

## **TECHNICAL WHITE PAPER: *IN SITU* HYBRIDIZATION**

Non-human primates provide experimentally tractable model systems that allow carefully controlled analysis of important developmental processes relevant for human health and disease. The goal of the NIH Blueprint Non-Human Primate (NHP) Atlas is to provide a detailed analysis of gene expression in the normal developing primate brain in brain regions associated with high-level cognitive, mnemonic and emotional functions as well as neurodevelopmental disorders linked to disruption of these capacities.

One of the data components of the NHP Atlas is a focused study of cellular resolution localization of gene expression at different developmental stages in specific anatomical regions and cell types, as well as throughout a complete adult hemisphere, using colorimetric *in situ* hybridization (ISH). This profiling was performed in two major phases. The first is a developmental timeseries spanning postnatal development in a set of specific forebrain regions. These regions match those assayed using DNA microarrays in the other major portion of the NHP atlas. The second is a cellular resolution analysis of a targeted gene set across entire hemispheres of the entire adult rhesus brain. A standardized processing pipeline was used to section tissue samples, perform histological staining, generate standardized colorimetric ISH data, and acquire images from the resulting stained tissue.

For cellular resolution localization of gene expression in specific anatomical regions and cell types across postnatal development, genes represented in this dataset comprised several categories of broad scientific and clinical interest, including highly restricted marker genes for specific anatomical regions and cell types, gene families important to neural function, disease-related genes and genes important in the comparative genomics field. ISH data are presented for medial prefrontal cortex, primary visual cortex, hippocampus, amygdala and ventral striatum, with genes for each structure assayed across the four timepoints neonate (birth), infant (3 months), juvenile (12 months), and post-pubertal adult (48 months) in three male specimens.

For cellular resolution localization of gene expression throughout the entire adult brain, ISH data are generated serially across complete hemispheres of three post-pubertal adult (48 months) male specimens, focusing on cellular marker genes, genes with cortical area specificity and gene families important to neural function. Complementing the ISH data across a complete hemisphere are the histological stains Nissl, acetylcholinesterase and SMI-32 that combined with the ISH data provide an unprecedented cellular resolution view into the post-pubertal adult brain.

This document describes details about the generation of these ISH data, starting with the specimen preparation through histological data generation and data processing as well as the methodology employed to score the ISH expression data for the Annotation Search.

### **PIPELINE OVERVIEW**

High-throughput processes for generation of ISH-based gene expression data were developed at the Allen Institute for Brain Science for the production of the [Allen Mouse Brain Atlas](#), a genome-wide atlas of gene expression in the adult mouse brain (Lein et al., 2007). The process, equipment and workflow for generation of gene expression data in macaque closely follows that described for generation of the Allen Mouse Brain Atlas (see Supplemental Methods 1 in Lein et al., 2007) with some adaptation to manage specific challenges posed by working with developing macaque tissue. For example, modifications were made to ISH

methodology (e.g., altering PK concentration for different ages, altering TSA (tyramide signal amplification) reagent), image acquisition and data processing capacity to accommodate larger tissue sections.

A Laboratory Information Management System (LIMS) was used to manage all information related to experimental design, slide, and image tracking as described previously (Lein et al., 2007). In addition to modifications to laboratory production processes mentioned above, the LIMS was updated to accommodate information management needs specific to macaque tissue samples. The LIMS was also used to view raw image data and perform image Quality Control (QC) following initial image acquisition.

Briefly, the workflow involved the following: Validated tissue and probes were coordinated into work packets, where each packet consisted of tissue and probes that were progressed as a discrete unit throughout the entire process of sectioning, ISH, image acquisition and image processing. QC metrics were established for image and data quality. Images passing QC were then released for public display. This workflow enabled the systematic generation of data for each gene, and where possible data for a particular gene/structure (multiple ages and replicates) were processed simultaneously to allow cross-comparison between samples.

All processes associated with data production, including tissue receipt and storage, solution preparation, probe preparation, colorimetric ISH, histological staining, and equipment and other laboratory maintenance functions were governed by Allen Institute Standard Operating Procedures (SOPs) with revision control, and complied with appropriate health and safety precautions.

## GENE SELECTION

For the ISH component of the project, gene selection was based on four major categories constituting thematically interesting datasets for a broad user community. These overlapping classes included the following, with candidate selection biased in favor of genes contained in multiple categories, and with evidence for developmental regulation:

- (1) **Gene markers delineating highly regionalized anatomical subdomains, subnuclei and/or cell types.** Since very little non-human primate transcriptional data was available, these markers were primarily identified through mining of extant data resources in rodents, including the [Allen Mouse Brain Atlas](#), other ISH and microarray-based public data repositories, and literature. These cellular markers generally aimed to sample the major cytoarchitectural features of each structure, including, for example, markers for cortical layers, hippocampal subfields or amygdalar subnuclei, as well as markers for specific cell classes with less discrete localization including interneurons, glial and vascular markers.
- (2) **Gene families important for neural function.** These gene families included ion channels, G-protein-coupled receptors (GPCRs), transporters, synaptic proteins, membrane proteins, and peptide or protein ligands. Particular emphasis was placed on ion channels and genes related to GABAergic neurotransmission.
- (3) **Disease-related genes.** Genes were selected by searching through available literature to identify genes (i) conferring disease susceptibility, (ii) known to be involved in physiological pathways implicated in disease, or (iii) encoding known drug targets. The diseases included autism, schizophrenia, epilepsy, microcephaly, neurodegenerative diseases, depression, anxiety, and intellectual and developmental disabilities.
- (4) **Comparative genomics.** This category comprised genes identified in the literature as showing accelerated evolution, being under positive selection, or showing microarray-based gene expression differences either between rodents and primates or humans.

The complete gene set is provided in the [Documentation](#) tab in the online application.

## TISSUE SPECIMENS

Frozen postmortem tissue samples from male rhesus macaque (*Macaca mulatta*) were provided by the California National Primate Research Center (CNPRC; <http://www.cnprc.ucdavis.edu>). For the purpose of generating histological and ISH data, as well as RNA analysis by microarrays, brain regions were systematically collected from well-characterized rhesus monkeys born and raised at the CNPRC in outdoor, half-acre enclosures that provide a naturalistic setting and normal social environment. Extensive health, family lineage and dominance information was maintained on all animals in the outdoor enclosures. All procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at UC Davis.

### Structure isolation and processing

After dissection, brains were sectioned into coronal slabs approximately 1 to 1.5 cm in thickness. These slabs were further dissected to yield blocks containing structures of interest (medial prefrontal cortex, primary visual cortex, hippocampus, amygdala and ventral striatum) contained within an area that will fit on a standard 1x3 inch microscope slide. These blocks were photo-documented, frozen on dry ice and stored at -80°C. For a subset of animals, regions of interest were selectively dissected from one hemisphere for RNA isolation and microarray analysis and frozen at -80°C until further processing.

### Complete hemisphere isolation and processing

Following magnetic resonance imaging (MRI), 48-month male rhesus macaques were prepared for brain acquisition. Animals were injected with heparin and then euthanized by an injection of a barbiturate. Room temperature saline was used for transcatheterial perfusion for 5 minutes followed by 4°C saline for 5 minutes. Brains were then extracted from the skull and bisected. After bisection, hemispheres were sectioned into coronal slabs approximately 1.7 to 2 cm in thickness. These blocks were photographed, frozen on dry ice and stored at -80°C before shipment to the Allen Institute.

### Tissue receipt and validation

Upon receipt at the Allen Institute, tissue was again photodocumented and information entered into LIMS for future tracking throughout the data pipeline. Tissue was stored at -80°C until removed for sectioning. Prior to sectioning for ISH, all tissue samples were tested to confirm that the region of interest was present based on expected cytoarchitecture. Tissue samples that failed these criteria were not used in the study.

## IMAGING

Magnetic resonance (MR) images were collected prior to dissection for anatomic visualization of each brain used for ISH on a complete hemisphere. Three 48-month male rhesus macaques underwent MRI using a GE Signa 1.5T MR Scanner and an SPGR sequence. The following parameters were used for the T1 weighted imaging:

Orientation: coronal  
Field of View (FOV): 160mm X 160mm  
Image size: 256x256  
Thickness: 1mm  
Spacing: 0  
TE: 7.9  
TR: 21  
FS: 15000  
Number of Excitations (NEX): 4  
Pixel size: 0.625mm

## PROBE DESIGN AND SYNTHESIS

For labeling target mRNA in tissue sections using ISH, digoxigenin-labeled riboprobes were designed and synthesized according to specific criteria. In general, the design and synthesis process followed previously described methods used to generate probes for the Allen Mouse Brain Atlas (Lein et al., 2007) with some modification. Briefly, probes were designed to be between 400-1000 bases in length (optimally >600 bases) and to contain no more than 200 bp with >90% homology to non-target transcripts, using sequence information obtained from NCBI RefSeq ([www.ncbi.nlm.nih.gov/RefSeq](http://www.ncbi.nlm.nih.gov/RefSeq)) and a semi-automated process using Primer3 software (Rozen & Skaletsky, 2000). In addition, to allow comparability of mouse and human gene expression datasets, probes were designed to overlap with the existing Allen Mouse Brain Atlas probe when the mouse and macaque genes were orthologous. Riboprobes were synthesized using standard *in vitro* transcription (IVT) reactions based on PCR templates prepared from cDNA clones (NIH Mammalian Gene Collection, Open Biosystems, Huntsville, AL) or cDNA synthesized from brain total RNA. cDNA was prepared from brain RNA from cerebellum using Superscript III RTS First-Strand cDNA Synthesis Kit (Invitrogen, Carlsbad, CA) to provide templates for PCR.

PCR primers were obtained from Integrated DNA Technologies (Coralville, IA) at a final concentration of 10  $\mu$ M, and designed with GC content between 42% to 62% and an optimal size of 22nt with lower and upper limits of 18 and 26 nt, respectively. For cDNA clones, the clone sequence was compared with RefSeq sequences and consensus sequences with >98% homology across 80% of the total length were used to develop probes. When a clone was used as a template, a single PCR was used requiring only a forward and reverse primer with an additional SP6 RNA polymerase binding sequence (GCGATTTAGGTGACACTATAG). When using brain cDNA as a template, probes were generated against sequences within a region 3000 bp from the 3' end using 3 primers: forward, reverse, and a nested reverse primer containing the SP6 RNA polymerase binding sequence. cDNA primers underwent a BLAST analysis to verify amplification of only target sequence. All cDNA reactions were run on the Bioanalyzer for quality control.

Standard conditions for PCR and IVT reactions were as described. IVT reactions were diluted to working stocks of 30 ng/ $\mu$ l with THE (0.1mM Sodium Citrate pH 6.4, Ambion). Aliquots were stored in single-use volumes to minimize freeze/thaw cycles. IVT dilutions were stored at -80°C. For hybridization, the probe was diluted 1:100 (to 300 ng/ml) into *in situ* hybridization buffer (Ambion) in 96-well ISH Probe Plates. Each well provided probe for one ISH slide. Probe plates were stored at -20°C until used in an ISH run.

All PCR and IVT products were run on the Bioanalyzer for size and peak characteristic quality control. Specifically, PCR products that were not of the correct size (+/- 100bp) or that showed multiple products were not used to generate riboprobes. IVT products that were shorter than their predicted size were not used. It is common to see IVT products that run slightly larger than their predicted molecular weight, or as multiple peaks, due to secondary structure of the RNA.

## TISSUE SECTIONING

### Individual structures

Frozen tissue samples were cryosectioned using Leica CM3050 S cryostats (object temperature, -10°C; chamber temperature, -15°C) at 18  $\mu$ m (amygdala and ventral striatum) or 20  $\mu$ m (hippocampus, cortical areas) thickness in the coronal plane from anterior to posterior. One or two sections were placed on a positively charged Superfrost Plus 1" x 3" microscope slide (Erie Scientific Co, Portsmouth, NH), pre-printed with a unique identifying barcode for tracking. Specimen numbers were also printed on to the slides for tracking purposes. For samples in which two sections/slide were used, sections were serially sectioned onto sets of 50 slides, such that sections 1 and 51 go on slide 1, sections 2 and 52 go on slide 2 ... sections 50 and 100 go on slide 50, sections 101 and 151 go on slide 51 and so on. Each gene was processed on a slide series, representing the same position on sequential sets of 50 slides (e.g., slide 1, 51, 101), thereby giving uniform sampling at ~1 mm spacing across the sample block (e.g., 50 serial sections x 20  $\mu$ m = 1 mm). In practice, each series consisted of 1 to 5 slides, determined by the tissue sample thickness. For anatomical and cytoarchitectural reference, two of the 50 sectioning series (series 2 and series 26) were designated for Nissl staining so that a Nissl reference was available every ~500  $\mu$ m throughout the tissue block. Two additional series were used for positive control genes and two series were reserved as backups for repeating data that might fail QC. Each of the remaining 44 series was hybridized to a single gene probe.

## Complete hemisphere

Frozen tissue samples were cryosectioned using Leica CM3050 S cryostats (object temperature,  $-10^{\circ}\text{C}$ ; chamber temperature,  $-15^{\circ}\text{C}$ ) at  $20\ \mu\text{m}$  thickness in the coronal plane from anterior to posterior. One section was placed on a 2" x 3" microscope slide, pre-printed with a unique identifying barcode for tracking. Specimen numbers were also printed on to the slides for tracking purposes. Sections were serially sectioned onto sets of 50 slides, such that sections 1 and 51 go on slide 1 and slide 51, respectively, sections 2 and 52 go on slide 2 and slide 52, respectively, and so on. Each gene from the hemisphere 46 gene set was processed on a slide series, representing the same position on sequential sets of 50 slides (e.g., slide 1, 51, 101), thereby giving uniform sampling at  $\sim 1\ \text{mm}$  spacing across a complete hemisphere (e.g., 50 serial sections x  $20\ \mu\text{m} = 1\ \text{mm}$ ). For anatomical and cytoarchitectural reference, two of the 50 sectioning series (series 1 and series 26) were designated for Nissl staining so that a Nissl reference was available every  $\sim 500\ \mu\text{m}$  throughout the hemisphere. In addition, two sectioning series (series 2 and series 27) were designated for acetylcholinesterase and SMI-32 staining, respectively, so that an acetylcholinesterase and SMI-32 reference was available every  $\sim 1\ \text{mm}$  throughout the hemisphere. Each of the remaining 46 series was hybridized to a single gene probe.

## Post-section processing

Following sectioning, slides designated for ISH were allowed to air dry and tissue was fixed, acetylated and dehydrated according to standard protocols as described (Lein et al., 2007). Briefly, tissue was fixed for 20 minutes in 4% neutral buffered paraformaldehyde (PFA) and rinsed in 1x PBS, acetylated for 10 minutes in 0.1M triethanolamine with 0.25% acetic anhydride, and subsequently dehydrated using a graded series of 50%, 70%, 95% and 100% ethanol. Slides that passed section QC examination were stored at room temperature in Parafilm-sealed slides boxes until use.

## HISTOLOGICAL STAINING

### Nissl

After brain tissue was sectioned, slides were stored at  $37^{\circ}\text{C}$  for 1 – 5 days and were removed 5 – 15 minutes prior to staining. Sections were defatted with xylene or the xylene substitute Formula 83, and hydrated through a graded series containing 100%, 95%, 70%, and 50% ethanol. After incubation in water, the sections were stained in 0.213% thionin, then differentiated and dehydrated in water and a graded series containing 50%, 70%, 95%, and 100% ethanol. Finally, the slides were incubated in xylene or Formula 83, and coverslipped with the mounting agent DPX. After drying, the slides were analyzed microscopically to ensure staining quality. Slides were stored at room temperature in slide boxes before being cleaned in preparation for digital imaging.

### SMI-32

Immunohistochemical staining using SMI-32, an antibody that reacts with an epitope on non-phosphorylated neurofilament H proteins, was used to visualize cell bodies and processes in single hemisphere tissue sections. Stored fresh frozen tissue sections on slides were taken out of storage at  $-80^{\circ}\text{C}$ , equilibrated to room temperature and fixed with 100% ice-cold acetone. Sections were then rehydrated in 1X PBS with potassium, pH 7.4 (1:10 dilution of 10X PBS, Ambion, Austin, TX). Non-specific binding was blocked with 5% horse serum (Vector Laboratories, Burlingame, CA) in PBS and permeabilized with 0.3% Triton-X. Endogenous peroxidase activity was blocked in 3% hydrogen peroxide in methanol. Sections were then incubated in 1:1000 dilution of mouse anti-SMI-32 (Covance, Inc., Princeton, NJ) for 1 hr. Sections were rinsed in PBS-Tween 20 (0.0005%), then incubated in 1:100 biotinylated horse anti-mouse IgG secondary antibody (Vector Laboratories), rinsed in PBS-Tween, and incubated for 30 min in ABC (Vectastain, Vector Laboratories). The reaction product was visualized with 0.5% DAB (Sigma-Aldrich), activated with 0.003% hydrogen peroxide for 6 minutes. Sections were dehydrated through graded alcohols (50%, 70%, 95% and 100% ethanol), cleared with Formula 83 and coverslipped with DPX. Slides were examined for stain quality and stored at room temperature prior to digital imaging.

### Acetylcholinesterase

A modified acetylcholinesterase (AChE) protocol was used to help delineate subcortical structures at high resolution in single hemisphere tissue sections. Unlike AChE staining in fixed tissue, staining in fresh frozen

tissue does not elucidate cholinergic fibers commonly seen in the cortex, but instead provides demarcation of various subcortical nuclei. AChE staining was performed using a direct coloring thiocholine method combined with a methyl green nuclear counterstain to improve tissue visibility (Karnovsky and Roots, 1964). Glass slides with fresh frozen tissue sections were removed from 4°C, allowed to equilibrate to room temperature, fixed in 10% neutral buffered formalin (NBF) and washed briefly in ultra-pure water. Sections were then incubated for 25 minutes in a solution of acetylthiocholine iodide, sodium citrate, cupric sulfate, and potassium ferricyanide in a 0.1M sodium acetate buffer (pH 6.0), washed in 0.1M Tris-HCl buffer (pH 7.2), and incubated with 0.5% diaminobenzidine (DAB) in 0.1M Tris-HCl with 0.03% hydrogen peroxide for 8 minutes. Slides were incubated in 0.2% methyl green, briefly dipped in 100% ethanol, cleared with Formula 83 and coverslipped with DPX.

## STANDARDIZED HIGH THROUGHPUT COLORIMETRIC ISH

An automated technology platform for standardized high throughput histological slide processing was used to generate cellular resolution ISH data on rhesus macaque tissue sections. Digoxigenin-based riboprobe labeling, coupled with TSA amplification and alkaline phosphatase-based colorimetric detection was used to label target mRNAs in expressing cells. Detailed descriptions of the high-throughput platform, protocols, and reagent preparation are available elsewhere (Lein et al., 2007, Supplemental Methods 1).

Briefly, slides containing tissue sections were placed in flow-through chambers on temperature-controlled racks on computer-controlled Tecan liquid handling platforms for sequential application of solutions. Initial steps in the protocol blocked endogenous peroxidase activity and permeabilized the tissue for probe penetration. Digoxigenin-labeled probes were subsequently hybridized to target mRNA, and after a series of washes to eliminate excess probe the remaining bound probe was subjected to a series of enzymatic reaction steps to detect and amplify the digoxigenin signal. First, a horseradish peroxidase (HRP)-conjugated anti-digoxigenin antibody was added, followed by biotin-coupled tyramide that was converted by HRP to an intermediate that binds to cell-associated proteins at or near the HRP-linked probe. Neutravidin conjugated with alkaline phosphatase (AP) was then bound to biotin and BCIP/NBT was added for colorimetric detection. A blue/purple particulate precipitate formed as a result of the enzymatic cleavage of BCIP by AP and subsequent indole reaction with NBT. Finally, the colorimetric reaction was terminated by washing with EDTA and fixation with 4% PFA. This entire process occurred over the course of approximately 23.5 hours on the Tecan automated platform. ISH data from medial prefrontal cortex, primary visual cortex, hippocampus, amygdala and ventral striatum was generated on 1x3 slides. ISH on a complete hemisphere was generated on 2x3 slides.

Each ISH run contained several positive controls. Macaque-specific probes against GAD1 and CALB1 were included as robust and moderately expressed control genes, respectively, for ISH on individual structures. For complete hemisphere ISH, GAP43 on adult human tissue was included as a robust expressed control gene. This control was routinely used on all 2x3 ISH runs. All ISH had a *Drd1a* positive control on mouse brain tissue (the same control used throughout the Allen Mouse Brain Atlas project) to provide verification of a successful ISH run. All ISH had a negative control (no probe) slide included as an indication of background for each ISH run.

To reduce background signal, an acid alcohol wash step was performed after completion of the hybridization process. Slides were rinsed 4 times (1 minute each) in acid alcohol (70%, adjusted to pH = 2.1 with 12N HCl) and rinsed 4 times in milliQ water (1 minute each). Acid alcohol and water solutions were refreshed every fourth rack to ensure that all slides were rinsed in clean solution.

## IMAGE ACQUISITION

Image acquisition was performed using ScanScope scanners (Aperio Technologies, Inc.; Vista, CA). The line scan camera continually adjusted for focus based on a variable number of focus points and provided advantages over tile-based image acquisition platforms for large tissue sections that tended to have more variation in height. The ScanScope scanner used a 20x objective that is downsampled in software to

minimize data volume acquired for this project. The downsampling provided comparable image resolution (approximately 1  $\mu\text{m}/\text{pixel}$ ) to the ICS scanning systems with 10x objectives used for the Allen Mouse Brain Atlas (Slaughterbeck et al., 2007).

## DATA PROCESSING

Once images were acquired, the Informatics Data Pipeline (IDP) managed image preprocessing, image QC, ISH expression detection and measurement, histological processing (Nissl, SMI-32, AChE), annotation QC and public display of information via the Web application. The IDP has been described in detail previously (Dang et al., 2007), and was modified slightly for processing rhesus macaque histological images.

For image preprocessing, scanned ISH and histological images were converted and background corrected to provide more consistent white background intensities and orientation across samples. ScanScope images were first converted to JPEG 2000 format, and then orientation-adjusted and white balanced. The final products were images in a JPEG 2000-compressed format for further pipeline processing and analysis.

A major goal of ISH data presentation was to provide users with a quantified representation of the data, similar to the color-coded “heat mask” representation used for the [Allen Mouse Brain Atlas](#). To detect and quantify expressing cells on ISH images of tissue sections from the individual structures, adaptive image processing techniques were applied to 10x full-resolution ISH images. An algorithm based on techniques used for the Allen Mouse Brain Atlas (Lein et al., 2007; Supplemental Methods 2) was significantly redesigned to accommodate the full image resolution needs that resulted from the macaque image sizes. The resulting module produced a mask of detected expressor objects and a set of numerical values describing the statistical attributes of gene expression. The mask image with measured intensity of expression was then pseudo-color coded and converted to AFF file format for Web display. The raw ISH data, quantified heat mask representation, and closest Nissl-stained section were presented for each tissue section.

## QUALITY CONTROL FOR WEB-BASED DATA PRESENTATION

Additional QC steps during the course of data processing described above ensure the release of accurate, high quality data to the publicly accessible web portal. Once image preprocessing was complete, an image quality control step ensured the images were in good focus, and provided an initial indication of the presence of ISH signal. If focus criteria were not met, the images were failed and the appropriate slides were rescanned. If focus criteria were met, the images were passed and processed through the IDP for ISH expression detection. From time to time, slides may be rescanned to improve image quality. In these cases images in the database are replaced with the most recent scan of the original slide.

## ANNOTATION SEARCH

In order to provide detailed spatiotemporal searching and data mining capabilities to the ISH data from the medial prefrontal cortex, primary visual cortex, hippocampus, amygdala and ventral striatum, a qualitative scoring strategy was used to discretize specific characteristics of the expression data at each developmental stage. This scoring was designed to provide a single set of expression scores for each individual structure/age for the purpose of allowing searches for differential gene expression between neuroanatomical structures and across ages, and therefore represents the consensus pattern across the three replicate data series for each structure and age rather than an independent set of scores for each replicate. For each major brain region the structure set used for scoring represented major definable subnuclei or laminar compartments, capturing both the functional architecture of the structure as well as the specific cellular components that make up that subdivision (e.g., GABAergic interneurons in non-pyramidal sublayers of CA subfields of the hippocampus).

The scoring paradigm was designed to capture two fundamental aspects of cellular labeling observed in ISH data. The first metric was the cellular expression level, related directly to the staining intensity observed in

labeled cells (on a 0-5 scale). As described below, this involved expert scoring of the representative cellular labeling based on algorithmically quantified signal “heat maps,” which are also available for visualization for each ISH image. The second metric was the density of labeled cells in a particular structure (on a 0-5 scale), designed to capture the proportion of labeled neurons observed within particular brain nuclei. This cellular distribution is highly meaningful, in that sparse and scattered cellular distributions (e.g., scored as density of 1 or 2) typically represent selective expression in glial cells or GABAergic interneurons, whereas higher densities typically represent primary excitatory neurons (or at least include these neurons). The combination of fine neuroanatomical distribution and relative density provided a great deal of predictive power for the cell types using a particular gene transcript, and therefore some indication of that gene’s function in the brain region as a whole. Differential expression across the developmental time points assayed in this data set can provide some insight into functional roles for genes in developmental processes associated with maturation of functional circuitry from birth through adulthood.

### Data Scoring Methods

To characterize detailed cellular gene expression patterns in the developing macaque brain, qualitative expert annotation of colorimetric ISH was performed for each isolated structure (medial prefrontal cortex, primary visual cortex, hippocampus, amygdala and ventral striatum) across the four developmental time points (0, 3, 12 and 48 months). For each major brain region a set of substructures (referred to below as regions of interest or ROIs) clearly identifiable in the Nissl-stained sections were established as the structural basis for the scoring paradigm.

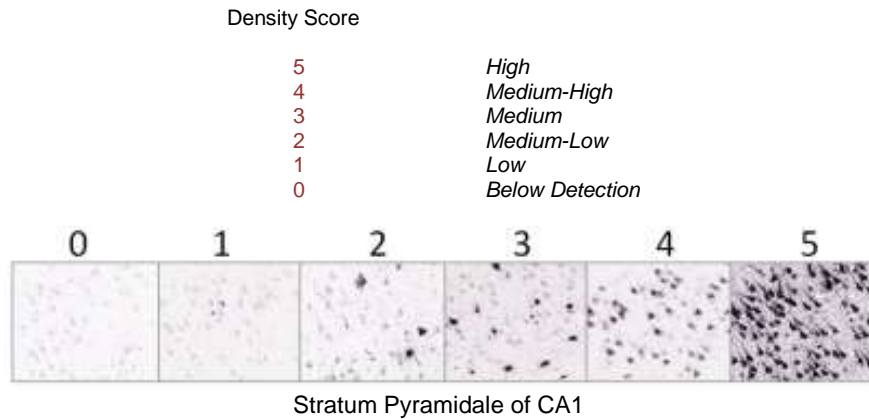
The goal of the scoring was to provide a quantized data set for each structure/age, with a single set of scores representing the expression characteristics of a representative specimen from three independent biological replicate series available for each age. For each specimen at a particular age, ROIs were initially identified based on the Nissl images and available reference literature and atlases. The corresponding ISH gene expression data were then assessed for tissue and stain quality as well as consistency between the replicates at that age, and one of the replicates was selected for scoring as an image series that provides a representative, high quality labeling pattern in a data set that includes all ROIs for that structure. For this image series, the density and intensity of expression within each ROI was assessed using a standardized semi-quantitative scoring system, allowing for direct comparisons across gene, region and/or age. Data for a particular gene was scored by a single annotator, with all time points scored together where possible to provide scoring consistency across ages. Clear ISH artifacts and individual data sets displaying different patterns from the other replicates were excluded.

For each specimen, the interleaved Nissl-stained sections were used to delineate the ROIs across the image series constituting the entire structure. These delineations were then used to aid identification of these ROIs for scoring of ISH patterns on nearly adjacent tissue sections, where the cellular labeling patterns are frequently insufficient for structural delineation by themselves. In order to provide a representative score for each structure, all slides containing that structure for a given gene were analyzed to provide a single representative score for that ROI.

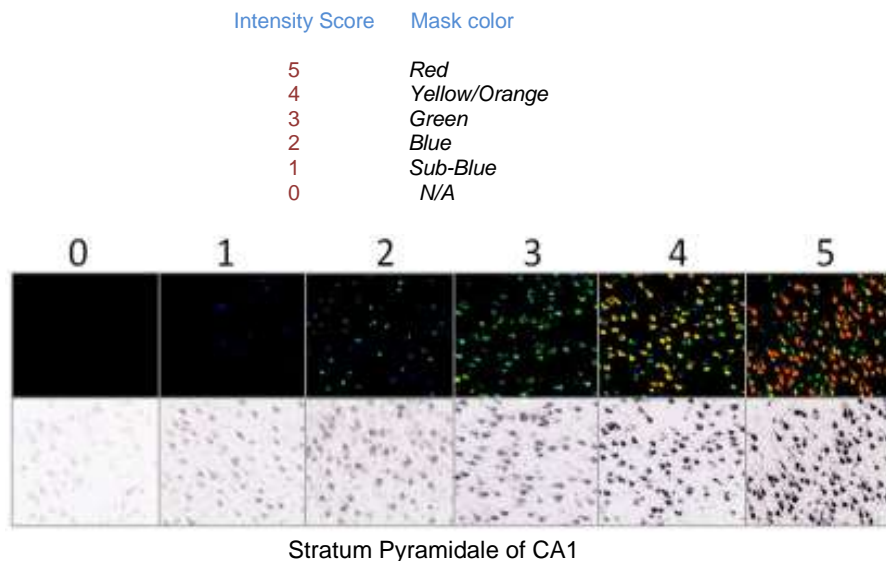
### Gene expression scoring

The density and intensity of gene expression (*D*, *I*) was assessed using a semi-quantitative scoring method. For expression **density**, scores ranged from 0 (below detection) to 5 (high density), and were assessed manually.





For expression **intensity**, scoring was based on a color-coded heat map image of the ISH data based on an automated quantification algorithm that segments the image to identify clusters of pixels (cells) with signal over background and color-codes these cells by labeling intensity. These color coded “heat masks” are also available for viewing with the ISH data. Scores ranged from 0 to 5 based on visual integration of representative cellular labeling in the ROI (see key below). N/A indicated either the absence of the ROI in the data set scored, or ambiguity for presence of artifacts.



For each structure, a scoring matrix was constructed based on the structural ontology and density and intensity (D, I) scores.

**REFERENCES**

Dang CN *et al.* (2007) The Allen Brain Atlas: Delivering Neuroscience to the Web on a Genome Wide Scale. In: *Data Integration in the Life Sciences* 17-26 (Springer Berlin).

Karnovsky MJ, Roots L (1964) A “Direct Coloring” Thiocholine Method for Cholinesterases. *J Histochem Cytochem* 12:219-221.

Lein ES *et al.* (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168-76.

Paxinos, G, Huang, XF, and Toga, AW (2000). *The Rhesus Monkey Brain in Stereotaxic Coordinates* (San Diego, CA; Academic Press).

Rozen S & Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology* 365-386 (Humana Press).

Slaughterbeck CR *et al.* (2007) High-throughput automated microscopy platform for the Allen Brain Atlas. *J. Assoc. Lab Automation* 12:377-383.