Supplemental Figures S1-S8

Figure S1.

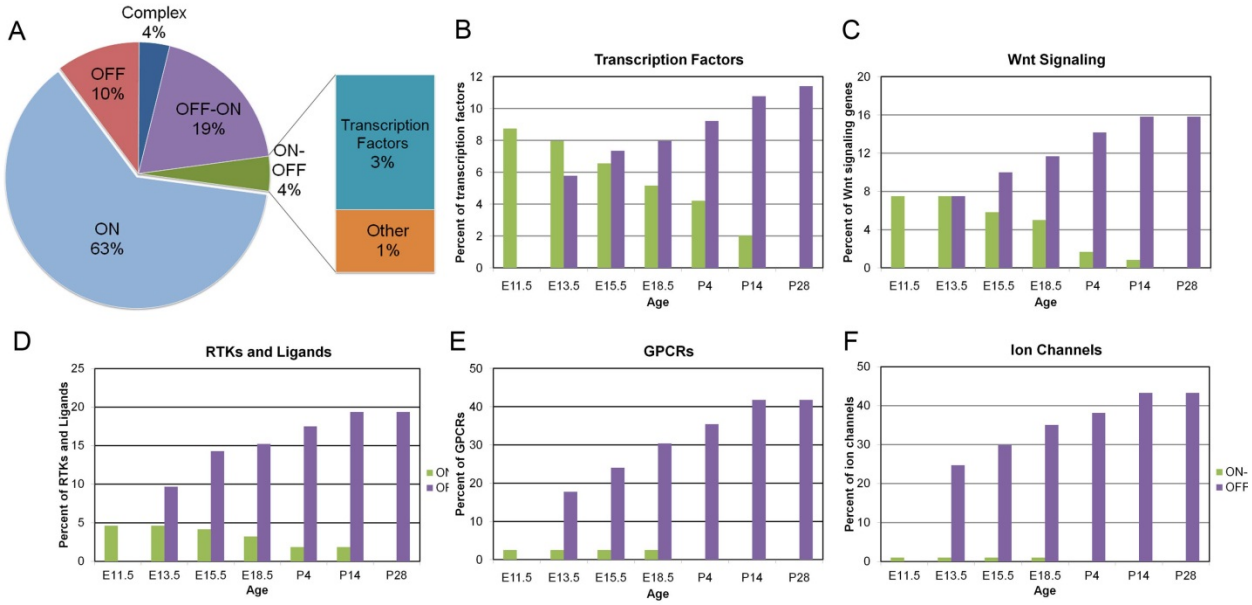

Figure S1. Temporal Gene Expression in the Brain by Class, related to Figure 3. (**A**) Percentage of genes by expression trend in brain: ON, on at all stages; OFF, off at all stages; OFF-ON, genes not expressed at E11.5 but eventually turned on; ON-OFF, genes that are expressed at E11.5 but eventually turned off; Complex patterns, e.g. OFF-ON-OFF. Note that transcription factors account for 75% of the genes in the ON-OFF category. (**B-F**), Distribution of OFF-ON (purple) and ON-OFF (green) genes shown as percentages from five gene functional categories or pathways over seven stages. (**B**) Transcription Factors, (**C**) Wnt Signaling, and (**D**) Receptor Tyrosine Kinases (RTKs) and ligands all exhibit both ON-OFF and OFF-ON trends of expression, whereas GPCRs (**E**) and Ion Channels (**F**), tend to be expressed largely later in development (OFF-ON).

Figure S2.

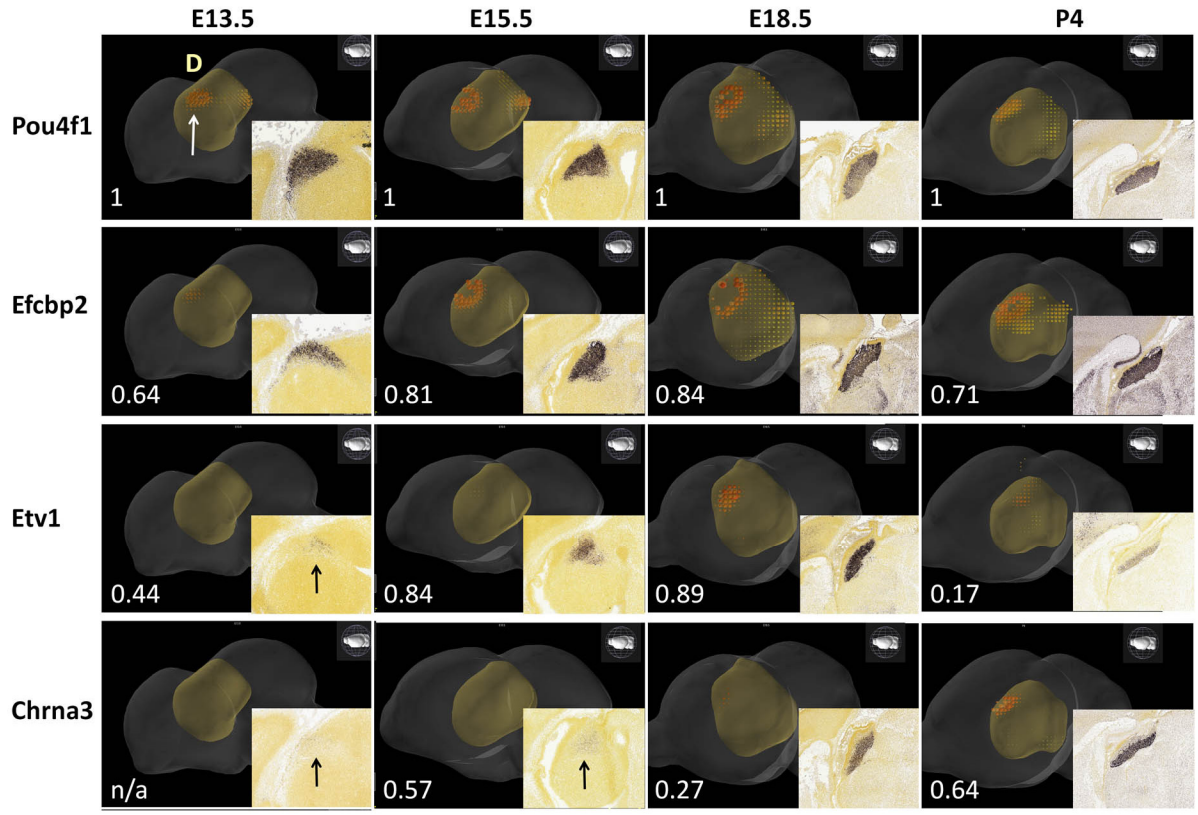

Figure S2. Using *NeuroBlast* tool to identify genes based on spatial correlation of expression, related to Figure 2. *NeuroBlast* was used to identify top search returns in the diencephalon (yellow 3D structure) at the indicated ages for the seed gene *Pou4f1/Brn3a*, a transcription factor. Expression of *Efcbp2, Etv1,* and *Chrna3* are highly correlated with *Pou4f1* across most ages. Insets show ISH in the sagittal plane for these genes in the habenula. *Etv1* at E13.5 and *Chrna3* at E13.5 and E15.5 are weakly expressed (arrow indicates area of incipient expression). The correlation between each gene and *Pou4f1/Brn3a* is given in the lower left.

Figure S3.

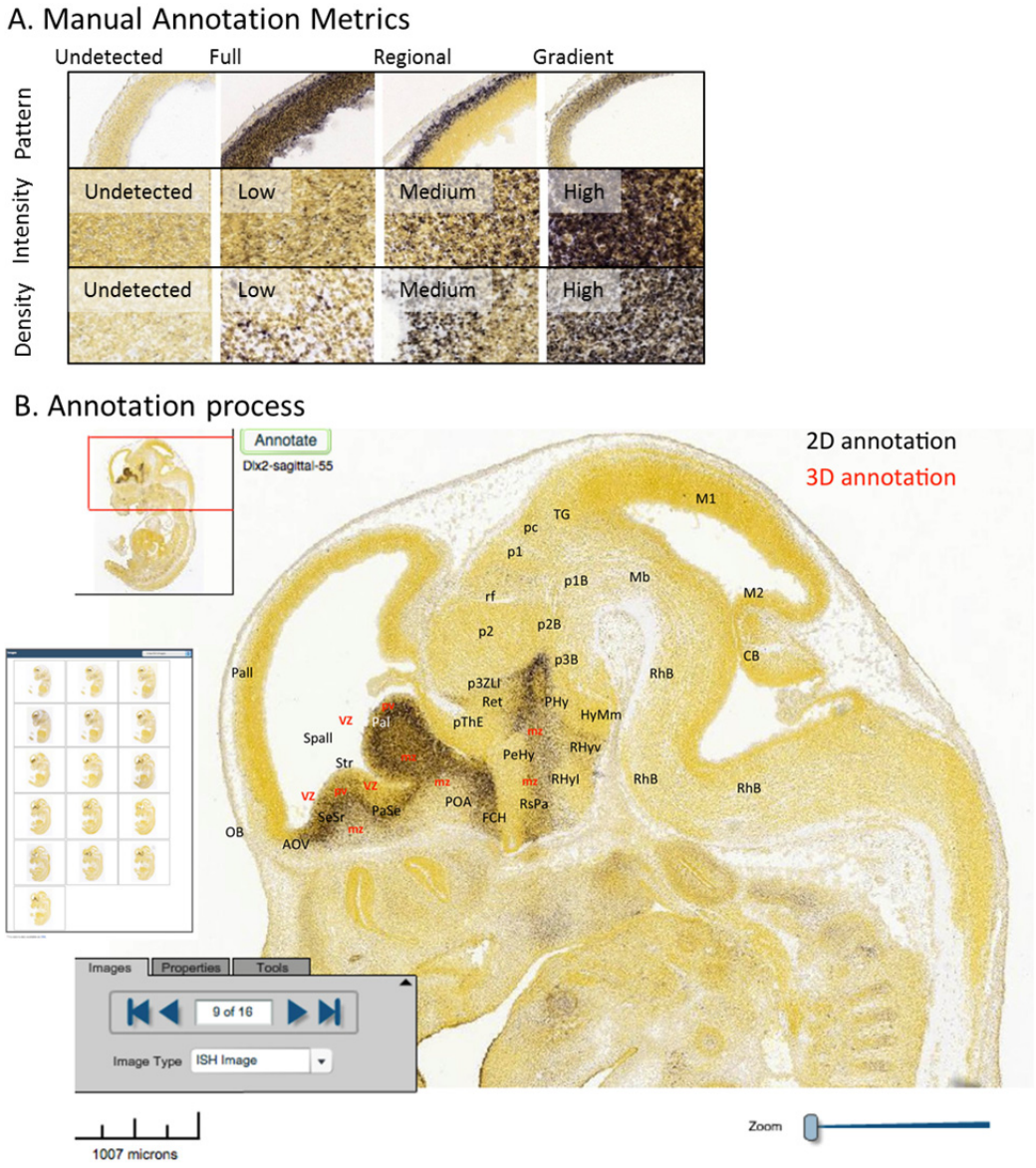

Figure S3. Manual annotation of embryonic gene expression data, related to Figure 2. (**A**) Metrics used to annotate expression patterns included density, intensity, and pattern. (**B**) The annotation process began with viewing all images in a series, and opening a separate window to record metrics for each structure in the ontology. The “2D annotation” refers to an initial assignment of expression to a structure, and the “3D annotation” refers to further assignment of expression to strata within that region (*e.g.*, ventricular zone, mantle zone).

Figure S4.

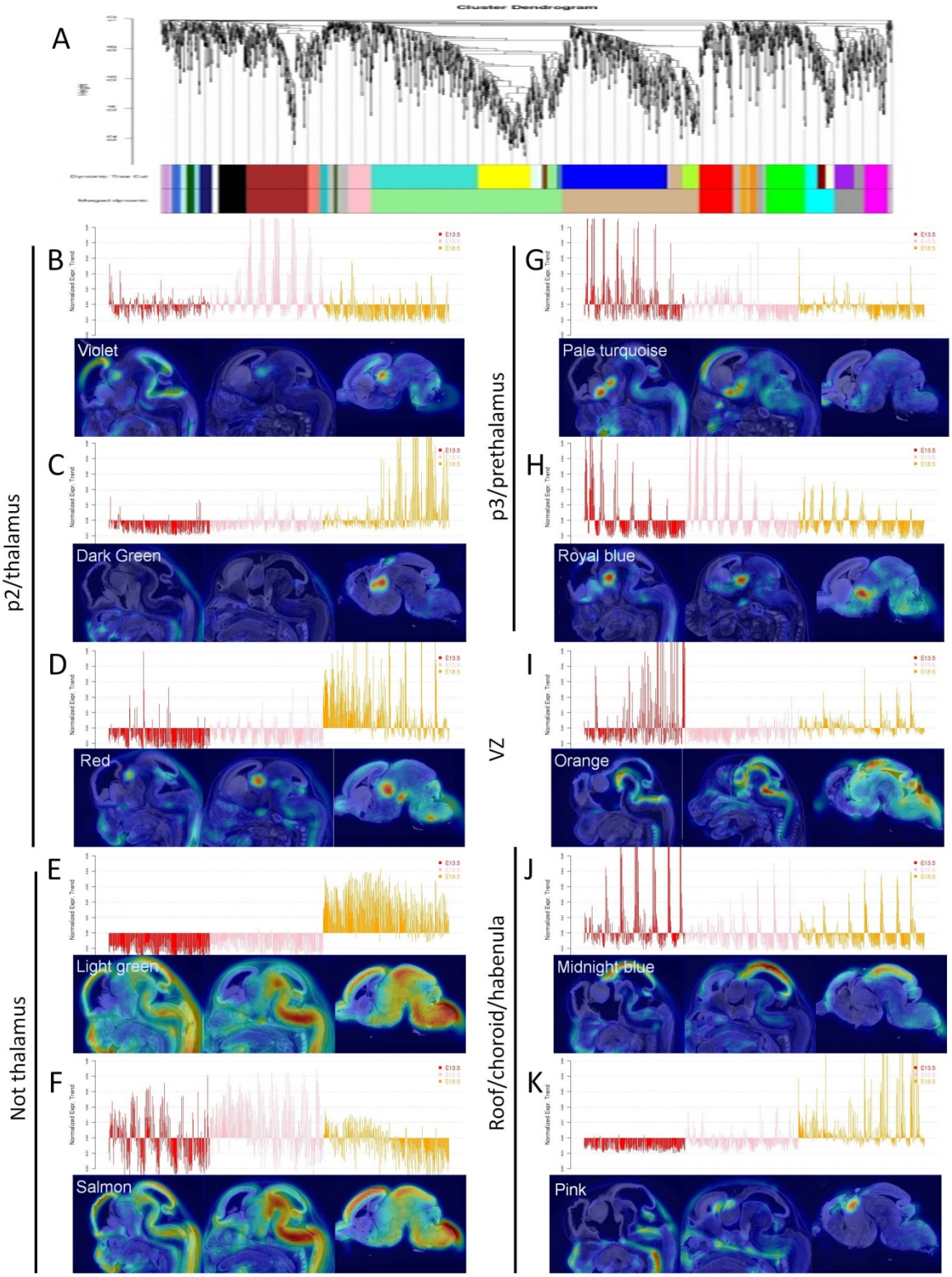
Cluster Dendrogram
A
B
Violet
C
Dark Green
D
Red
E
Light green
F
Salmon
G
Pale turquoise
H
Royal blue
I
Orange
J
Midnight blue
K
Pink
p2/thalamus
Not thalamus
p3/prethalamus
VZ
Roof/choroid/habenula

Figure S5.

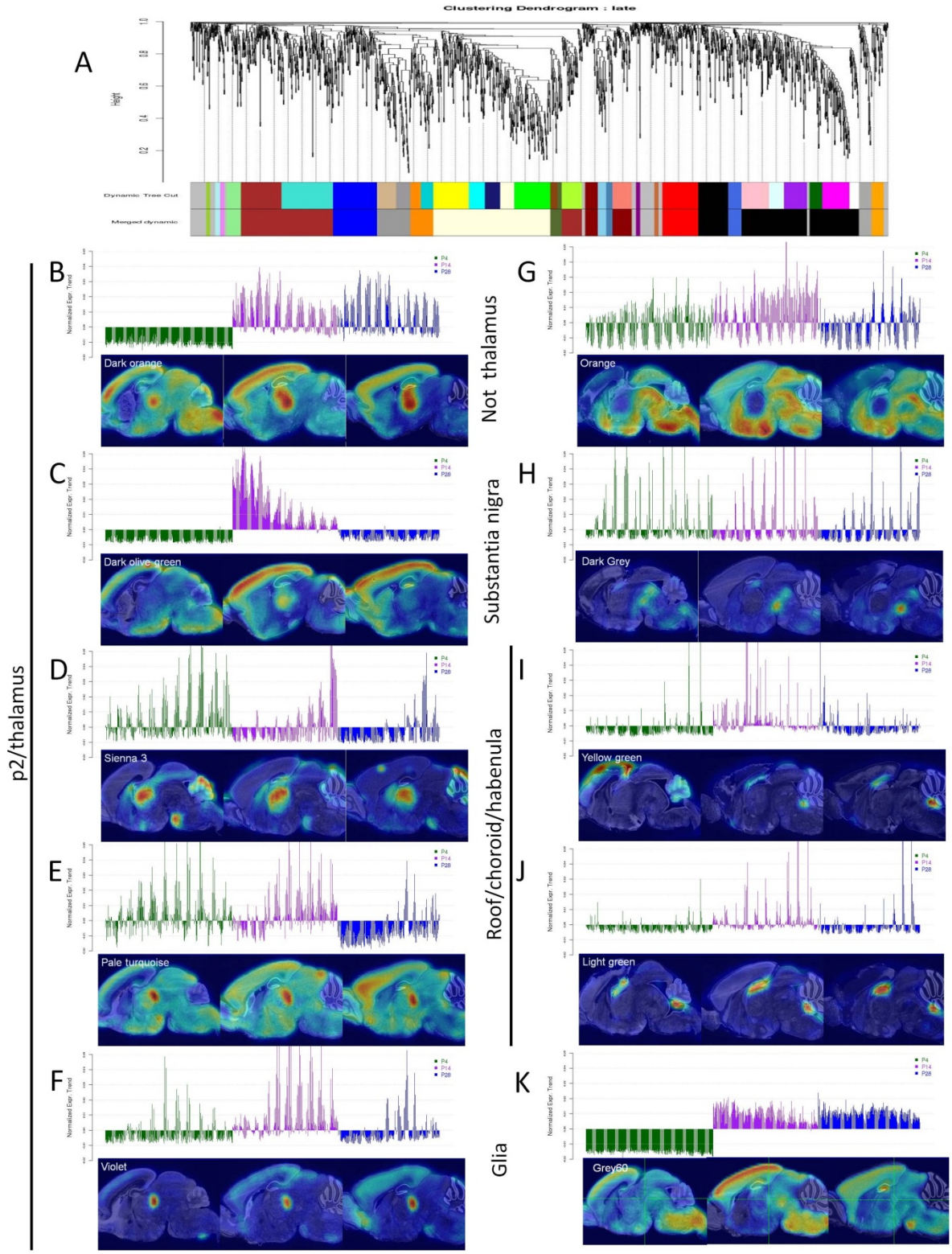
A
Clustering Dendrogram : late
Height
Dynamic Tree Cut
Merged dynamic
B
Dark orange
C
Dark olive green
D
Sienna 3
E
Pale turquoise
F
Violet
p2/thalamus
G
Orange
H
Dark Grey
Substantia nigra
Not thalamus
I
Yellow green
J
Light green
Roof/choroid/habenula
K
Grey60
Glia
P4
P14
P28

Figure S6.

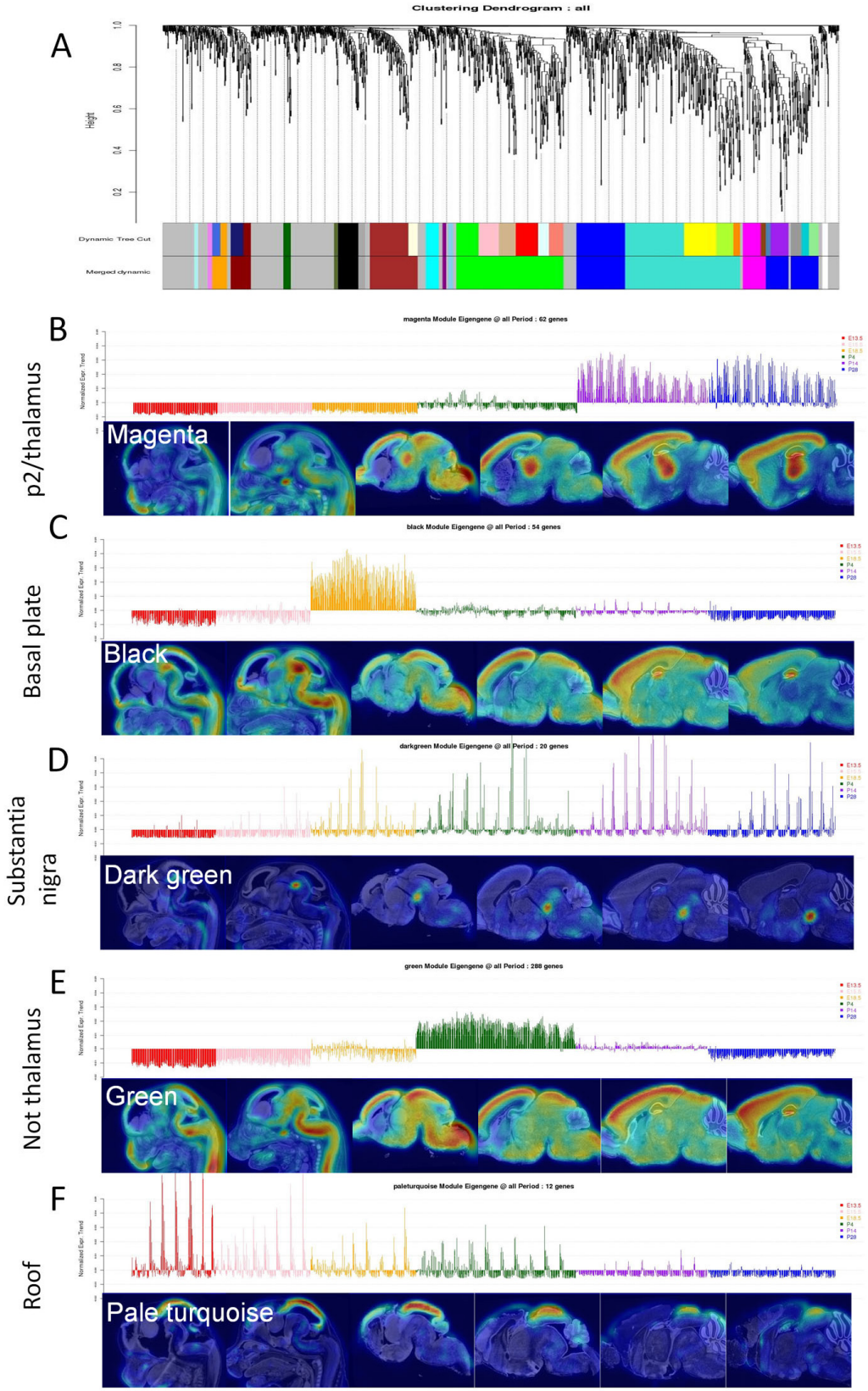
Clustering Dendrogram : all
A
Height
Dynamic Tree Cut
Merged dynamic
B
magenta Module Eigengene @ all Period : 62 genes
p2/thalamus
Magenta
C
black Module Eigengene @ all Period : 54 genes
Basal plate
Black
D
darkgreen Module Eigengene @ all Period : 20 genes
Substantia nigra
Dark green
E
green Module Eigengene @ all Period : 288 genes
Not thalamus
Green
F
paleturquoise Module Eigengene @ all Period : 12 genes
Roof
Pale turquoise

Figure S4. Spatial and Temporal Expression Profiles across E13.5, E15.5, and E18.5, related to Figure 4. (**A**) Cluster dendrogram groups genes into distinct modules using a vector of expression energy across all diencephalon voxels spanning the time window E13.5, E15.5, and E18.5 with the y-axis corresponding to co-expression distance between genes and the x-axis to genes. Two colorbars label the modules assigned by dynamic tree cutting (top) and by dynamic tree cutting followed by merging close modules (bottom), which is used in analyses. For the early period, expression levels for the diencephalon voxels at E13.5, E15.5, and E18.5 time points were concatenated as a vector for each gene. (**B-K**) Spatial realization of examples of clusters, showing the eigengene of voxels over time (top, E13.5 (red), E15.5 (pink), and E18.5 (orange)) and the plot of the average spatial expression of the cluster genes on the Nissl atlas (bottom, ordered E13.5, E15.5, and E18.5). (**B-D**) p2/thalamus modules violet, darkgreen, and red are shown. Not thalamus modules light green and salmon are shown in **E** and **F**. p3/prethalamus modules pale turquoise and royal blue are shown in **G** and **H**. A ventricular zone module (orange) is shown in **I**. Midnight blue and pink are modules with expression in the roof, choroid, and habenula (**J** and **K**).

Figure S5. Spatial and Temporal Expression Profiles across P4, P14 and P28, Related to Figure 4. (**A**) Cluster dendrogram groups genes into distinct modules using a vector of expression energy across all diencephalon voxels spanning the time window P4, P14, and P28 with the y-axis corresponding to co-expression distance between genes and the x-axis to genes. Two colorbars label the modules assigned by dynamic tree cutting (top) and by dynamic tree cutting followed by merging close modules (bottom), which is used in analyses. (**B-J**) Spatial realization of examples of clusters, showing the eigengene of voxels over time (top, P4 (green), P14 (purple), and P28 (blue)) and the plot of the average spatial expression of the cluster genes on the Nissl atlas (bottom, ordered P4, P14, and P28). (**B-F**) p2/thalamus modules dark orange, dark olive green, sienna 3, pale turquoise, and violet are shown. Not thalamus module orange is shown in **G**. Substantia nigra module dark grey is shown in (**H**). Roof, choroid plexus, and habenula modules yellow green and light green are shown in (**I**) and (**J**). A glial module grey60 is shown in (**K**).

Figure S6. Spatial and Temporal Expression Profiles across "all" timepoints, related to Figure 4. (**A**) Cluster dendrogram groups genes into distinct modules using a vector of expression energy across all diencephalon voxels spanning the time window P4, P14, and P28 with the y-axis corresponding to co-expression distance between genes and the x-axis to genes. Two colorbars label the modules assigned by dynamic tree cutting (top) and by dynamic tree cutting followed by merging close modules (bottom), which is used in analyses. (**B-F**) Spatial realization of examples of clusters, showing the eigengene of voxels over time (top, E13.5 (red), E15.5 (pink), E18.5 (yellow), P4 (green), P14 (purple), and P28 (blue)) and the plot of the average spatial expression of the cluster genes on the Nissl atlas (bottom, same order as top panel).

Figure S7.

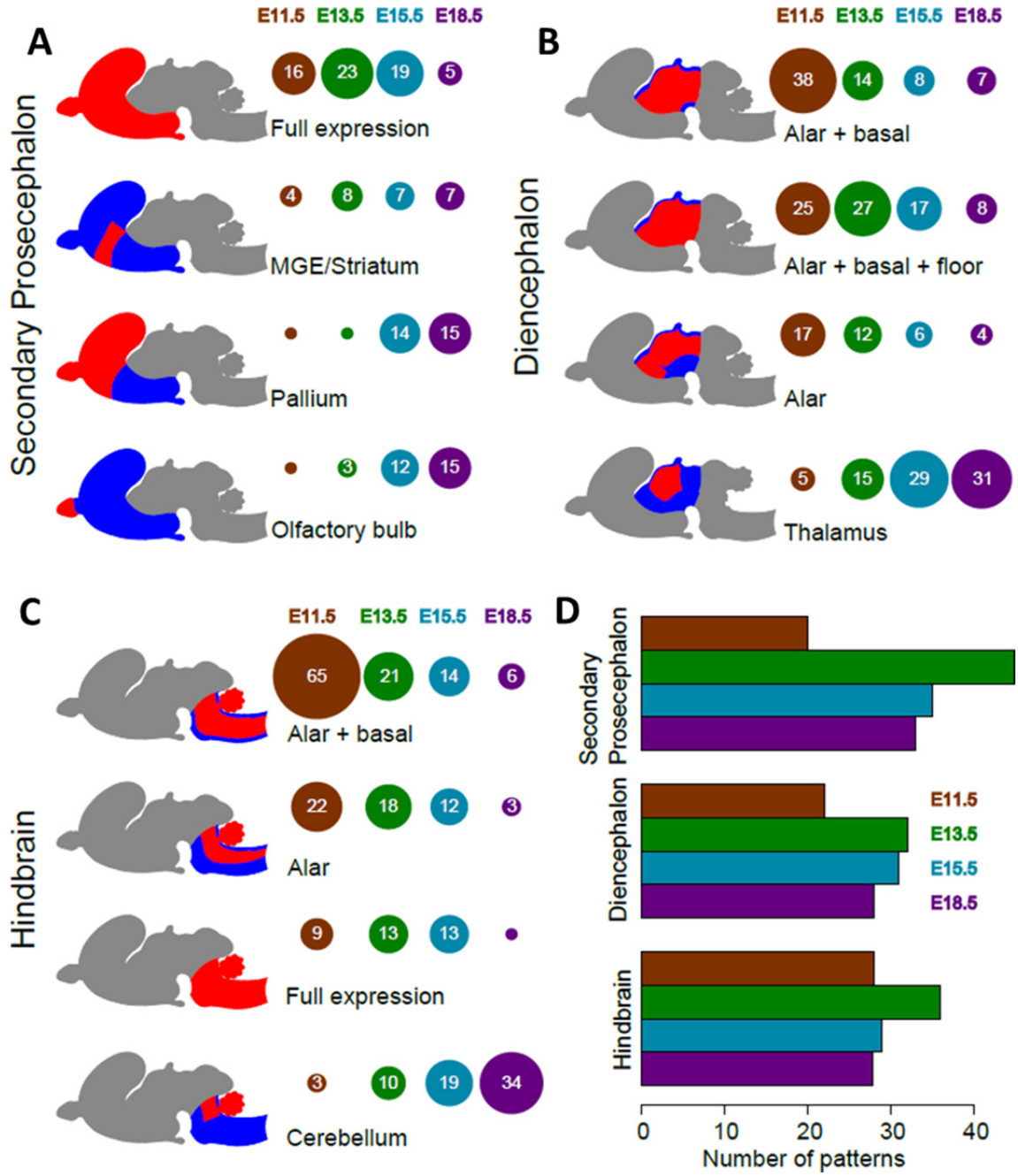

Figure S7. Expression patterns in three brain regions, related to Figure 8. (**A-C**) Most common patterns in Secondary Prosencephalon, Diencephalon, and Hindbrain using manual annotation data; number of subregions per structure was 16, 12, and 16, respectively. Brain schematics indicate expression pattern (red, expressed; blue, not expressed). Circles show number of genes per pattern. (**D**) Number of patterns observed; note that patterns that differ by expression in a single subregion are grouped together as a single pattern to allow for possible error.

A
No expression
Regional expression
Widespread expression
Forebrain
Midbrain
Hindbrain, Roof plate
Hindbrain, Alar plate
Hindbrain, Basal plate
Hindbrain, Floor plate
B
E11.5
E13.5
E15.5
E18.5

Figure S8. A minimal transcription factor code with maximal discrimination between age-specific brain regions, related to Figure 8. (**A**) Brain regions are labeled on the right and the color bar provides the atlas-region color (also demonstrated in schematic at the top). Royal blue denotes a gene is “widely expressed” and white denotes “not expressed”. Cyan indicates “locally expressed” within a structure. Only genes that are *widely expressed* or *not expressed* can be used to discriminate structures. (**B**) Brain region pairs for each age that are indistinguishable based upon transcription factor expression.

Supplemental Table Descriptions

Table S1. Gene information for Allen Developing Mouse Brain Atlas, related to Figure 2. Genes selected for *in situ* hybridization probes are listed. The table shows the gene symbol, gene name, Entrez Gene ID, gene classification, and weight gene co-expression analysis (WGCNA) module for the early (E13.5, E15.5, and E18.5), middle (E18.5 and P4), and late (P4, P14, and P28) periods as well as the correlation to the eigengene for each gene within each period. Gene class categories (column D) include axon guidance pathway, cell adhesion, ion channel, neurotransmitter pathway, Notch signaling, RTKs and RTKs ligands (receptor tyrosine kinase), transcription factor activity, bHLH TF (basic helix loop helix transcription factor), forkhead TF, homeobox TF, nuclear receptor, POU domain genes, and Wnt signaling.

Table S2. Temporal gene expression in the brain by class, related to Figure 3. Genes were evaluated for expression over development in brain: ON, on at all stages; OFF, off at all stages; OFF-ON, genes not expressed at E11.5 but eventually turn on; ON-OFF, genes that are expressed at E11.5 but eventually turn off; Complex patterns, *e.g.* OFF-ON-OFF. Gene symbol, Entrez Gene ID, expression (on or off for E11.5, E13.5, E15.5, E18.5, P4, P14, and P28), class (on, off, off-on, etc.), and classification of gene as a transcription factor, Wnt Signaling, receptor tyrosine kinase (RTK) and ligand, ion channel or GPCR are provided.

Table S3. Gene ontology results from early (E13.5, E15.5, and E18.5) WGCNA modules, related to Figure 4. Gene ontology (GO) analysis was carried out on individual modules using DAVID with GO biological process 5, GO molecular function 5, PANTHER biological process all, PANTHER molecular function all, KEGG pathway, and PANTHER pathway with the entire early gene list as the background gene list. A GO significant summary is provided containing any DAVID result with BH (Benjamini-Hochberg) p-value < 0.1. For each individual module, all DAVID results are shown separately.

Table S4. Gene ontology results from late (P4, P14, and P28) WGCNA modules, related to Figure 4. Gene ontology (GO) analysis was carried out on individual modules using DAVID with GO biological process 5, GO molecular function 5, PANTHER biological process all, PANTHER molecular function all, KEGG pathway, and PANTHER pathway with the entire late gene list as the background gene list. A GO significant summary is provided containing any DAVID result with BH (Benjamini-Hochberg) p-value < 0.1. For each individual module, all DAVID results are shown separately.

Table S5. Gene ontology results from all timepoints (E13.5, E15.5, E18.5, P4 and P14) WGCNA modules, related to Figure 4. Gene ontology (GO) analysis was carried out on individual modules using DAVID with GO biological process 5, GO molecular function 5, PANTHER biological process all, PANTHER molecular function all, KEGG pathway, and PANTHER pathway with the entire middle gene list as the background gene list. A GO significant summary is provided containing any DAVID result with BH (Benjamini-Hochberg) p-value < 0.1. For each individual module, all DAVID results are shown separately.

Supplemental Experimental Procedures

IN SITU HYBRIDIZATION PIPELINE OVERVIEW

The Allen Developing Mouse Brain Atlas (http://developingmouse.brain-map.org) study utilizes the production processes as developed for the Allen Mouse Brain Atlas (http://www.brain-map.org), a genome scale atlas of gene expression in the adult mouse brain[1], with adaptations including: 1) addition of a yellow counterstain to enhance analysis of the ISH; 2) changes in tissue embedding processes for embryonic tissue; 3) adjusted proteinase K concentrations optimized for each age; 4) adjusted Nissl protocols for some time points; and 5) utilization of dual image acquisition platforms.

Probe Design and Synthesis
Procedures described in the Allen Mouse Brain Atlas Data Production Processes[1] were used for generating probes for the Allen Developing Mouse Brain Atlas. The three sources of probe templates were cDNA clones, pooled cDNA from mouse brain, and synthetic cDNA clones.

cDNA clones. When cDNA clones were available from MGC[2] (Mammalian Gene Collection, NIH), they were used as direct templates for PCR. Clone sequences were verified by comparison to RefSeq sequences. Consensus sequences with >98% homology across 80% of the total length were used to generate probes.

cDNA templates. Pooled cDNA reactions made from mouse brain total RNA were also used as a template source. Probes were generated against sequences within 3,000 bp from the cDNA 3' end.

Synthetic cDNA clones. In rare instances splice variants made cDNA template design impossible, synthetic plasmids from GenScript (Piscataway, NJ) were ordered. Sequence information was obtained from the NCBI RefSeq database and GenScript synthesized the 400-500 bp clone insert. The resulting synthetic plasmid was then used as direct template for PCR amplification.

Mouse brain cDNA preparation. Total RNA was isolated from homogenized C57BL/6J mouse whole brain tissue using Ambion's ToTALLY RNA kit (Life Technologies, Grand Island, NY), or Trizol and Ambion's MagMaxExpress protocol. Typical yield was 120 µg total RNA per brain. Invitrogen's Superscript III RTS first-strand cDNA synthesis kit (Life Technologies) was used for cDNA reactions in a 96-well format, using 5 µg anchored olig-dT-25.

Primer design. Gene-specific forward and reverse primers were designed in the following way. BLAST[3] was used to identify regions of homology in other genes/family members, and repetitive and/or homologous sequences were masked. Primer3 software (MIT) was used for primer design with the following criteria:

1. Optimal size was 18-20 nt for clone templates, 22-24 nt for cDNA templates
2. GC content was between 42% to 62%
3. PCR product size was between 300-1200 nt (optimal > 600 nt)
4. Probe location was within the gene (No bias for clone templates, within 3,000 bp of polyA for cDNA)

The top primer pair was chosen with the lowest penalty score. A nested reverse primer was also designed for cDNA templates. The SP6 RNA polymerase binding sequence was added to the reverse/nested primer. Primers were ordered from IDT (Integrated DNA Technologies, Coralville, IA) in 96-well format at 10 µM concentration.

All gene sequences were "blasted" against available collections of transcript sequences including RefSeq, MGC, Celera, TIGR, RIKEN (FANTOM3), and UniGene. Regions of homology greater than 70% for regions over 100 bp were identified and excluded from probe design. (For a subset of genes in families with high homology these standards were relaxed to >90% homology over 120 bp). Within the remaining sequence, primers were designed using Primer3 software. A nested approach was used for generation of probes from mouse brain cDNA, using three primers: a forward, a reverse, and a nested primer. An initial

polymerase chain reaction (PCR) was performed using forward and reverse primers. The purified product was then used as a template for a second PCR using the same forward primer with the nested primer. When a cDNA clone was used as a template, a single PCR reaction was used with a single set of forward and reverse primers. All PCR products generated from cDNA templates were sequenced from both ends, using the forward primer and SP6.

In vitro Transcription (IVT). Standard IVT reactions were performed using Roche's 10X DIG RNA Labeling Mix (Roche, Indianapolis, IN). All reactions were done in 96-well format for 2 hours at 37°C, with 30 µl total volume. Purified PCR products served as the template, using SP6 RNA polymerase (NEB, Ipswich, MA). IVT reactions were purified using Millipore's Montage 96 filter plate (Millipore, Billerica, MA), and were eluted with 90 µl of THE (0.1 mM sodium citrate pH 6.4, Ambion) following 30 minute room temperature incubation. IVT reactions were quantified using RiboGreen HIGH assay (Molecular Probes, Life Technologies) and the SpectraMax-M2 plate reader (1.0 µl in 200 µl total volume) (Molecular Devices, Sunnyvale, CA). Each IVT reaction (1.0 µl) was analyzed on Agilent's Bioanalyzer 2100 (Agilent, Santa Clara, CA) for size confirmation and quantification. IVT reactions were stored at -80°C. IVT reactions were diluted to working stocks of 30 ng/µl with THE. For hybridization, probes were diluted to 300 ng/ml with hybridization buffer (Ambion) in 96-well plates and were stored at -20°C until used.

Quality Control. PCR products were evaluated for expected size and homogeneity; PCR products with multiple products were discarded. IVT products that were shorter than their predicted size were also discarded; however, it was common to see IVT products slightly larger than their predicted molecular weights, or to see multiple peaks, due to RNA secondary structure. IVT products with multiple bands were not used for ISH unless the additional bands were determined to result from secondary structure.

Specimen Preparation

The C57BL/6J mouse strain was chosen for the Allen Developing Mouse Brain Atlas because the strain is widely used in the research community, and to permit direct comparison to both the adult P56 mouse brain atlas and to the mouse spinal cord atlas. Seven time points were chosen for broad characterization of gene expression, including four embryonic time points and three postnatal time points. The embryonic time points were chosen to span a wide range of developmental events in utero, beginning with embryonic day (E) E11.5 and ending with postnatal day (P) P28.

Breeding. Specimens were derived from breeding pairs for all time points except for P28. Breeding animals were either purchased from The Jackson Laboratory West (Sacramento, CA) or were derived from purchased animals. Personnel monitored the presence of vaginal plugs at 12 hour intervals (6 am and 6 pm). In order to harvest embryonic specimens with accuracy to 0.5 days, only dams with visible plugs at 6 am were used to obtain embryonic time points. For postnatal time points, births were recorded at 12 hour intervals (6 am and 6 pm). Animal handling was reduced as much as possible for P4 and P14, and animals were maintained on a 12 hour light-dark schedule. The P28 animals were either bred in-house and weaned at day 21, or obtained from Jackson Laboratory, with shipping at P21, receipt of animals at P23, and maintained under normal housing conditions for 5 days prior to dissection.

Aged animals. For aging studies of between 18 months (18M) and 24 months of age (24M), mice were selected from the Jackson Laboratory C57BL/6J colony at 10 weeks of age. They were housed 5-10 per cage and aged at Jackson Laboratory until 17 months of age. Mice were then shipped to the Allen Institute and aged an additional month for 18M time point or 5 months for the 24M time point. These mice were housed in the same groupings as at Jackson Laboratory, but no more than 5 per cage. Environmental enrichment was provided for all animals. Early pilot work was performed using a small subset of brains obtained from 30-33 month mice from a colony at University of Washington, and early pilot work on 24M animals was performed on mice aged entirely at the Allen Institute.

Specimen criteria. In order to obtain the most homogenous set of specimens for each time point, certain criteria were established (see **Table 1**). For each embryonic time point, specific Theiler stages (TS;[4] were chosen for production, representing the most prevalent Theiler stage occurring at that age, as determined using major criteria from Karl Theiler. For E11.5 and E13.5 embryos, gender was not determined. Male specimens were used for E15.5 through adult. For P14 collection, all pups within the harvested litter

displayed eye opening. Specific specimen criteria were not established for brains obtained for E12.5, E14.5, E16.5, E17.5, P1, P2, and P7, except for selection of gender (male for E16.5, E17.5, P1, P2, and P7 and either sex for E12.5 and E14.5).

Specimen embedding. Whole embryos aged E11.5-E15.5 were dissected into a 12-well plate containing chilled 4% sucrose/PBS for 15 min. During sucrose equilibration, embryos were analyzed under a dissecting scope to determine Theiler stages.. Embryos meeting the Theiler stage criteria were then stepped through 50:50 OCT: 4% sucrose/PBS (phosphate buffered saline) for 2 minutes prior to freezing in OCT (Tissue-Tek® Optimal Cutting Temperature, Thermo Fisher Scientific, Waltham, MA) in a dry ice alcohol bath. For E16.5 and E17.5 specimens, whole heads were removed, degloved, then equilibrated and embedded. E18.5 and postnatal brains were dissected and frozen in OCT. Brains were frozen at -80ºC prior to sectioning.

Table 1. Specimen criteria

Age	TS	Major criteria	Gender (method)
E11.5	TS19	Clear anterior footplate; No sign of fingers; Pigment not uniformly distributed around eye	Undetermined
E13.5	TS21	Anterior footplate indented; Posterior footplate beginning to indent; Fingers not yet separated distally; Pinna forms crest pointing 90 degrees out of head	Undetermined
E15.5	TS24	Eyelids closing; Anterior part of the back is straight; Nail primordial on toes; Head is rounded not angular	Male (tail-tip, genotype)
E18.5	TS26	Long whiskers; Pinna fused to skin; Males are showing a lump and dark pigment between the anus and genitals	Male (visual)
P4	N/A	Body weight > 1.8 grams	Male (visual)
P14	N/A	Body weight > 6 grams; Eyes open	Male (visual)
P28	N/A	Body weight > 10 grams	Male (visual)
≥18M	N/A	Exclude for health issues during visual inspection and necropsy (skin lesions, abscesses, abnormal organs)	Male (visual)

Cryosectioning

Specimens were sectioned at 20 or 25 µm (see **Table 2**) on a Leica 3050 S cryostat (Leica Microsystems, Wetzlar, Germany) with adjacent sections placed across different slides to produce series of slides such that when a given series was used for ISH or Nissl it would result in the sampling shown in **Table 2**. Nissl series were generated for every specimen at ages P4, P14, and P28, and these associated Nissl images are available on the website.

Fixation, Acetylation, Dehydration

Fixation, acetylation, and dehydration (F/A/D) were performed as described for the Allen Mouse Brain Atlas. For all ages except E14.5 and E16.5-P7, sections were allowed to air dry on slides for 30 minutes prior to F/A/D. For E14.5, and E16.5 to P7 tissue, the tissue was first air-dried and then baked overnight in a 37°C oven prior to F/A/D, in order to improve tissue adhesion to the slide. The tissue was then fixed in 4% neutral buffered paraformaldehyde (PFA) for 20 minutes and rinsed for 3 minutes in 1x PBS (phosphate buffered saline). Acetylation was performed to reduce non-specific probe binding to tissue sections. Tissue was equilibrated briefly in 0.1 M triethanolamine and acetylated for 10 minutes in 0.1 M triethanolamine with 0.25% acetic anhydride. Immediately following acetylation, tissue was dehydrated through a graded series containing 50%, 70%, 95%, and 100% ethanol. Finally, each slide was analyzed microscopically to ensure section quality. Slides that pass quality control were stored at room temperature in Parafilm-sealed slide boxes for up to one month.

***In Situ* Hybridization**

In situ hybridization (ISH) processes were performed as described in the Allen Mouse Brain Atlas, using a non-radioactive, digoxigenin (DIG) based technique to label cells expressing a particular mRNA sequence. The ISH protocol was performed using a Tecan robot (Tecan Group, Mannedorf, Switzerland)

with GenePaint technology developed by Dr. Gregor Eichele's Laboratory at the Max Planck Institute and Baylor College of Medicine[5]. In situ hybridization (ISH) processes were performed as described in the Allen Mouse Brain Atlas, using a non-radioactive, digoxigenin (DIG) based technique to label cells expressing a particular mRNA sequence. The ISH protocol was performed using a Tecan robot (Tecan Group, Mannedorf, Switzerland) with GenePaint technology developed by Dr. Gregor Eichele's Laboratory at the Max Planck Institute and Baylor College of Medicine[5]. The ISH protocol executed on the Tecan platform is listed at the end of this document. Slides were integrated into flow-through chambers on a temperature-controlled rack, and placed on a Tecan Genesis liquid handling platform. Reagents were applied using a liquid handling system. Prior to hybridization, the fixed, acetylated, and dehydrated tissue has undergone steps designed to block endogenous peroxidase activity and to increase permeability of the tissue, allowing penetration and hybridization of the labeled probe to its complementary target mRNA. Tissue was incubated with digoxigenin-labeled riboprobe for 5.5 hours at 63.5°C. Once hybridization was complete, tissue was treated with a sequence of increasingly stringent washes containing decreasing salt concentrations.

Table 2a. Sectioning schemes for major time points

Ages	Specimen type	Section width	Genes per brain	Sampling per gene	Nissl sampling
E11.5	embryo	20 µm	4	80 µm*	N/A
			* sagittal only; coronal is 120 µm		
E13.5	embryo	20 µm	5	100 µm	N/A
E15.5	embryo	20 µm	6	120 µm	N/A
E18.5	brain	20 µm	7	140 µm	N/A
P4	brain	20 µm	7	160 µm	160 µm
P14	brain	25 µm	7	200 µm	200 µm
P28	brain	25 µm	7	200 µm	200 µm
P56	brain	25 µm	6	200 µm	100 µm
18M	brain	25 µm	6	200 µm	100 µm
24M	brain	25 µm	6	200 µm	100 µm

Table 2b. Sectioning schemes for additional time points

Ages		Section width	Genes per brain	Sampling per gene	Nissl sampling
E12.5	embryo	20 µm	4	80 µm	N/A
E14.5	embryo	20 µm	5	100 µm	N/A
E16.5	head	20 µm	6	120 µm	N/A
E17.5	head	20 µm	7	140 µm	N/A
P1	brain	20 µm	6	140 µm	140 µm
P2	brain	20 µm	7	160 µm	160 µm
P7	brain	20 µm	8	180 µm	180 µm

Detection of the bound probe was a multi-step procedure. First, a succession of blocking steps inhibited endogenous protein activity from interfering with the colorimetric enzymatic reactions. The colorimetric reaction itself was a four-part process, starting with addition of a horseradish peroxidase (HRP)-conjugated anti-digoxigenin antibody. A tyramide signal amplification (TSA) step was utilized to maximize sensitivity. The tissue was incubated with a biotin-coupled tyramide. Tyramide was converted to HRP into a highly reactive oxidized intermediate which binds rapidly and covalently to cell-associated proteins by or near the HRP-linked probe, resulting in amplification of bound biotin molecules available for detection up to a hundred fold. These biotin molecules were then bound to neutravidin-AP. A colorimetric reaction occurred when the alkaline phosphatase (AP) conjugated to neutravidin enzymatically cleaved the phosphate from 5-bromo-4-chloro-3-indolyl phosphate (BCIP), and two of the resulting indoles have undergone a redox reaction with nitroblue tetrazolium (NBT) to produce a blue-purple precipitate at the site of probe binding. Once this process was completed, the tissue was treated with a wash buffer containing ethylenediaminetetraacetic acid (EDTA) followed by fixation with 4% PFA to halt the colorimetric reaction.

Table 3. Proteinase K treatment

Ages	Proteinase K (U/mL)	Incubation Time
E11.5	0.0014 U/mL	1 x 10 min
E12.5-P7	0.0025 U/mL	1 x 10 min
P14-adult	0.0175 U/mL	2 x 10 min

The major modification to the original ISH protocol was the optimization of proteinase K concentrations to obtain the highest *in situ* hybridization signal while retaining tissue integrity. The proteinase K

concentrations are listed below (**Table 3**); postnatal time points P14 and P28 were processed under the same conditions as the P56 tissue in the Allen Mouse Brain Atlas.

Every ISH run had three control slides on age-matched tissue: two positive controls and one negative control. Positive control genes were selected by the following criteria: 1) expression at all time points examined; 2) expression across many brain regions, such that most brain sections will show evidence of expression; and 3) expression of varying intensity at each time point including areas of no expression, low/medium expression, and high expression. Positive controls were Cannabinoid receptor 1 (*Cnr1*) and Calbindin 2 (*Calb2*).

Yellow Counterstain
The Feulgen-HP yellow DNA counterstain is a nuclear stain that adds definition to the tissue for the purpose of analyzing and understanding the gene expression data. This counterstain was used in conjunction with ISH for all data produced for the Allen Developing Mouse Brain Atlas, except for P56, in order to provide tissue context to the ISH signal which is otherwise difficult to discern due to the very light tissue background for embryonic ISH. The counterstain also enabled better tissue detection and focus during automated image acquisition.

After colorimetric ISH was completed on the Tecan robots, slides were removed and were subjected to an acid alcohol wash (70% ethanol adjusted to pH 2.1) to reduce background, 5N hydrochloric acid washes to prepare the tissue for HP yellow counterstain, followed by HP yellow counterstain (Catalog #869, Anatech, Battle Creek, MI), and two final acid alcohol washes to remove non-covalently bound HP yellow. Slides were coverslipped using Hydro-Matrix® Mounting Medium (Micro Tech Lab, Graz, Austria). The yellow counterstain and acid alcohol washes were carried out using a Leica CV5030 coverslipper. Coverslipped slides were incubated overnight at 37°C to solidify the mounting media. Prior to scanning, slides were cleaned to remove excess mounting media and other debris.

Nissl Staining
In adult and postnatal brains, Nissl staining served as a cytoarchitectural reference to help identify specific cell populations; however, at earlier times in brain development, this stain gives no more information than a nuclear stain, such as the Feulgen-HP yellow counterstain present on all ISH datasets. Nissl sets were generated for all postnatal time points at sampling densities indicated in **Table 2**.

The Nissl protocol using 0.25% thionin stain for the Allen Mouse Brain Atlas was used for P14 and P28 tissue. For P1-P7 tissue, the only modification that was made to the protocol was the substitution of 0.72% cresyl violet/60 mM sodium acetate, pH 3.4 for the thionin stain. Briefly, after sectioning, a set of slides from each P4, P14, P28, P56, 18M, or 24M brains was baked at 37°C for 1-5 days. Sections are defatted with xylene substitute Formula 83 (CBG Technologies, Columbus, OH) and hydrated through a graded ethanol series (100%, 95%, 70%, and 50% ethanol). After incubation in water, slides were stained in either thionin or cresyl violet, differentiated and dehydrated in water and a graded ethanol series (50%, 70%, 95%, and 100% ethanol). Finally, slides were incubated in Formula 83 and coverslipped in DPX mounting medium. Slides were air-dried in a fume hood at room temperature.

Image Acquisition
Slides for P14, P28, P56, 18M, and 24M tissue were initially scanned on the same Image Capture System (ICS) platform developed for use for the Allen Mouse Brain Atlas. All other time points (E11.5 to P7) were scanned using the ScanScope® automated slide scanner (Aperio Technologies, Vista, CA) equipped with a 20x objective and Spectrum software, and whole slide images were downsampled to a resolution of 1.0 µm/pixel. Later in the project, all slides from all time points were scanned with the ScanScope® system.

Data Processing
The automated image processing workflow leverages the informatics data processing pipeline generated for the Allen Mouse Brain Atlas with specific informatics modules created for this project. Following image acquisition on the ICS platform, individual section images were “stacked” or combined into a single slide image, which then enters the same pipeline as the Aperio images. All slide images were white-balanced,

and a tissue detection algorithm assigned bounding boxes to individual tissue sections, which were manually assessed and adjusted when necessary. A segmentation algorithm created the expression mask, which is provided as a colorized view of expression levels across the tissue. The position of each section in the brain or specimen was calculated to a master section index which provides a framework for section position across all time points. For P4, P14, P28, 18M, and 24M, the closest Nissl section was calculated for each ISH section.

Quality Control

Quality control measures were implemented throughout the process. There was a quality control step for section quality post-fixation just prior to ISH. Post-ISH quality control consisted of examination of both positive and negative control slides as well as a random sampling of experimental slides. After image acquisition, image quality was assessed for focus and bounding box. Finally, the data analysis team ensured that all passed images presented to the public met an acceptable standard for consistency and were of sufficient overall quality for public release. First, specimens were inspected for anatomic anomalies, dissection damage, Theiler stage, midline, and orientation. If a specimen did not pass this initial evaluation then the entire specimen and all related image series were failed and rerun. Second, each image series was reviewed for sectioning quality, ISH quality, and scanning artifacts. Slides that contain sections with artifacts that can be mitigated were sent back to an earlier stage to be remedied, after which the image series was reassessed by the data analysis team. Third, any individual sections that were unanalyzable were failed and not released to the public. When greater than 20 percent of the sections contained artifacts the image series was not released and was regenerated.

INFORMATICS DATA PROCESSING

The informatics data processing pipeline developed by the Allen Institute enables the navigation and analysis of the Allen Developing Mouse Brain Atlas large and complex dataset to identify gene expression with precise spatial and temporal regulation.

In particular, informatics data processing supports the following features in the Web application:

1. The **"Expression Summary"** is a heatmap representation of gene expression for a given gene by age and by atlas structure.

2. A cross-plane and cross-time, point-based "**Synchronize**" feature in the Zoom and Pan (Zap) Image Viewer allows multiple image series to be synchronized to the same approximate position in the brain based on a linear alignment of the images to a set of 3D reference models. An image series is an indexed set of images spanning a single specimen where sections are treated with the same stain, such as an ISH for a particular gene or a Nissl stain.

3. Visualization of gene expression in a 3D format uses "**Brain Explorer 2**".

4. The "**Anatomic Search**" feature enables users to discover genes that are predominantly enriched within a brain structure at a specific age.

5. The "**Temporal Search"** feature allows users to search for genes that exhibit higher expression at a particular age for a specific brain region.

6. "**Developmental AGEA**", or "Developmental Anatomic Gene Expression Atlas" provides users the ability to explore the spatial and temporal relationships in the developing brain based on gene expression and search for genes expressed at a given voxel in the brain.

7. The "**Neuroblast**" feature allows the user to search for genes whose expression patterns are highly correlated to the seed gene.

The informatics data processing pipeline consists of the following components: a set of 3D reference models, an Alignment module, an Expression Gridding module, a Structure Unionizer module, an Anatomic Search strategy, a Temporal Search strategy, a Gene-to-Gene Correlation module (to support NeuroBlast similarity search) and a Voxel-to-Voxel Correlation module (to support Developmental AGEA). These are described in more detail below.

3D Reference Models
The cornerstone of the automated pipeline is a set of 3D reference models. For each time point, a specimen was sectioned to span a nearly complete specimen and slides were either Nissl or Feulgen-HP yellow stained to form one high density image series. Images were reassembled to form a consistent 3D volume. Structural delineation from the 2D reference atlas images were inserted into the 3D model and interpolated to create 3D structural delineations. The 3D reference spaces were then co-registered and scaled into a common space such that brains of different ages can be roughly compared for the purpose of the "Synchronize" feature.

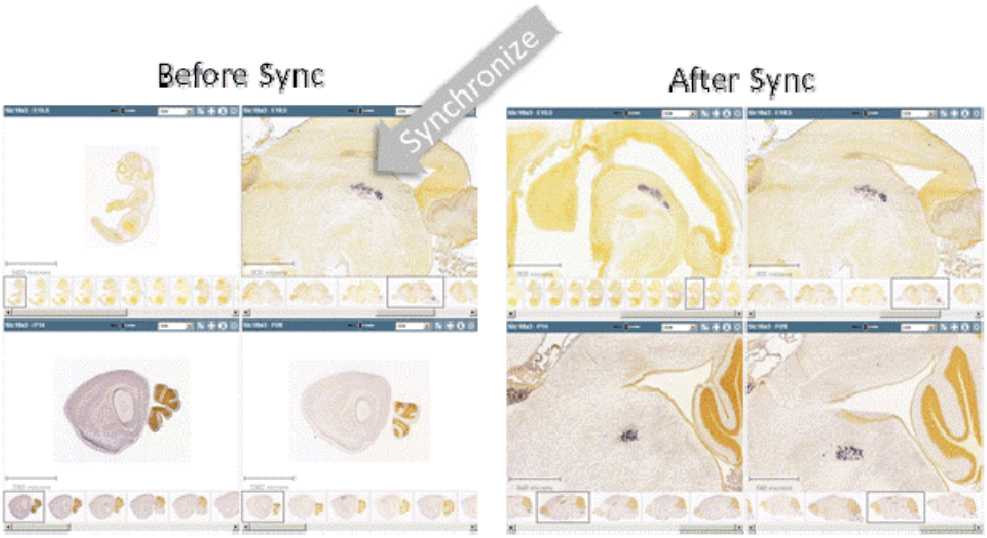

Figure 1. Point-based image synchronization. Multiple image-series in the Zoom and Pan (Zap) Image Viewer can be synchronized to the same approximate location within and across time points. Before and after synchronization screenshots show gene *Slc18a3* at ages E15.5, E18.5, P14 and P28 (ordered top left to lower right for each panel).

Alignment Module
The Alignment module operates on a per-specimen basis where all image series from a specimen are combined as one super series. Based on maximization of image correlation, the module interleaves the sections from different gene image series, reconstructing the specimen as a consistent 3D volume with co-registration to the 3D reference model. Once registration is achieved, information from the 3D reference model can be transferred to the reconstructed specimen and vice versa. The resulting transform information is saved in the database to support the image synchronization feature in the Zap viewer. Because reference models for each time point are also co-registered, synchronization is possible between specimens of different ages (**Figure 1**).

Expression Gridding Module
A detection algorithm was applied to each ISH image to create a mask identifying pixels in the high resolution image which corresponds with gene expression. The aim of the Gridding module was to create a low resolution 3D summary of the gene expression and project the data to a common coordinate space of the 3D reference model to enable spatial comparison between data from different specimens. The expression data grids were used for downstream search and analysis, and they can also be viewed directly as 3D volumes in Brain Explorer 2 (similar to Brain Explorer[6]), alongside the 3D version of the Allen Developing Mouse Brain Reference Atlas. Resolution of the data grids varies with age and corresponds with the sampling density for that time-point: ranging from 80 µm for E11.5 to 200 µm for P28 (**Figure 2**). For the purpose of search and analysis pixel-based statistics of sum and average number of expressing pixels and sum and average expression intensity per grid voxel were collected.

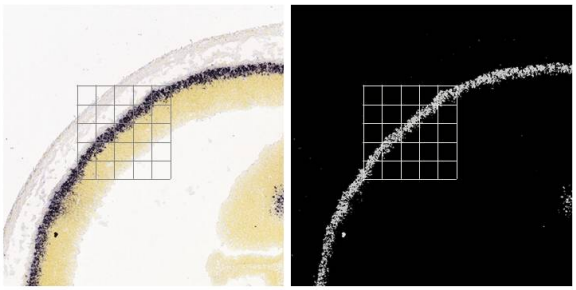

Figure 2. Expression grid sizes per age. An example of a 100 µm grid on an ISH section from an E13.5 embryo and the resulting mask section. Grid sizes are determined by the interval between sections for ISH image series. The resolution of the data grids varies with age and corresponds with the sampling density for that time-point ranging from 80 µm/side for E11.5 to 200 µm/side for P28.

Age	E11.5	E13.5	E15.5	E18.5	P4	P14	P28
3D Grid Size (µm/side)	80	100	120	140	160	200	200

Structure Unionizer

Expression statistics were computed for each structure in the reference atlas by combining or "unionizing" grid voxels with the same 3D structural label. *Expression energy for brain region R* is defined as the sum of expressing pixel intensities in R divided by the total number of pixels that intersect R. While the reference atlas is annotated at ontological Level 08, statistics at lower levels (Levels 00 to 05) were obtained by further combining measurements of the hierarchical children to obtain statistics for "parent" structures. The computed structure-based expression statistics were displayed as an expression summary on the image series or gene page, and used to enable Anatomic and Temporal Search.

Anatomic Search

The goal of Anatomic Search is to enable users to discover genes that are predominantly enriched within a particular brain region, with results provided for a specific developmental age. Our approach was to define an enrichment measure that will permit the ranking of different genes for their specificity in the brain structure of interest as compared to a "contrast" brain region (**Figure 3**).

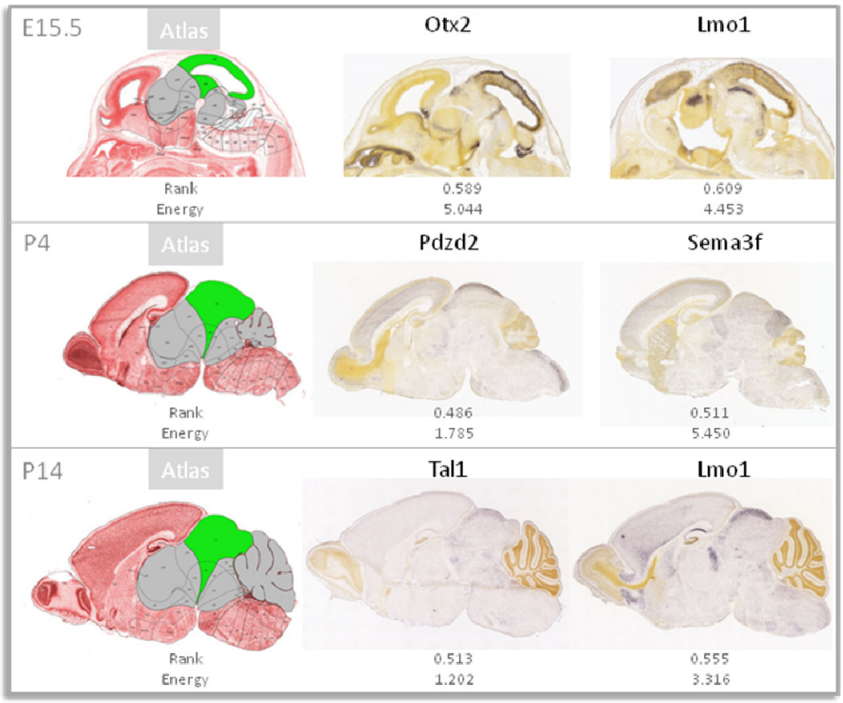

Figure 3. Anatomic Search examples. Genes are ranked by specificity to midbrain by ratio of expression in midbrain (green area) over expression in a denominator set of diencephalon, midbrain, and prepontine hindbrain (green plus gray). Atlas schematics (left column) shown are false-colored images. Gene names are shown at the top of each ISH image (two right columns). For each gene, the specificity *rank*, defined as the ratio of sum of expressing pixel intensities in the numerator set over the sum of expressing pixel intensities in the denominator set, is computed. The rank and energy for each gene for this query is provided below the ISH image.

Specifically, we defined a set of non-overlapping brain structures as the "numerator" set. Typically, the "numerator" set will simply be the brain structure of interest. The flexibility to incorporate other areas to the numerator was useful for example for the E11.5 embryos where in many areas the brain is just a thin

wall surrounding large ventricles. Slight misalignment may cause the expression to be excluded from a structure; the inclusion of adjacent ventricle areas (indicated here with a "v_" prefix) in the numerator may help to mitigate alignment errors.

For each search, we also defined a set of non-overlapping brain structures as the "denominator" set. Many genes exhibit region-specific enrichment, albeit in multiple areas; thus using the whole brain (or neural plate) as a "contrast" brain region does not necessarily provide a full list of genes, due to local specificity of the gene in multiple regions of the brain. In order to identify genes with local specificity to an anatomic region, a "contrast" region was used as the denominator to determine local specificity. The spatial span of the "numerator" set must be within the spatial span of the "denominator" set, or more simply, the denominator region is inclusive of the numerator. For each gene, the specificity *rank* defined as the ratio of sum of expressing pixel intensities in the numerator set over the sum of expressing pixel intensities in the denominator set was computed. Theoretically, rank can range from 0 (no expression in the numerator) to 1 (ideal specificity to the numerator). Genes were sorted in descending rank order to generate the Anatomic Search return lists. The maximum observed rank varies per structure and age.

Table 4. Calculations used to identify genes expressed in 12 brain regions

Age	Structure	Numerator	Denominator	Energy Threshold
E13.5, E15.5, E18.5, P4, P14, P28	Tel	Tel	F	1
E13.5, E15.5, E18.5, P4, P14, P28	PedHy	PedHy	p3, PedHy, RSP	1
E13.5, E15.5, E18.5, P4, P14, P28	CSPall	CSPall	Tel	1
E13.5, E15.5, E18.5, P4, P14, P28	DPall	DPall	Tel	1
E13.5, E15.5, E18.5, P4, P14, P28	MPall	MPall	Tel	1
E13.5, E15.5, E18.5, P4, P14, P28	p3	p3	p3, p2	1
E13.5, E15.5, E18.5, P4, P14, P28	p2	p2	F	1
E13.5, E15.5, E18.5, P4, P14, P28	p1	p1	p2, p1, M	1
E13.5, E15.5, E18.5, P4, P14, P28	M	M	D, M, PPH	1
E13.5, E15.5, E18.5, P4, P14, P28	PPH	PPH	M, PPH, PH	1
E13.5, E15.5, E18.5, P4, P14, P28	PH	PH	PPH, PH, PMH	1
E13.5, E15.5, E18.5, P4, P14, P28	PMH	PMH	PH, PMH, MH	1
E13.5, E15.5, E18.5, P4, P14, P28	MH	MH	PMH, MH, SpC	1
E11.5	Tel	Tel, v_Tel	RSP, Tel, p3, v_RSP, v_Tel, v_p3	1
E11.5	PedHy	PedHy, v_PedHy	p3, PedHy, RSP, v_p3, v_PedHy, v_RSP	1
E11.5	D	D, v_D	SP, v_SP, D, v_D, M, v_M	1
E11.5	M	M, v_M	p1, M, PPH, v_p1, v_M, v_PPH	1
E11.5	H	H, v_H	NP, ventricles	1

Abbreviations: CSPall, central subpallium (striatum/pallidum); D, diencephalon; DPall, dorsal pallium (isocortex and entorhinal cortex); F, forebrain; H, hindbrain; M, midbrain; MH, medullary hindbrain (medulla); MPall, medial pallium (hippocampus, taenia tecta, subiculum); NP, neural plate; p1, prosomere 1 (pretectum); p2, prosomere 2 (thalamus); p3, prosomere 3 (prethalamus); PedHy, peduncular hypothalamus; PH, pontine hindbrain (pons proper); PMH, pontomedullary hindbrain; PPH, prepontine hindbrain; RSP, rostral secondary prosencephalon; SP, secondary prosencephalon; SpC, spinal cord; Tel, telencephalic vesicle.

To minimize false positives due to artifacts, genes with expression in the numerator below a specified expression energy threshold are excluded from the return list. Each of the search returns was then verified in a quality control step, and search returns resulting from artifacts were removed from the search. **Table 4** lists the numerator, denominator and expression energy threshold for the anatomic searches available via the Web application.

Temporal Search
The goal of Temporal Search is to allow users to search for genes that exhibit higher expression at a particular age, with results returned for a specific brain region. Note that while the temporal search provides results for a particular anatomic region, the results are provided regardless of the anatomic specificity of the gene expression. For each gene and brain region of interest R, a simple ranking metric was computed at age A where rank is defined as the ratio of the expression energy of R at age A over the sum of expression energy of R over all ages (the seven standard time points). Theoretically, rank can range from 0 (no expression at age A) to 1 (ideal specificity to the age A). For each age, genes were sorted in descending rank order to generate the Temporal Search return lists. Temporal Search lists are provided for Tel, D, M and H (telencephalic vesicle, diencephalon, midbrain, and hindbrain) for all seven ages.

To minimize false positives due to artifacts, genes with expression energy below the specified threshold or genes with "widespread" expression were excluded from the return list. In order to identify and remove "widespread" genes, a metric based on the coefficient of variation (standard deviation/mean) of the expression energy of voxels spanning the whole brain was used. The threshold for removing "widespread" genes was determined for each individual time point by manual assessment of search returns. The Temporal Search results provided on the Web were manually verified in a quality control process by a team of data analysts.

NeuroBlast
NeuroBlast is a search tool to help identify genes with similar 3D spatial gene expression profiles. While searching for genes using the conventional "Anatomic Search" strategy is a natural approach to identify genes of interest expressed in a particular region, greater search power may sometimes be obtained by starting with a particular expression pattern and inquiring whether there exist other genes with a similar pattern of expression. For example, in order to identify genes expressed in a particular cell type which is distributed throughout the brain (*e.g.,* astrocytes, oligodendrocytes), a region-based approach may not be useful. Instead, one might use a gene which is a canonical cell type marker to initiate a correlational search to identify genes with a similar expression pattern; the results may be enriched in genes also expressed in the desired cell type.

To support NeuroBlast, Pearson's correlation coefficient was computed for each pair of image series (at the same age) using the 3D expression data grids. In particular, correlation was computed using expression energy of each voxel and over five regions of interest: NP, Tel, D, M, and H (neural plate, telencephalic vesicle, diencephalon, midbrain, and hindbrain). For database efficiency, the top 250 most similar image series were archived and presented on the Web application. NeuroBlast is accessible on the gene search results page to the right of each image series.

Developmental AGEA
The Developmental Anatomic Gene Expression Atlas (Developmental AGEA) is a new relational atlas that allows users to explore the spatiotemporal relationships in the developing mouse brain based on the expression patterns of ~2,000 genes. Similar to the AGEA for the adult mouse brain[7], Developmental AGEA is based on interactive visualization of 3D correlation maps rendered as false color images. The value at a spatial location (voxel) of a map represents the Pearson's correlation coefficient (cc) of the voxel with respect to a "seed" voxel. Correlation is computed over a "gene vector" whose elements represent the expression energy for a gene at the voxel of interest. 3D correlation maps are generated for each possible seed voxel (265,621 in total over 7 ages).

Figure 4 illustrates the construction of an intra-age (within the same time point) and an inter-age (across two time points) correlation map. Correlation values in intra-age maps are typically of higher value than those in inter-age maps. In an intra-age correlation computation, corresponding elements in the gene vector are derived from the same ISH experiment (image-series) while in an inter-age computation, corresponding elements are derived from different experiments from specimens of different ages. The semi-quantitative nature of ISH, inter-experiment variability, and natural developmental differences results in lower correlation values in inter-age maps. Typically, inter-age maps should be interpreted with respect to relative correlation within the brain of the "map" age and not absolute comparisons between ages.

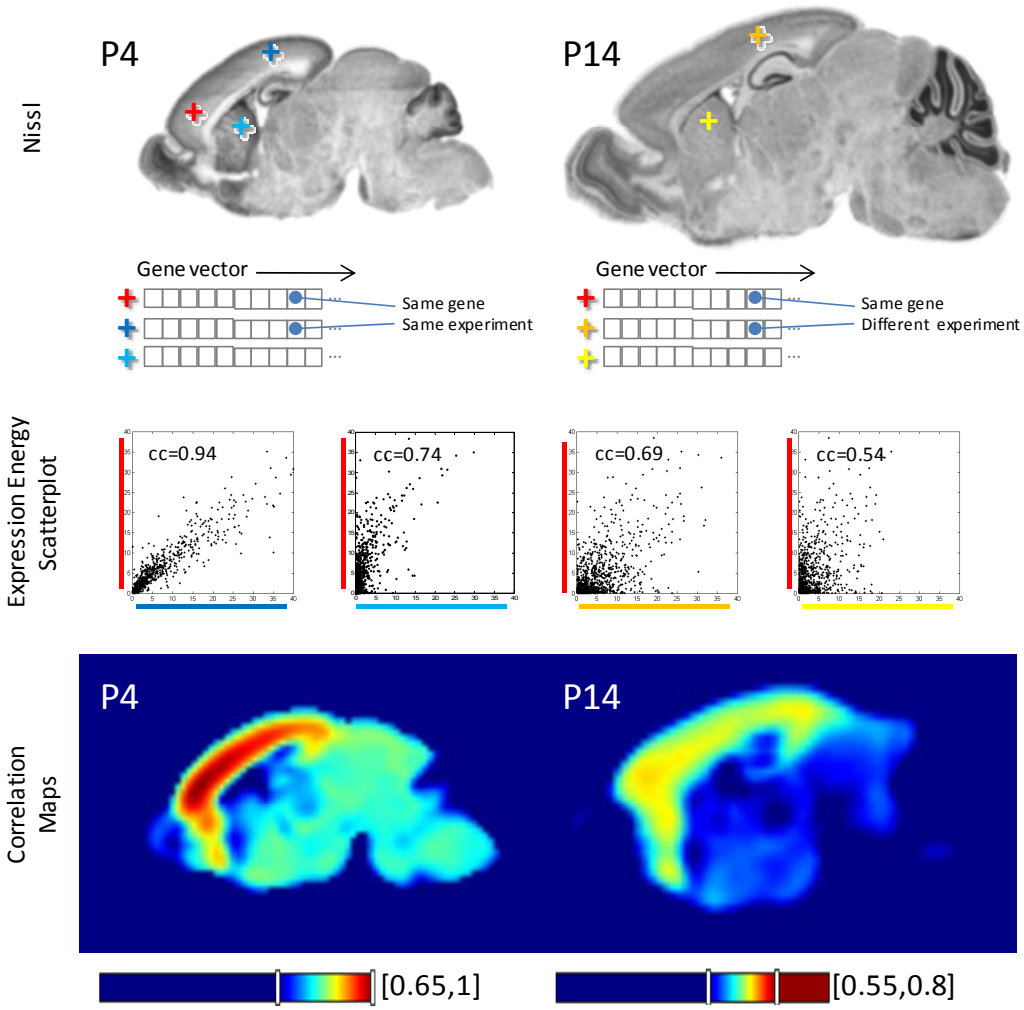

Figure 4. Construction of an intra-age (P4) and inter-age (P4-P14) AGEA correlation map for a seed voxel in the isocortex. The red cross within the isocortex in the P4 Nissl image represents the "seed" voxel. The "gene vector" at this location is correlated with the corresponding values at two other intra-age locations (light blue and dark blue crosses). Scatter plots of the expression energy over ~2,000 genes between the "seed" and "target" voxels shows that the cortical target (dark blue) is more correlated (cc=0.94) to the seed location than the striatal target (light blue) (cc= 0.74). For the chosen seed, correlation is computed at every other voxel in the P4 brain and visualized simultaneously as a false color map that can be thresholded for significance by the user. Cool colors represent lower correlation values while warmer colors represent higher correlation values. The P4 correlation map show strong correlation between the seed and cortical areas and lower correlation with subcortical areas. A similar correlation map can be generated between the "seed" voxel (at P4) and the voxels in the P14 mouse brain. Scatter plots of the expression energy between the "seed" and the inter-age targets show that the cortical target (orange cross) is more correlated (cc=0.69) to the seed location than the striatal target (yellow cross) (cc=0.54). A complete 3D map can be generated by computing the correlation at every P14 voxel with respect to the "seed" voxel. The P14 correlation map shows a similar stronger correlation between the seed and cortical areas and lower correlation with subcortical areas.

Gene Finder

The Gene Finder function of Developmental AGEA uses spatial correlation maps to generate a search "space" to find genes enriched in the correlation region surrounding a seed voxel. For each correlation

map, voxels were assigned to numerator and denominator spaces, similar to Anatomic Search. For each correlation map, let t_m be the maximum correlation value. A "denominator" threshold (t_d) was determined such that the number of voxels greater than t_d spans approximately 1/3 of the brain. A "numerator" threshold (t_n) is defined as ($0.6 * t_m + 0.4 * t_d$). All voxels greater than t_d form the "denominator" space and all voxels greater than t_n form the "numerator" space. Note that by definition, the "denominator" space is inclusive of the "numerator" space. For each gene, a specificity rank was defined as the sum of expression energy in the numerator space over the sum of expression energy in the denominator space. Genes were sorted in descending rank order and the top 100 for each map was archived and presented on the Web application. For an inter-age search, the rank between the "seed" age map and "map" age map was averaged to produce a combined rank (see **Figure 5**).

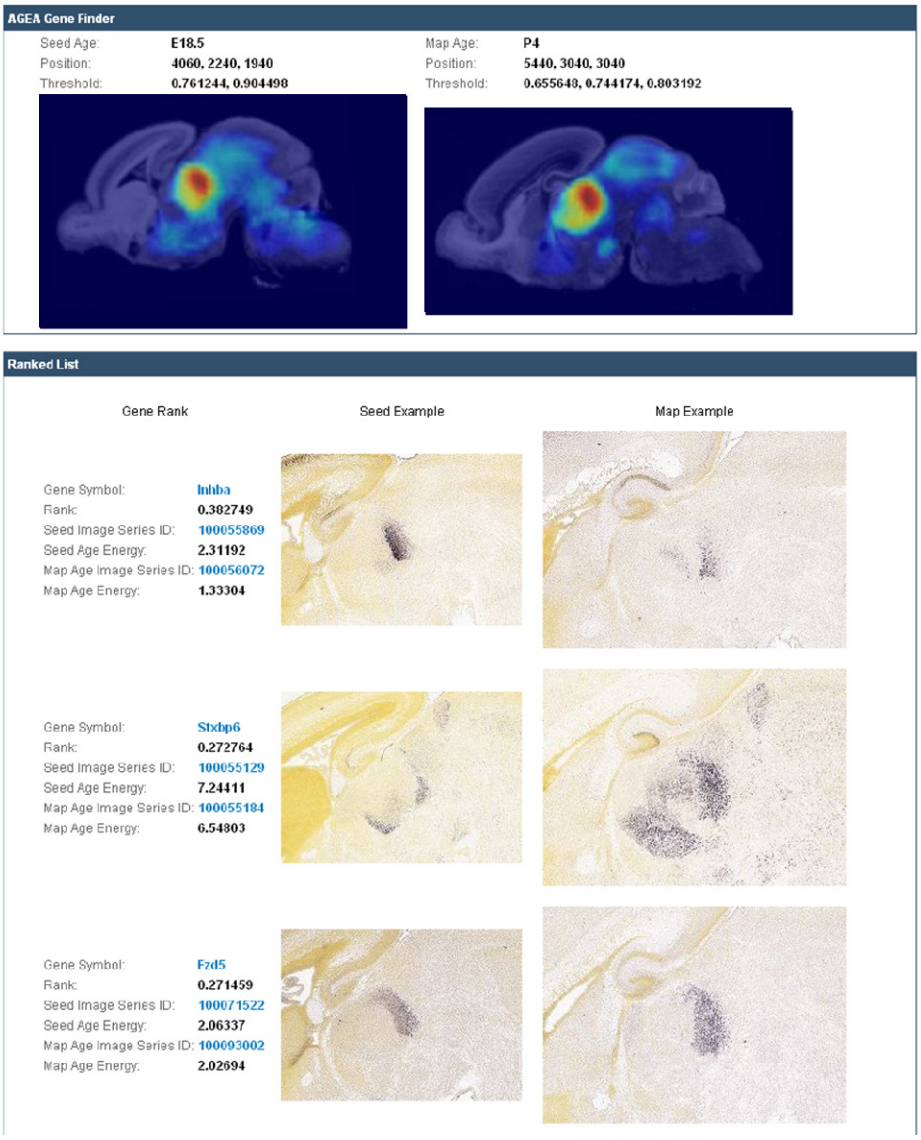

Figure 5. Example of inter-age Gene Finder return list for a seed voxel in the diencephalon at E18.5. The intra-age (E18.5) and inter-age (E18.5-P4) correlation maps have been thresholded with the lower bound set to "denominator" threshold of each map and the upper bound set to the maximum correlation of the map. In the maps the warm-colored areas roughly correspond with the "numerator" mask with the cool areas being the contrast region for search. Each row in the table corresponds to a gene in rank order. The region of interest in the zoomed in thumbnail was computed using the transform obtained in the Alignment module centered at the corresponding position to the maximum correlation location.

Brain Explorer

Brain Explorer® 2 is a desktop software application for viewing brain anatomy and gene expression data in 3D. It is integrated with the Allen Developing Mouse Brain Atlas in which you can view spatially registered gene expression data in 3D, view a fully interactive version of the associated reference atlas, view ISH expression data from different mouse developmental ages side-by-side, view ISH expression data from multiple genes superimposed on each other in 3D, navigate the high-resolution 2D ISH images, and link to associated gene metadata on the associated Web application (see **Figure 6**).

There are two main ways to obtain gene expression data for display in Brain Explorer: either perform a search and click on the "3D File" link in search results list or search for genes from Brain Explorer itself. The results in Brain Explorer shows the age, gene name, gene symbol along with a 3D summary thumbnail of the expression for each experiment matching the search. The expression thumbnail represents a maximum expression projection rendering: the denser the expression in a region the more "solid" the appearance. Reference atlas colors are additionally layered on top. Seven reference time points are available to view to Brain Explorer.

The main Brain Explorer window is divided into 3 main areas. The 3D view shows brain anatomy and gene expression in 3D. The structural ontology panel shows the color-coded hierarchy. The ontology can be switched between hierarchical and alphabetical mode. Each 3D view is based upon one underlying 3D volume; depending on the data resource, the volume maybe a 3D reconstruction of down-sampled serial sections, such as Nissl stained histology. In addition to the 3D polygonal structures, you can also turn on planar views of the underlying 3D volume and associated annotations.

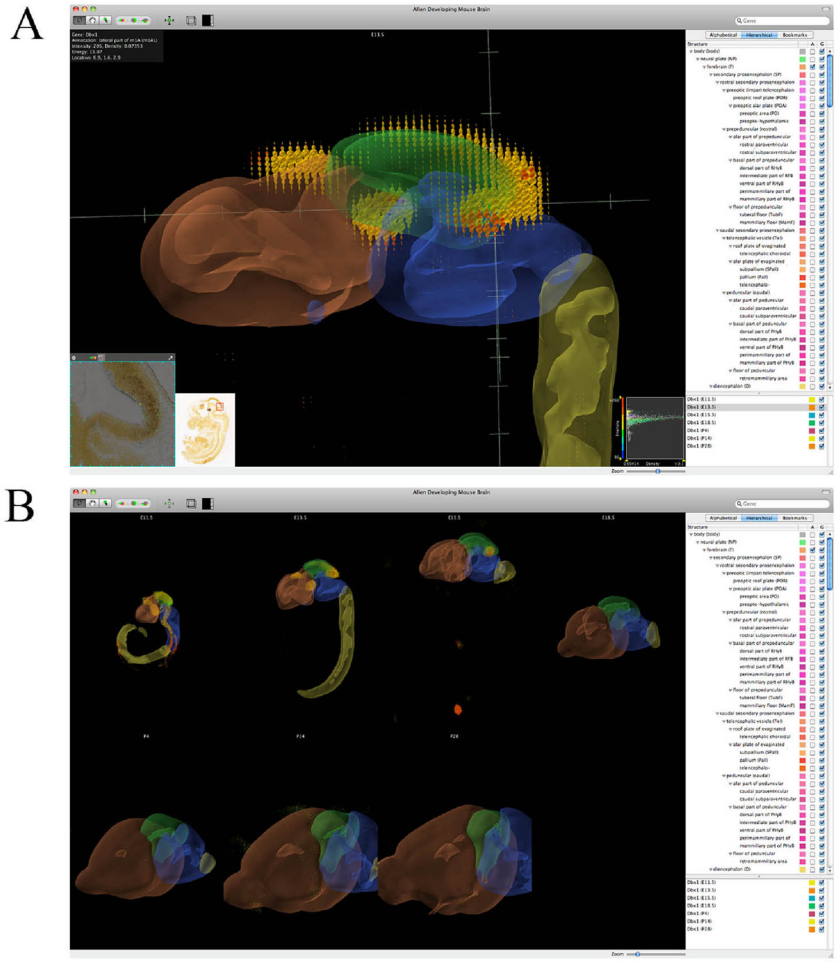

Figure 6. The Brain Explorer desktop application. (**a**) 3D surface views are shown for the level 1 structures in the reference atlas at each time point: forebrain (orange), midbrain (green), hindbrain (blue), and spinal cord (yellow). Expression of the gene *Dbx1* at the 3D grid resolution used for search and analysis is plotted as spheres at each time point. The size of the spheres corresponds to expression density and the color is mapped to expression intensity (blue-green is low, yellow-orange is medium, and red is high). (**b**) The user can click on spheres to view the original image from which the grid values were calculated.

References

1. Lein, E.S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168-76 (2007).
2. Temple, G. *et al.* The completion of the Mammalian Gene Collection (MGC). *Genome Res* **19**, 2324-33 (2009).
3. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J Mol Biol* **215**, 403-10 (1990).
4. Theiler, K. *The house mouse : atlas of embryonic development*, 178 p. (Springer-Verlag, New York, 1989).
5. Tecan. The GenePaint System. *Mol Biotechnol* **25**, 103-4 (2003).
6. Lau, C. *et al.* Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. *BMC Bioinformatics* **9**, 153-163 (2008).
7. Ng, L. *et al.* An Anatomic Gene Expression Atlas of the Adult Mouse Brain *Nature Neuroscience* **12**, 356-362 (2009).

Appendix 1 - *In situ* hybridization protocol

Day 1

Cycles	Time (min)	Volume (µl)	Reagent	Temp (°C)	Total Time (min)	Process
5	5	300	3% H2O2 in MeOH	25	25	Pre-Hyb
7	5	300	PBS (1)	↓	35	↓
2	5	300	0.2M HCl	↓	10	↓
4	5	300	PBS (2)	↓	20	↓
1	5	400	PK(+) buffer	↓	5	↓
1-2*	10	300	Proteinase K *	↓	10-20*	↓
7	5	300	PBS (3)	↓	35	↓
2	10	300	4% PFA (1)	↓	20	↓
7	5	300	PBS (4)	↓	35	↓
2	15	300	Hyb Solution	↓	30	↓
_	15	_	Temperature ramp-up	63.5	15	↓
1	_	300	Riboprobe addition	↓	_	Hybridization
1	330	_	Hybridization	↓	330	

Day 2

Cycles	Time (min)	Volume (µl)	Reagent	Temp (°C)	Total Time (min)	Process
5	5	300	5 x SSC	61	25	Stringency Wash
5	10	350	Formamide I	↓	50	↓
5	12	350	Formamide II	↓	60	↓
4	8	300	0.1 x SSC (1)		24	↓
1	8	300	0.1 x SSC (2)	25	8	
4	5	300	NTE (1)	↓	20	Post Hyb Blocking
3	5	300	20 mM Iodoacetamide (1)	↓	15	↓
3	5	300	20 mM Iodoacetamide (2)	↓	15	↓
4	5	300	NTE (2)	↓	20	↓
2	5	300	TNT (1)	↓	10	↓
3	5	300	4% Sheep serum (1)	↓	15	↓
3	5	300	4% Sheep serum (2)	↓	15	↓
4	5	200	TNT (2)	↓	20	↓
2	10	300	TNB blocking buffer	↓	20	↓
2	5	200	TNT (3)	↓	10	↓
2	5	300	Maleate wash buffer (1)	↓	10	↓
2	10	350	Blocking reagent	↓	20	↓
2	5	300	Maleate wash buffer (2)	↓	10	↓
2	5	250	TNT (4)	↓	10	↓
3	5	350	TMN	↓	15	↓
4	5	200	TNT (5)	↓	20	↓
4	10	300	TNB blocking buffer	↓	40	↓
2	30	350	Anti-DIG-POD	↓	60	Colorimetric Detection
6	5	250	TNT (6)	↓	30	↓
1	30	250	Tyramide-biotin	↓	30	↓
6	5	300	Maleate wash buffer (3)	↓	30	↓
2	20	350	Neutravidin	↓	40	↓
6	5	300	Maleate wash buffer (4)	↓	30	↓
4	5	250	TNT (7)	↓	20	↓
2	5	400	TMN	↓	10	↓
2	15	350	BCIP/NBT	↓	30	↓
1	10	350	BCIP/NBT	↓	10	↓
3	_	400	System liquid (1)	↓	10	↓
1	_	300	NTE (3)	↓	5	↓
1	10	250	4% PFA (2)	↓	10	↓
1	_	400	System liquid (2)	↓	10	↓

* Concentration and time dependent upon time point (see **Table 3**)