freeman-Anderson, N.E., Templeton, J.P., Nassr, M., Wang, X., Gu, W., Jiao, Y., Williams, R.W., 2009. Gene expression in the mouse eye: an online resource for genetics using 103 strains of mice. Mol. Vis. 15, 1730–1763.

Gold, D.A., Baek, S.H., Schork, N.J., Rose, D.W., Larsen, D.D., Sachs, B.D., Rosenfeld, M. G., Hamilton, B.A., 2003. RORalpha coordinates reciprocal signaling in cerebellar development through sonic hedgehog and calcium-dependent pathways. Neuron 40, 1119–1131.

Goldowitz, D., Hamre, K., 1998. The cells and molecules that make a cerebellum. Trends Neurosci. 21, 375–382.

Guo, A.Y., Webb, B.T., Miles, M.F., Zimmerman, M.P., Kendler, K.S., Zhao, Z., 2009. ERGR: an ethanol-related gene resource. Nucleic Acids Res. 37, D840–D845.

Ha, T.J., Swanson, D.J., Kirova, R., Yeung, J., Choi, K., Tong, Y., Chesler, E.J., Goldowitz, D., 2012. Genome-wide microarray comparison reveals downstream genes of Pax6 in the developing mouse cerebellum. Eur. J. Neurosci. 36, 2888–2898.

Hadley, D., Murphy, T., Valladares, O., Hannenhalli, S., Ungar, L., Kim, J., Bucan, M., 2006. Patterns of sequence conservation in presynaptic neural genes. Genome Biol. 7, R105.

Hatten, M.E., Alder, J., Zimmerman, K., Heintz, N., 1997. Genes involved in cerebellar cell specification and differentiation. Curr. Opin Neurobiol. 7, 40–47.

Heintz, N., 2004. Gene expression nervous system atlas (GENSAT). Nat. Neurosci. 7, 483.

Hevner, R.F., Hodge, R.D., Daza, R.A., Englund, C., 2006. Transcription factors in glutamatergic neurogenesis: conserved programs in neocortex, cerebellum, and adult hippocampus. Neurosci. Res. 55, 223–233.

Hirsch M.W., Smale S., Devaney R.L. (2013) Differential Equations, Dynamical Systems, and an Introduction to Chaos. Academic Press.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13.

Huber, K.M., 2006. The fragile X-cerebellum connection. Trends Neurosci. 29, 183–185.

Jansen, R.C., Nap, J.P., 2001. Genetical genomics: the added value from segregation. Trends Genet. 17, 388–391.

Jay, J.J., Eblen, J.D., Zhang, Y., Benson, M., Perkins, A.D., Saxton, A.M., Voy, B.H., Chesler, E.J., Langston, M.A., 2012. A systematic comparison of genome scale clustering algorithms. BMC Bioinform. 13, S7.

Johnson, M.B., Kawasawa, Y.I., Mason, C.E., Krsnik, Z., Coppola, G., Bogdanovic, D., Geschwind, D.H., Mane, S.M., State, M.W., Sestan, N., 2009. Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron 62, 494–509.

Klevecz, R.R., Li, C.M., Bolen, J.L., 2007. Signal processing and the design of microarray time-series experiments. Methods Mol. Biol. 377, 75–94.

Koekkoek, S.K., Yamaguchi, K., Milojkovic, B.A., Dortland, B.R., Ruigrok, T.J., Maex, R., De Graaf, W., Smit, A.E., VanderWerf, F., Bakker, C.E., Willemsen, R., Ikeda, T., Kakizawa, S., Onodera, K., Nelson, D.L., Mientjes, E., Joosten, M., De Schutter, E., Oostra, B.A., Ito, M., De Zeeuw, C.I., 2005. Deletion of FMR1 in Purkinje cells enhances parallel fiber LTD, enlarges spines, and attenuates cerebellar eyelid conditioning in Fragile X syndrome. Neuron 47, 339–352.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., Chen, L., Chen, L., Chen, T.M., Chin, M. C., Chong, J., Crook, B.E., Czaplinska, A., Dang, C.N., Datta, S., Dee, N.R., Desaki, A. L., Desta, T., Diep, E., Dolbeare, T.A., Donelan, M.J., Dong, H.W., Dougherty, J.G., Duncan, B.J., Ebbert, A.J., Eichele, G., Estin, L.K., Faber, C., Facer, B.A., Fields, R., Fischer, S.R., Fliss, T.P., Frensley, C., Gates, S.N., Glattfelder, K.J., Halverson, K.R., Hart, M.R., Hohmann, J.G., Howell, M.P., Jeung, D.P., Johnson, R.A., Karr, P.T., Kawal, R., Kidney, J.M., Knapik, R.H., Kuan, C.L., Lake, J.H., Laramee, A.R., Larsen, K.D., Lau, C., Lemon, T.A., Liang, A.J., Liu, Y., Luong, L.T., Michaels, J., Morgan, J.J., Morgan, R.J., Mortrud, M.T., Mosqueda, N.F., Ng, L.L., Ng, R., Orta, G.J., Overly, C. C., Pak, T.H., Parry, S.E., Pathak, S.D., Pearson, O.C., Puchalski, R.B., Riley, Z.L., Rockett, H.R., Rowland, S.A., Royall, J.J., Ruiz, M.J., Sarno, N.R., Schaffnit, K., Shapovalova, N.V., Sivisay, T., Slaughterbeck, C.R., Smith, S.C., Smith, K.A., Smith, B.I., Sodt, A.J., Stewart, N.N., Stumpf, K.R., Sunkin, S.M., Sutram, M., Tam, A., Teemer, C.D., Thaller, C., Thompson, C.L., Varnam, L.R., Visel, A., Whitlock, R.M., Wohnoutka, P.E., Wolkey, C.K., Wong, V.Y., et al., 2007. Genome-wide atlas of gene expression in the adult mouse brain. Nature 445, 168–176.

Li, C.Y., Mao, X., Wei, L., 2008. Genes and (common) pathways underlying drug addiction. PLoS Comput. Biol. 4, e2.

Machold, R., Fishell, G., 2005. Math1 is expressed in temporally discrete pools of cerebellar rhombic-lip neural progenitors. Neuron 48, 17–24.

Martou, G., De Boni, U., 2000. Nuclear topology of murine, cerebellar Purkinje neurons: changes as a function of development. Exp. Cell Res. 256, 131–139.

Miale, I.L., Sidman, R.L., 1961. An autoradiographic analysis of histogenesis in the mouse cerebellum. Exp. Neurol. 4, 277–296.

Nayate, A., Bradshaw, J.L., Rinehart, N.J., 2005. Autism and Asperger's disorder: are they movement disorders involving the cerebellum and/or basal ganglia? Brain Res. Bull. 67, 327–334.

Ouyang, Z., Song, M., Guth, R., Ha, T.J., Larouche, M., Goldowitz, D., 2011. Conserved and differential gene interactions in dynamical biological systems. Bioinformatics 27, 2851–2858.

Palla, G., Derenyi, I., Farkas, I., Vicsek, T., 2005. Uncovering the overlapping community structure of complex networks in nature and society. Nature 435, 814–818.

Perkins, A.D., Langston, M.A., 2009. Threshold selection in gene co-expression networks using spectral graph theory techniques. BMC Bioinform. 10.

Pernet, C.R., Poline, J.B., Demonet, J.F., Rousselet, G.A., 2009. Brain classification reveals the right cerebellum as the best biomarker of dyslexia. BMC Neurosci. 10, 67.

Sasik, R., Iranfar, N., Hwa, T., Loomis, W.F., 2002. Extracting transcriptional events from temporal gene expression patterns during Dictyostelium development. Bioinformatics 18, 61–66.

Sato, A., Sekine, Y., Saruta, C., Nishibe, H., Morita, N., Sato, Y., Sadakata, T., Shinoda, Y., Kojima, T., Furuichi, T., 2008. Cerebellar development transcriptome database (CDT-DB): profiling of spatio-temporal gene expression during the postnatal development of mouse cerebellum. Neural Netw. 21, 1056–1069.

Shimogori, T., Lee, D.A., Miranda-Angulo, A., Yang, Y., Wang, H., Jiang, L., Yoshida, A. C., Kataoka, A., Mashiko, H., Avetisyan, M., Qi, L., Qian, J., Blackshaw, S., 2010. A genomic atlas of mouse hypothalamic development. Nat. Neurosci. 13, 767–775.

Siegert, S., Scherf, B.G., Del Punta, K., Didkovsky, N., Heintz, N., Roska, B., 2009. Genetic address book for retinal cell types. Nat. Neurosci. 12, 1197–1204.

Storey, J.D., Xiao, W., Leek, J.T., Tompkins, R.G., Davis, R.W., 2005. Significance analysis of time course microarray experiments. Proc. Natl. Acad. Sci. USA 102, 12837–12842.

Tsai, P.T., Hull, C., Chu, Y., Greene-Colozzi, E., Sadowski, A.R., Leech, J.M., Steinberg, J., Crawley, J.N., Regehr, W.G., Sahin, M., 2012. Autistic-like behaviour and cerebellar dysfunction in Purkinje cell Tsc1 mutant mice. Nature 488, 647–651.

Vadakkan, K.I., Li, B., De Boni, U., 2006. Cell-type specific proximity of centromeric domains of one homologue each of chromosomes 2 and 11 in nuclei of cerebellar Purkinje neurons. Chromosoma 115, 395–402.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., Luscombe, N.M., 2009. A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. 10, 252–263.
Visel, A., Thaller, C., Eichele, G., 2004. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. Nucleic Acids Res. 32, D552–D556.
Voy, B.H., Scharff, J.A., Perkins, A.D., Saxton, A.M., Borate, B., Chesler, E.J., Branstetter, L.K., Langston, M.A., 2006. Extracting gene networks for low-dose radiation using graph theoretical algorithms. PLoS Comput. Biol. 2, e89.
Wang, D.D., Kriegstein, A.R., 2009. Defining the role of GABA in cortical development. J. Physiol. 587, 1873–1879.
Wang, V.Y., Zoghbi, H.Y., 2001. Genetic regulation of cerebellar development. Nat. Rev. Neurosci. 2, 484–491.
Zhang, B., Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. Statist. Appl. Genet. Mol. Biol. 4 (article 17).
Zhang, P., Mourad, R., Xiang, Y., Huang, K., Huang, T., Nephew, K., Liu, Y., Li, L., 2012. A dynamic time order network for time-series gene expression data analysis. BMC Syst. Biol. 6, S9.