

# ALLEN Developing Mouse Brain Atlas

## TECHNICAL WHITE PAPER: EXPERT ANNOTATION OF ISH DATA

### OVERVIEW

The Allen Developing Mouse Brain Atlas is an extensive resource on gene expression over the course of brain development from embryonic through postnatal stages, providing both spatial and temporal information about gene expression. *In situ* hybridization (ISH) data has been generated for approximately 2,000 genes at each of seven timepoints and the data is presented online as image data.

Sophisticated tools have been developed to automatically detect gene expression in these images and assign it to its spatial location as either part of a defined reference atlas structure or as spatial coordinates in a 3-D space. These methods allow a basic search for candidate genes in the Atlas, acting as the basis for the tools NeuroBlast (to find genes with similar profiles), AGEA (to find genes associated with expression at a particular voxel) and Anatomic and Temporal searches (to find genes with particular enrichment at a given brain region or age). While these automatically-generated semi-quantitative tools provide an informative global view of gene expression in the brain, these techniques provide information primarily about larger brain regions, with limited resolution for precise brain areas.

Users of the Allen Developing Mouse Brain Atlas who are unfamiliar with mouse developmental anatomy may also benefit from an expert annotation of the ISH data, providing an interpretation between the ISH signal and the reference atlas ontology. Therefore, expert-guided manual annotation of the ISH data has been performed to accurately assign gene expression metrics to specific brain areas. This work has been performed using the expertise of Dr. Salvador Martinez at the Neurosciences Institute, Miguel Hernandez University and CSIC in Alicante, Spain.

### EXPRESSION PATTERN ANNOTATION

#### Brain Ontology

The reference atlas associated with the Allen Developing Mouse Brain Atlas forms the basis for this annotation effort. This reference atlas utilizes a single unified ontology for all ages, based upon a topological ontogenetic viewpoint, such that brain structures in different developmental timepoints can be roughly related by ontology despite the existence of transient or migratory structures. The overall ontology consists of “levels” of organization from 0-13, with Level 0 representing neural plate, and Levels 11-13 representing individual brain nuclei. A description of the basic organization ontology can be found in the *Reference Atlas* white paper in the [Documentation](#) section of the Allen Developing Mouse Brain Atlas. The Allen Developing Mouse Brain Atlas reference atlas ontology was updated for the June 2013 release to include additional brain structures and to correct ontological classification based upon more recent concepts of neurodevelopment. However, the ontology utilized for the annotation effort relied upon an intermediate version of the reference atlas ontology. The annotation ontology is completely consistent with the reference atlas ontology down to Level 5; after Level 5, there are a variety of differences in the organization and composition of the ontology. In order to

enable comparison of the two ontologies, a mapping between structures of the two ontologies has been made available under [supplementary material](#).

It is not possible to identify and annotate the brain uniformly down to the level of nuclei at the earliest ages of embryonic development in this dataset, due to the timing of maturation of many brain regions. For the purpose of annotation of embryonic data for E11.5, E13.5, E15.5 and E18.5, the annotators determined the equivalency of regions for these ages across the levels of the ontology levels. Therefore, the annotation effort was largely focused on providing annotation to Level 9 and 10, which represent the 3-D complexity of brain presumptive territories identifying neural regions and subregions distributed in the different “strata”, such as the ventricular zone and mantle zone (including periventricular, intermediate, and superficial strata).

### Annotation of Expression

The gene expression revealed by *in situ* hybridization was annotated by experts in developmental neuroanatomy. Complete sets of image series of E11.5, E13.5, E15.5 and E18.5 developmental stages were manually annotated. Three metrics were used to describe gene expression: intensity, density, and pattern. These metrics were scored for each brain region according to a standard scheme (see below) and entered into the hierarchically organized ontology of anatomical structures.

At each developmental stage, annotation was performed for anatomic structures belonging to the most detailed level of the ontology (down to Level 10) that were identifiable as exhibiting differential expression. For example, if the pallium exhibited one homogenous expression pattern but the subpallium exhibited a different pattern, annotation would be recorded for each of these structures individually. When a given brain structure was annotated, that annotation data was intended to represent the complete set of “child” or “descendent” structures of this level in the hierarchical tree (corresponding with an anatomical region), such that the expression call for pallium would then apply to its children: medial pallium, lateral pallium, ventral pallium, and dorsal pallium.

It is important to note that manual annotation was not performed for every structure at every level of the ontology, which amounts to over 1,500 brain regions. Instead, the annotation strategy was meant to ensure that every “branch” of the ontological tree was covered by annotation. For example, for the four major parts of the brain: forebrain, midbrain, hindbrain, and spinal cord, a given gene may have expression in only diencephalon. Therefore, midbrain, hindbrain, and spinal cord would be annotated as “undetected”, and the expression in forebrain may be addressed by providing the actual expression information for diencephalon, while provided an “undetected” call for the sibling structure, secondary prosencephalon.

### Expression Levels and Patterns

The figures below illustrate the different expression metrics levels used to characterize the ISH data in the Allen Developing Mouse Brain Atlas.

Four types of expression **patterns** were identified (Figure 1):

- Undetected
- Full
- Regional
- Gradient

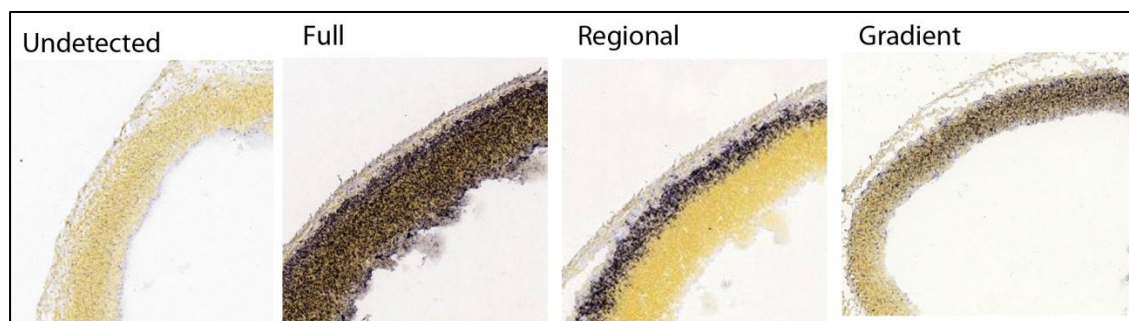


Figure 1. Four types of expression patterns used to characterize ISH data in the Allen Developing Mouse Brain Atlas.

Four different levels of expression **density** are defined (Figure 2):

- Undetected
- Low
- Medium
- High

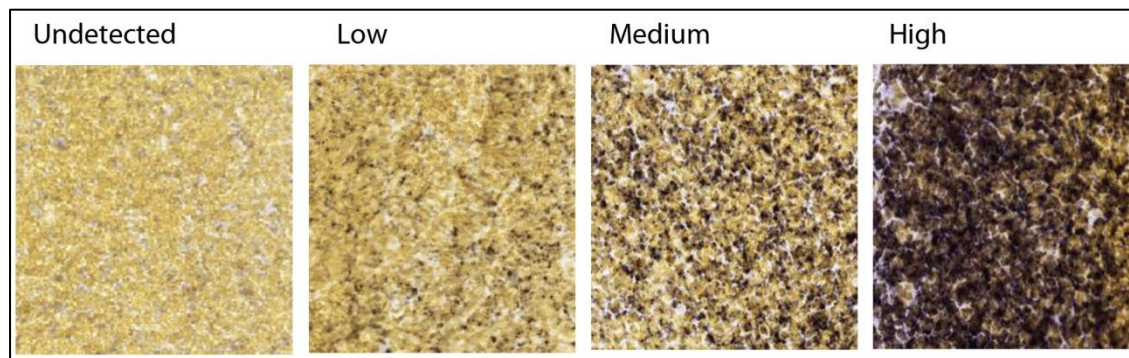


Figure 2. Four levels of expression density used to characterize ISH data in the Allen Developing Mouse Brain Atlas.

Four different levels of expression **intensity** are defined (Figure 3):

- Undetected
- Low
- Medium
- High

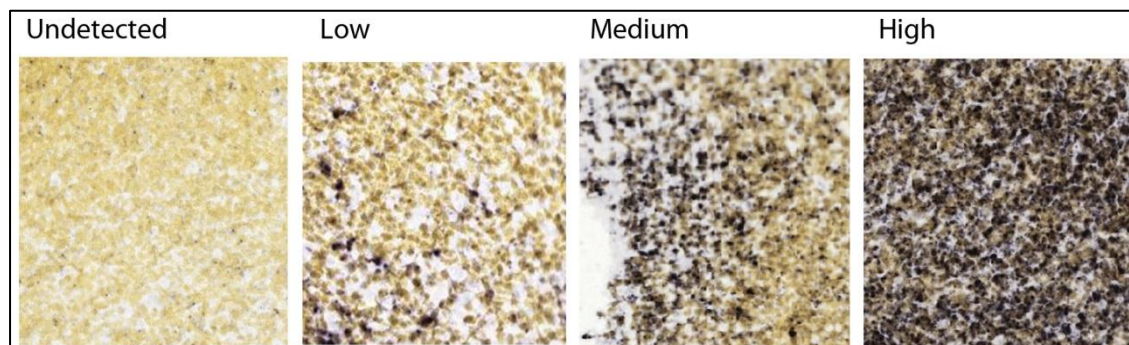
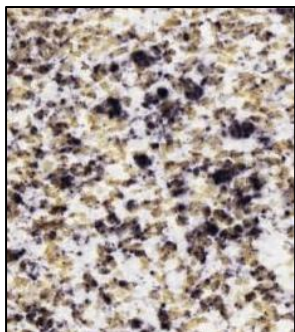


Figure 3. Four levels of expression intensity used to characterize ISH data in the Allen Developing Mouse Brain Atlas.

Additionally, when the annotated structure showed heterogeneous cell population (Figure 4), we annotated the cell population with the highest intensity value, and provided an indication that the annotation in this region applied to a “mixed population”.



**Figure 4. Example of a mixed population.**

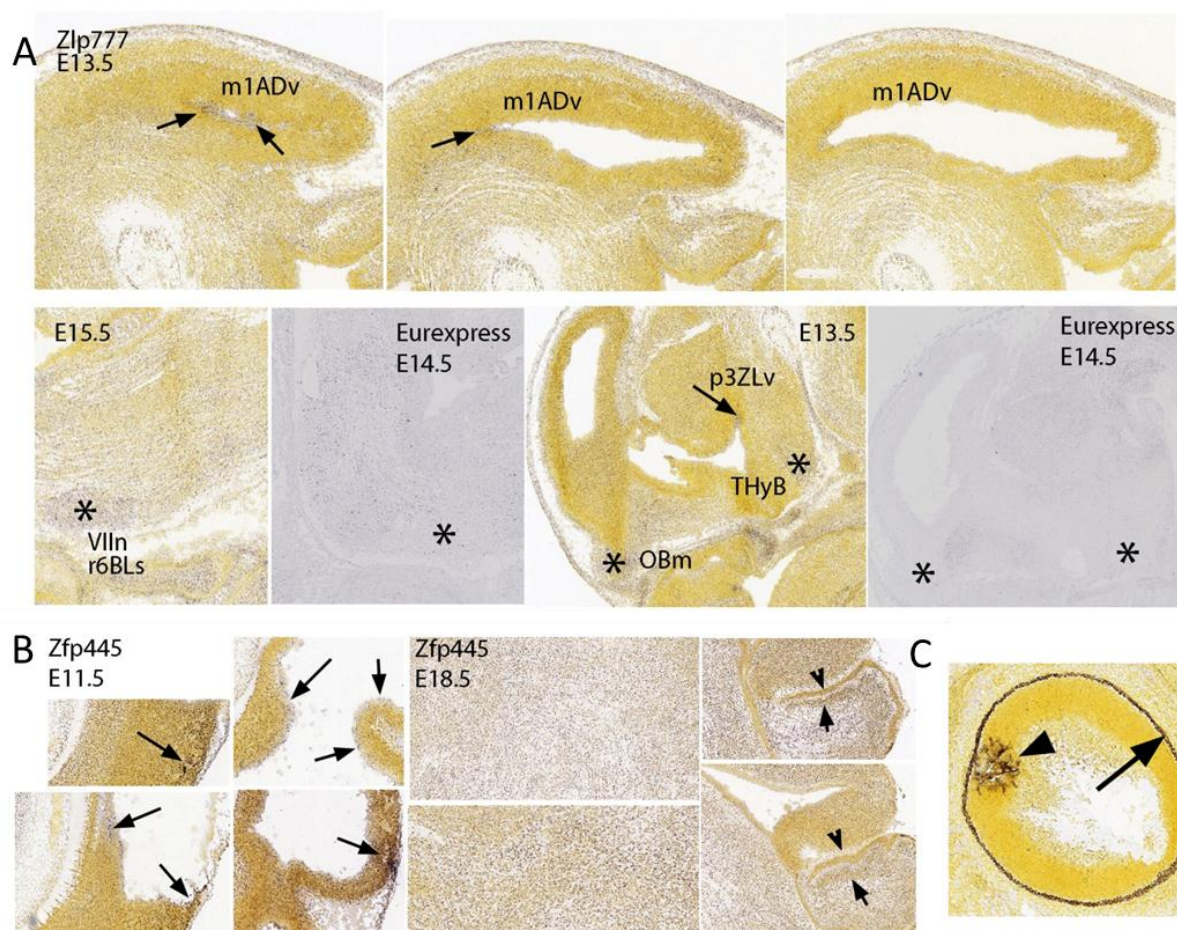
In some cases, weak signal was interpreted as background and scored as “undetected”. An example is shown below in Figure 5.

### **Quality Control of Annotated Image Series**

Weekly meetings of the annotators were organized to homogenize scoring criteria and to discuss individual image series with specific characteristics and complexity of expression patterns. Group agreement was used to finalize the annotation of these image series.

Manual quality control of anatomical identification was also performed by randomly opening and revising scores and annotation plan report. Approximately 10% of genes have been revised. When some discrepancy was detected the image series was re-annotated and discussed in the group agreement meeting.





**Figure 5. Examples of weak or artifactual staining.**

**A**, *Zfp777* shows a very weak labeling in which some staining accumulates in the ventricular corners (arrows), but is not expressed in other ventricular regions, and is considered an artifact. Labeling that also accumulates in some nuclear domains with high cellular density (e.g., Vln, OBm, and THyB) is also considered as an artifact; in these cases, comparison to EurExpress ISH data at E14.5 cannot confirm true ISH signal in these high density regions. Once artifactual staining has been identified in some areas, despite weak labeling in the striatum that corresponds to weak labeling in EurExpress data, the area is scored as “undetected” because it is difficult to distinguish between real and artifactual staining. **B**, *Zfp445* exhibits homogenous weak labeling in almost all cells (brain and organs). Slight differences can be attributed to cell density. At E11.5, some artifacts were clearly detectable suggesting a precipitate (arrows). At E18.5, some structures, like the external granular layer (arrowhead) or Purkinje cell layer (small arrows) of developing cerebellum were inconsistently labeled between sequential sections, suggesting non-specific labeling. **C**, The pigmentary epithelium of the retina typically exhibits non-specific labeling due to inherent properties of this tissue (arrow), and is not considered true signal. A labeling artifact is also visible over the anterior neural retina (arrowhead).

## PROCESSING OF ANNOTATION DATA FOR SEARCH

For the sake of efficiency, annotation was performed with the goal of covering every branch of the anatomic ontology tree. Therefore, annotation was not recorded directly for each of the >1,500 brain regions in the ontology, and different brain regions or levels of the ontology were annotated for different experiments. Therefore, the manually-entered annotation data required post-processing to support a website search for gene expression across the entire ontology. The manual annotation was “propagated” from a given annotated structure to fill in expression information for both child structures and for parent structures, based upon a simple logic (Figure 6). An example of the summary of the annotation for a gene at Level 5 is shown in Figure 7, with the directly annotated structures shown in Table 1.

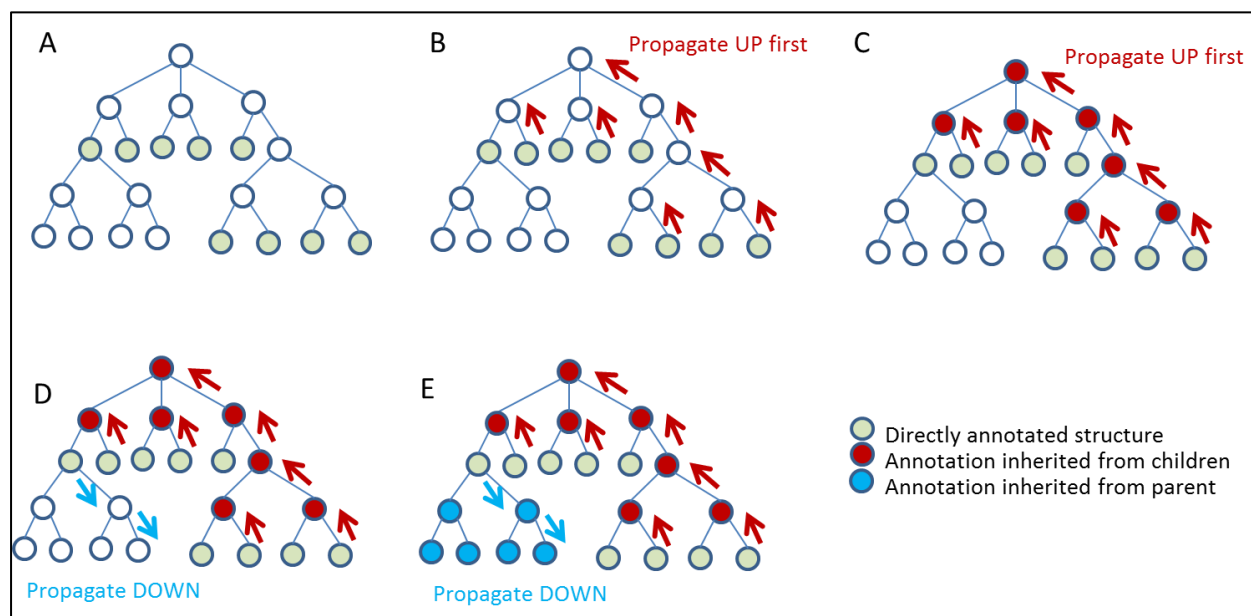
### Propagation of Annotation to Parent Structures

First, “child” structures were used to calculate the expression pattern of the parent. A first requirement was that all child structures must be annotated in order for the parent to inherit annotation. For level and density, the parent structure received the *highest* expression value of any of its children. The pattern metric was calculated as follows:

- If all children = full, then the parent = full.
- If all children = undetected, then the parent = undetected.
- If the children possess different pattern values, then the parent = regional.

### Propagation of Annotation to Child Structures

If a given structure was annotated, that expression data was assumed to apply to all child structures. If a parent has a pattern = full or undetected, the children inherit all annotation from that parent. If the parent has a pattern = regional or gradient, then the children inherit “can’t annotate”, because it is not possible to determine which child structures should receive the expression calls.



**Figure 6. Schematic illustrating propagation of annotation throughout anatomic ontology.**

This process is a step-wise operation in the order from A to E. In A, the manual annotation is filled in. In B and C, the values are calculated for parent structures based upon the annotation of their children, starting at the bottom of the tree and working all the way up to the top. In D and E, values are calculated for child structures based upon annotation of parents.

### Structures listed as “cannot annotate”

There are multiple reasons for a structure to be labeled “cannot annotate”:

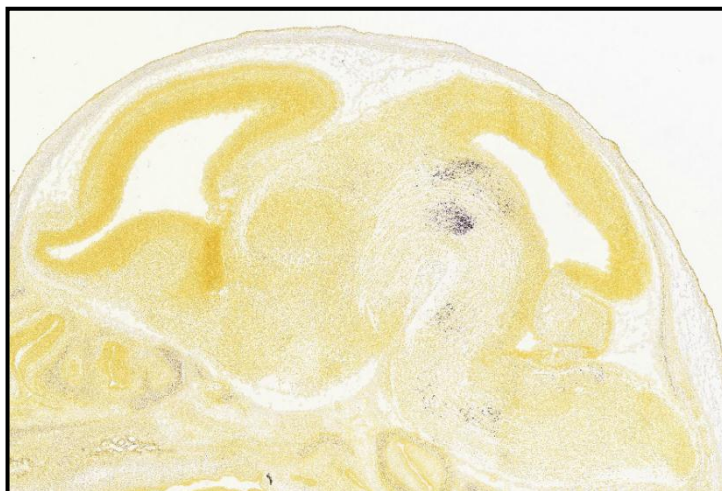
- A given structure was not available to assess in the ISH experiment. In some cases, this may be because a particular section of tissue containing the region was not available. For example, at E18.5, the brain was dissected, and thus eye vesicle structures are not available to annotate.
- The parent structure was labeled as “regional”. In this case, child structures cannot receive an annotation.

### Quality Control of Annotation

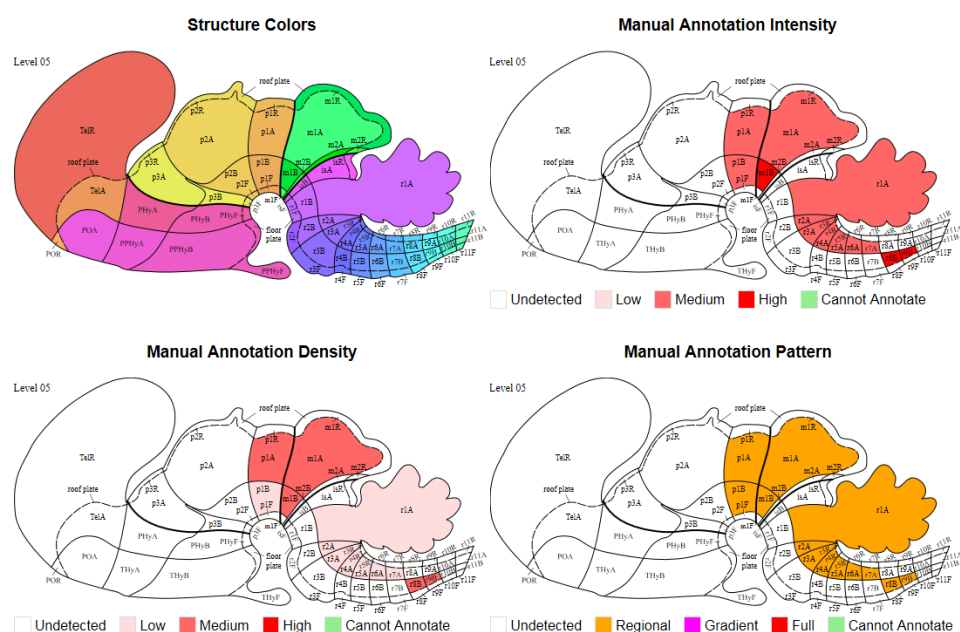
The annotation data was evaluated for different types of errors by querying the database, and these errors were then corrected by the annotation group. These errors included:

- 1) Inconsistent annotation within a given brain structure. This included instances where for a given brain region, there existed a combination of annotations that both indicated expression and lack of expression. For example, if an intensity call was listed as “high” but the pattern call was “undetected”, this is a contradiction.
- 2) Inconsistent annotation in the ontological tree. This included instances where a parent was annotated as “undetected” but a child structure was annotated as “detected”.
- 3) Missing annotation. Any branch of the ontological tree that was missing annotation was identified and corrected.

## ISH data for *C1ql2*



### Expression for *C1ql2* (E15.5)



**Figure 7. Example annotation for gene *C1ql2* at E15.5.**

Top, ISH image. Bottom, schematic of brain ontology showing the metrics annotated for each region for intensity, density, and pattern. Note that the annotation given in Table 1 is summarized here by Level 5 structures which include alar, basal, roof, and floor divisions of basic anteroposterior segments of the brain.

**Table 1. Annotation for Gene *C1q2* at E15.5.**

Region	Pattern	Density	Intensity	Mixed population
cerebellar hemisphere (CbH - cerebellar hemisphere)	Undetected	Undetected	Undetected	No
colliculus superior (SC - colliculus superior)	Undetected	Undetected	Undetected	No
commissural pretectal domain (CoP - commissural pretectal domain)	Regional	Low	Medium	No
dorsal part of m1A (m1AD - dorsal part of m1A)	Undetected	Undetected	Undetected	No
floor plate of m1 (m1F - floor plate of m1)	Undetected	Undetected	Undetected	No
floor plate of p1 (p1F - floor plate of p1)	Undetected	Undetected	Undetected	No
inferior colliculus (IC - inferior colliculus)	Undetected	Undetected	Undetected	No
intermediate stratum of PcPD (PcPDi - intermediate stratum of PcPD)	Regional	Medium	Medium	No
intermediate stratum of PcPL (PcPLi - intermediate stratum of PcPL)	Regional	Medium	Medium	No
intermediate stratum of PcPV (PcPVi - intermediate stratum of PcPV)	Regional	Medium	Medium	No
intermediate stratum of r8BI (r8Bli - intermediate stratum of r8BI)	Regional	Medium	High	No
intermediate stratum of r9BI (r9Bli - intermediate stratum of r9BI)	Regional	Medium	High	No
isthmus (is - isthmus)	Undetected	Undetected	Undetected	No
juxtacommissural pretectal domain (JcP - juxtacommissural pretectal domain)	Undetected	Undetected	Undetected	No
landmark tract (tracts - landmark tract)	Undetected	Undetected	Undetected	No
lateral part of r8B (r8BL - lateral part of r8B)	Undetected	Undetected	Undetected	No
lateral part of r9B (r9BL - lateral part of r9B)	Undetected	Undetected	Undetected	No
liminar part of alar r1 (r1Lim - liminar part of alar r1)	Regional	Low	Medium	No
liminar part of r2 alar plate (r2Lim - liminar part of r2 alar plate)	Regional	Low	Medium	No
mantle zone of m1BL (m1BLm - mantle zone of m1BL)	Regional	Medium	High	No
mantle zone of m1BM (m1BMm - mantle zone of m1BM)	Regional	Medium	High	No
mantle zone of m1Lim (m1Limm - mantle zone of m1Lim)	Regional	Medium	Medium	No
mantle zone of p1B (p1Bm - mantle zone of p1B)	Regional	Low	Medium	No
mantle zone of TG (TGm - mantle zone of TG)	Regional	Medium	Medium	No



Region	Pattern	Density	Intensity	Mixed population
medial part of r8B (r8BM - medial part of r8B)	Undetected	Undetected	Undetected	No
medial part of r9B (r9BM - medial part of r9B)	Undetected	Undetected	Undetected	No
mesomere 2 (preisthmus) (m2 - mesomere 2 (preisthmus))	Undetected	Undetected	Undetected	No
oculomotor nerve root (3N - oculomotor nerve root)	Undetected	Undetected	Undetected	No
periventricular stratum of PcPD (PcPDp - periventricular stratum of PcPD)	Undetected	Undetected	Undetected	No
periventricular stratum of PcPL (PcPLp - periventricular stratum of PcPL)	Undetected	Undetected	Undetected	No
periventricular stratum of PcPV (PcPVp - periventricular stratum of PcPV)	Undetected	Undetected	Undetected	No
periventricular stratum of r8BI (r8BIp - periventricular stratum of r8BI)	Undetected	Undetected	Undetected	No
periventricular stratum of r9BI (r9BIp - periventricular stratum of r9BI)	Undetected	Undetected	Undetected	No
prosomere 2 (posterior parencephalon) (p2 - prosomere 2 (posterior parencephalon))	Undetected	Undetected	Undetected	No
prosomere 3 (p3 - prosomere 3)	Undetected	Undetected	Undetected	No
r1 basal plate (r1B - r1 basal plate)	Undetected	Undetected	Undetected	No
r1 floor plate (r1F - r1 floor plate)	Undetected	Undetected	Undetected	No
r1 part of trigeminal sensory column (r1Tr - r1 part of trigeminal sensory column)	Undetected	Undetected	Undetected	No
r1 part of vestibular sensory column (r1Ve - r1 part of vestibular sensory column)	Undetected	Undetected	Undetected	No
r1 roof plate (r1R - r1 roof plate)	Undetected	Undetected	Undetected	No
r2 basal plate (r2B - r2 basal plate)	Undetected	Undetected	Undetected	No
r2 floor plate (r2F - r2 floor plate)	Undetected	Undetected	Undetected	No
r2 part of cochlear sensory column (r2Co - r2 part of cochlear sensory column)	Undetected	Undetected	Undetected	No
r2 part of trigeminal sensory column (r2Tr - r2 part of trigeminal sensory column)	Undetected	Undetected	Undetected	No
r2 part of vestibular sensory column (r2Ve - r2 part of vestibular sensory column)	Undetected	Undetected	Undetected	No
r2 roof plate (r2R - r2 roof plate)	Undetected	Undetected	Undetected	No
r3 alar plate (r3A - r3 alar plate)	Regional	Low	Medium	No
r3 basal plate (r3B - r3 basal plate)	Undetected	Undetected	Undetected	No
r3 floor plate (r3F - r3 floor plate)	Undetected	Undetected	Undetected	No

Region	Pattern	Density	Intensity	Mixed population
r3 roof plate (r3R - r3 roof plate)	Undetected	Undetected	Undetected	No
r4 alar plate (r4A - r4 alar plate)	Regional	Low	Medium	No
r4 basal plate (r4B - r4 basal plate)	Undetected	Undetected	Undetected	No
r4 floor plate (r4F - r4 floor plate)	Undetected	Undetected	Undetected	No
r4 roof plate (r4R - r4 roof plate)	Undetected	Undetected	Undetected	No
r5 alar plate (r5A - r5 alar plate)	Regional	Low	Medium	No
r5 basal plate (r5B - r5 basal plate)	Undetected	Undetected	Undetected	No
r5 floor plate (r5F - r5 floor plate)	Undetected	Undetected	Undetected	No
r5 roof plate (r5R - r5 roof plate)	Undetected	Undetected	Undetected	No
r6 alar plate (r6A - r6 alar plate)	Regional	Low	Medium	No
r6 basal plate (r6B - r6 basal plate)	Undetected	Undetected	Undetected	No
r6 floor plate (r6F - r6 floor plate)	Undetected	Undetected	Undetected	No
r6 roof plate (r6R - r6 roof plate)	Undetected	Undetected	Undetected	No
r7 alar plate (r7A - r7 alar plate)	Regional	Low	Medium	No
r7 basal plate (r7B - r7 basal plate)	Undetected	Undetected	Undetected	No
r7 floor plate (r7F - r7 floor plate)	Undetected	Undetected	Undetected	No
r7 roof plate (r7R - r7 roof plate)	Undetected	Undetected	Undetected	No
r8 alar plate (r8A - r8 alar plate)	Undetected	Undetected	Undetected	No
r8 floor plate (r8F - r8 floor plate)	Undetected	Undetected	Undetected	No
r8 roof plate (r8R - r8 roof plate)	Undetected	Undetected	Undetected	No
r9 alar plate (r9A - r9 alar plate)	Undetected	Undetected	Undetected	No
r9 floor plate (r9F - r9 floor plate)	Undetected	Undetected	Undetected	No
r9 roof plate (r9R - r9 roof plate)	Undetected	Undetected	Undetected	No
rhombomere 1 (r1 - rhombomere 1)	Not Yet Annotated	Not Yet Annotated	Not Yet Annotated	No
rhombomere 10 (r10 - rhombomere 10)	Undetected	Undetected	Undetected	No
rhombomere 11 (r11 - rhombomere 11)	Undetected	Undetected	Undetected	No
roof plate of m1 (m1R - roof plate of m1)	Undetected	Undetected	Undetected	No
roof plate of p1 (p1R - roof plate of p1)	Undetected	Undetected	Undetected	No
roof plate of p3 (p3R - roof plate of p3)	Not Yet Annotated	Not Yet Annotated	Not Yet Annotated	No
secondary prosencephalon (SP - secondary prosencephalon)	Undetected	Undetected	Undetected	No
spinal cord (SpC - spinal cord)	Undetected	Undetected	Undetected	No
superficial stratum of PCPD (PcPDs - superficial stratum of PCPD)	Full	Medium	Medium	No

Region	Pattern	Density	Intensity	Mixed population
superficial stratum of PcPL (PcPLs - superficial stratum of PcPL)	Full	Medium	Medium	No
superficial stratum of PcPV (PcPVs - superficial stratum of PcPV)	Full	Medium	Medium	No
superficial stratum of r8BI (r8BIs - superficial stratum of r8BI)	Undetected	Undetected	Undetected	No
superficial stratum of r9BI (r9BIs - superficial stratum of r9BI)	Undetected	Undetected	Undetected	No
ventricular zone of m1BL (m1BLv - ventricular zone of m1BL)	Undetected	Undetected	Undetected	No
ventricular zone of m1BM (m1BMv - ventricular zone of m1BM)	Undetected	Undetected	Undetected	No
ventricular zone of m1Lim (m1Limv - ventricular zone of m1Lim)	Undetected	Undetected	Undetected	No
ventricular zone of p1B (p1Bv - ventricular zone of p1B)	Undetected	Undetected	Undetected	No
ventricular zone of PcPD (PcPDv - ventricular zone of PcPD)	Undetected	Undetected	Undetected	No
ventricular zone of PcPL (PcPLv - ventricular zone of PcPL)	Undetected	Undetected	Undetected	No
ventricular zone of PcPV (PcPVv - ventricular zone of PcPV)	Undetected	Undetected	Undetected	No
ventricular zone of r8BI (r8BIv - ventricular zone of r8BI)	Undetected	Undetected	Undetected	No
ventricular zone of r9BI (r9BIv - ventricular zone of r9BI)	Undetected	Undetected	Undetected	No
ventricular zone of TG (TGv - ventricular zone of TG)	Undetected	Undetected	Undetected	No
ventrolateral part of alar p1 (p1AVL - ventrolateral part of alar p1)	Undetected	Undetected	Undetected	No