The landscape of regulatory genes in brain-wide neuronal phenotypes of a vertebrate brain

Hui Zhang,1,2,3 Haifang Wang1,3, Xiaoyu Shen1,3, Xinling Jia1, Shuguang Yu1,2, Xiaoying Qiu1, Yufan Wang1,2, Jiulin Du1*, Jun Yan1*, Jie He1*

1 State Key Laboratory of Neuroscience, Institute of Neuroscience, Shanghai Institutes for Biological Sciences, Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China
2 University of Chinese Academy of Sciences, Beijing, China
3 Equal contributions
*Correspondence: jiehe@ion.ac.cn (lead contact)

Abstract
Multidimensional landscapes of regulatory genes in neuronal phenotypes at whole-brain levels in the vertebrate remain elusive. We generated single-cell transcriptomes of ~67,000 region- and glutamatergic/neuromodulator-identifiable cells from larval zebrafish brains. Hierarchical clustering based on effector gene profiles (“terminal features”) distinguished major brain cell types. Sister clusters at hierarchical termini displayed similar terminal features. It was further verified by a population-level statistical method. Intriguingly, glutamatergic/GABAergic sister clusters mostly expressed distinct transcriptional factor (TF) profiles (“convergent
pattern”), whereas neuromodulator-type sister clusters predominantly expressed the same TF profiles (“matched pattern”). Interestingly, glutamatergic/GABAergic clusters with similar TF profiles could also display different terminal features (“divergent pattern”). It led us to identify a library of RNA-binding proteins that differentially marked divergent pair clusters, suggesting the post-transcriptional regulation of neuron diversification. Thus, our findings reveal multidimensional landscapes of transcriptional and post-transcriptional regulators in whole-brain neuronal phenotypes in the zebrafish brain.

### Introduction

The vertebrate brain harbors highly diverse neuronal types that are specifically interconnected to form functional circuits (Kepecs and Fishell 2014, Armananzas and Ascoli 2015, Moffitt, Bambah-Mukku et al. 2018). The brain comprises conserved major cell types (neurons, progenitors, glia, endothelial cells, etc.) (Marques, Zeisel et al. 2016, Saunders, Macosko et al. 2018, Hodge, Bakken et al. 2019). Previous studies have indicated that neuronal types could be well characterized by their features, including electrophysiological properties, neurotransmitter/modulator identity, synaptic connectivity, brain region identity, and cellular morphology (Zeisel, Munoz-Manchado et al. 2015, Pandey, Shekhar et al. 2018, Zeisel, Hochgerner et al. 2018). How these neuronal features are molecularly encoded at the whole-brain level is an important issue in neuroscience.
With the advent of single-cell RNA-sequencing (scRNA-seq) technology, recent studies have begun to characterize neuronal diversity by analyzing single-cell transcriptomes of large populations in the whole brain or specific brain regions of mice and humans (Darmanis, Sloan et al. 2015, Blue B. Lake, Rui Liu et al. 2016, Poulin, Tasic et al. 2016, Tasic, Menon et al. 2016, Chen, Wu et al. 2017, Tasic, Yao et al. 2018). These studies have provided compelling evidence that individual neuronal types could be well characterized by their transcriptomes (Poulin, Tasic et al. 2016, Tasic, Yao et al. 2018, Sugino, Clark et al. 2019). For instance, scRNA-seq analyses of half a million cells from the mouse brain showed that neurons could be classified by genes related to neuronal connectivity, synaptic transmission, and membrane conductance (Zeisel, Hochgerner et al. 2018). Besides, the scRNA-seq analysis of nearly 200 genetically marked mouse neuronal populations also showed that neurons could be classified by the expression level of various transcription factors (TFs), ion channels, synaptic proteins, and cell adhesion molecules (Sugino, Clark et al. 2019). These studies provided extensive information on the molecules that could be used to define neuronal types.

Generally, these molecules could be sub-divided into two primary categories: regulatory genes and effector genes. Regulatory genes encode proteins involved in gene transcription and translation (e.g., TFs and post-transcriptional regulators), while effector genes encode proteins serving specific neuronal terminal features (e.g., synaptic proteins, ion channels, transporters, and receptors) (Kratsios, Pinan-Lucarre

Remarkably, this phenotypic convergence has also been reported in the Drosophila’s optic lobe (Konstantinides, Kapuralin et al. 2018). However, the multidimensional landscapes of regulatory genes in vertebrate whole-brain neuronal phenotypes remain elusive.

The larval zebrafish brain comprises only about 100 thousand cells, thus providing an outstanding vertebrate model for studying cell diversity within the entire brain, using single-cell transcriptome analysis with full cell coverage by the 10x Genomics platform. Previous scRNA analysis of the zebrafish nervous system has elegantly demonstrated the temporal dynamics of brain cell development (Raj, Wagner et al. 2018, Raj, Farrell et al. 2020). In this study, we generated the multidimensional landscape of regulator genes in effector gene-based neuronal phenotypes (terminal features) at the whole-brain level by combining single-cell transcriptome data obtained from the whole brain, specific brain regions, as well as neurotransmitter- and neuromodulator-defined neuronal populations. We found that, at the transcriptional and post-transcriptional levels, glutamatergic/GABAergic neurons with the same
terminal features could express different TF profiles, while those with different terminal features could express the same TF but different RNA-binding protein profiles. In contrast, neuromodulator-type neurons that display particular terminal features expressed unique TF profiles. Thus, our findings reveal multidimensional landscapes of transcriptional and post-transcriptional regulators in the whole zebrafish brain.

**Results**

**Molecular Classification of Whole-brain Cells in Zebrafish**

To uncover the transcriptomic profiles of diverse cell types with regional identity in the larval zebrafish whole brain at single-cell resolution, we dissected and dissociated cells from the whole brain (n = 4), four specific brain regions (n = 2 each), including the forebrain (Fore), optic tectum (OT), hindbrain (Hind), and the region underneath the optic tectum (sub-OT) in the 8-days post fertilization (dpf) zebrafish (Figure 1-figure supplement 1A). We performed scRNAseq of these cells using the 10x Genomics Chromium 3’ v2 platform. The libraries were sequenced to a mean depth of 126,651 reads per library, with a median of 1,891 UMI and 866 genes per cell (Figure 1-Source Data 1). Reproducibility of transcription analysis was shown by the finding that replicates of whole-brain samples and individual brain regions were primarily overlapped in the t-distributed stochastic neighbor embedding (t-SNE) plots (Figure 1-figure supplement 1B-C). We obtained the transcriptomes of 65,253 cells and selected 45,746 cells for further analysis after filtering out low-quality data (Materials
and methods). Then aggregated all cells into 68 clusters using high-variance genes (n = 1,402) in t-SNE plot (Figure 1A). The Jaccard index-based analysis showed that most clusters had a high Jaccard index greater than 0.6, indicating the robustness of these clusters (Details in Materials and methods) (Tang, Kaymaz et al. 2020) except for a few neuronal clusters from sub-OT (cluster 52, 66), and non-neuronal clusters (neuroprogenitors: cluster 37, 56; radial astrocytes: 65, Figure 1-figure supplement 1E).

Each cluster was annotated according to cell type-specific marker genes from the literature or ZFIN (Zebrafish Information Network) database (Figure 1-Source Data 2). To assign the brain region identity for each cluster, cells from each of the four specific regions (Fore, OT, sub-OT, Hind) were found to cover multiple but non-overlapping clusters and all 68 clusters could be assigned with their brain region origins (Figure 1-figure supplement 2A-B, and Figure 1-Source Data 3). Furthermore, to identify potential brain-region-specific markers that exist in all cell types, we identified all genes that were differentially expressed in each region for all major cell types (vglut+, glutamatergic neurons; gad+, inhibitory neurons; pcna+, neuroprogenitors; cx43+, radial astrocytes). The differentially expressed genes shared by all cell types were considered to represent the targeted region-specific markers (Figure 1C). We indeed identified a small set of genes specific to each brain region independent of cell type. For instance, foxg1a, en2a, and hoxb3a were explicitly expressed in all cell types of the Fore, OT, and Hind, respectively (Figure 1D).
Interestingly, these brain-region-specific genes also exhibited a conserved region-specific expression pattern in the mouse brain (Mark Hanks 1995, Miguel Manzanares 2001, Kumamoto and Hanashima 2017). We found no specific gene for the sub-OT, probably due to the diverse brain structures in this region. These region-specific genes, which may be involved in forming regional identity during brain development, could be used to study region-specific neuronal connectivity and function.

To assign the neurotransmitter/modulator identity for each cluster, we used the marker genes that were specific to primary neurotransmitter/modulator phenotypes, including slc17a6b (glutamatergic), gad1b (GABAergic), slc6a5 (glycinergic), th (dopaminergic, DA), tph2 (serotonergic, 5-HT), and chata (cholinergic, ChAT) (Figure 1-figure supplement 2C, Figure 1-Source Data 3). The ratio of glutamatergic to GABAergic neurons was the highest in the forebrain and lowest in the hindbrain, indicating that glutamatergic neurons predominantly belonged to the forebrain, whereas glycinergic neurons mainly resided in the hindbrain (Figure 1-figure supplement 2D). These regional patterns of neurons expressing different neurotransmitter types were validated using the transgenic fishlines: Tg (vglut2a:loxp-DsRed-loxp-GFP), Tg (gad1b:EGFP), and Tg (glyT2:GFP), each exhibiting distinct labeling of glutamatergic, GABAergic, and glycinergic neurons, respectively (Figure 1-figure supplement 2E).

Moreover, the Lawson-Hanson algorithm for non-negative least squares (NNLS)
analysis using cluster-specific marker genes (top 20, Figure 1-Source Data 4) showed that these 68 clusters exhibited a high overlap with their counterparts in the juvenile zebrafish brain recently reported (Raj, Wagner et al. 2018). Meanwhile, clusters with different regional origins or cell types also exhibited a high correlation with their counterparts in the juvenile zebrafish (Figure 1E and Figure 1-figure supplement 2F). Thus, our analysis indicated that the brain at 8 dpf mostly represented cellular diversity in the juvenile brain.

Furthermore, Gene Ontology (GO) analysis of 1,402 variable genes used for the classification of all 68 clusters showed that the majority of these genes (78.4%, n=1,099) were effector genes, which could be classified as neuropeptides, receptors, transporters, ion channels, synaptic proteins, and cell adhesion molecules (Figure 1-figure supplement 2G). This result suggests the importance of effector gene profiles in brain cell classification, which has also been appreciated by previous studies in different species (Anirban Paul 2017, Hodge, Bakken et al. 2019).

We thus generated the hierarchical classification of all 68 cell clusters using the profiles of these 1,099 effector genes. All 68 clusters were first segregated into two groups, neuronal cells (48 clusters, 37,880 cells) and non-neuronal cells (20 clusters, 7,866 cells) (Figure 1F). Among non-neuronal cells, oligodendrocytes (Clusters 40, 42; olig2+, sox10+), microglia (Cluster 55; apoeb+, mpeg1.1+), endothelial cells (Clusters 48, 57; fxyd1+, rbp4+), erythrocytes (Cluster 67; hbbe1.2+, hbae5+), radial astrocytes...
(Clusters 21, 46, 65; cx43*, glua*), and neuroprogenitors (Clusters 8, 14, 18, 28, 29, 33, 34, 37, 53, 56; penta*, cdk1*) were identified according to putative marker genes (Figure 1-Source Data 5). On the other hand, among neuronal cells, the first segregation defined three classes of neurons (branch I, cerebellum and habenula; branch IIa, glutamatergic neurons, branch IIb, inhibitory neurons). Branch I consisted of granule cells (Clusters 19), torus longitudinals (Clusters 45), cranial ganglions (Clusters 19, 50, 63), dorsal and ventral habenula neurons (Cluster 35 and Cluster 59). Branch IIa included 22 subclasses of excitatory glutamatergic neurons (vgluta2a*-gad1b*-glyt2*), whereas Branch IIb inhibitory neurons included 17 subclasses of GABAergic neurons (vgluta2a'/gad1b+) and 1 subclass of glycinergic neurons (vgluta2a'-gad1b*-glyt2*, Cluster17, Figure 1-Source Data 5).

**Molecular Classification of Neuromodulator-type Neurons**

To further examine the expression of neuromodulators at the whole-brain level, we performed the scRNA-seq analysis of neuromodulator neurons sorted from the whole brain of Tg (ETvmat2:GFP) transgenic fish (Figure 2A). Using this fishline, we could examine transcriptomes of dopaminergic (DA) neurons, serotoninergic (5-hydroxytryptamine, 5-HT) neurons, and norepinephrinergic (NE) neurons (Wen, Wei et al. 2008). After the filtering procedures described above, we obtained a total of 5,368 vmat2-expressing cells (Materials and methods). The analysis aggregated these cells into 22 clusters in t-SNE plots (Figure 2B), more than those found using whole-brain samples (2 dopaminergic, 1 serotonergic, and 3 cholinergic; Figure
To further validate the stability of clusters, we used the Jaccard index, which showed that the majority of clusters (20/22) were stable using mean/median Jaccard index > 0.6 as cutoff (Figure 2-figure supplement 1A).

According to region-specific marker genes identified above and known marker genes for neuromodulator neurons, we assigned 7 clusters (Fore: Clusters 18-19; sub-OT: Clusters 16, 20; OT: Cluster 22; and Clusters 17, 21 without specific regional identity) as DA neurons, 14 clusters (Hind: Cluster 10; sub-OT: Clusters 1, 2, 6, 7; and Clusters 3-5, 8-9, 11-14 without regional identity) as 5-HT neurons, one cluster (Hind: Cluster 15) as NE neurons (Figure 2-figure supplement 1C-D, Figure 2-Source Data 1).

Further examination of these neuromodulator clusters for their expression of specific neurotransmitters showed that the majority of serotoninergic (8/14) and dopaminergic clusters (5/7) expressed GABAergic markers gad1b/gad2 (Figure 2C). This result was further validated by Tg (ETvmat2:GFP::gad1b:gal4::uas:mCherry) (Figure 2-figure supplement 1E). Only three serotoninergic clusters (Clusters 11-13) expressed glutamatergic marker vglut3, two dopaminergic clusters (Clusters 16, 17) and one norepinephrinergic cluster (Cluster 15) expressed glutamatergic markers, vglut2a and vglut2b (Figure 2C). These results are consistent with previous studies (Filippi, Mueller et al. 2014). Besides, we also found 3 clusters (Clusters 24, 41, and 64) in whole brain showed choline and glutamate preferential co-expression (Figure 1-Source Data 2). Thus, our analysis provided a whole-brain characterization of the
co-expression patterns of neurotransmitters and neuromodulators. In sum, dopaminergic and serotonergic neurons preferentially expressed GABAergic markers, whereas norepinephrinergic and cholinergic neurons mostly expressed glutamatergic markers.

The TF Regulatory Landscape of Whole-brain Glutamatergic/GABAergic Neuron Clusters

In the hierarchical classification based on effector gene profiles (Figure 1F), glutamatergic/GABAergic clusters (n = 39) at the terminus pairs represented the ones with the most similar terminal features, termed “sister clusters”. In addition, the certainty of this hierarchical classification was verified by the bootstrap re-sampling analysis using pvclust v.2.0 (Figure 3-figure supplement 1A) (Suzuki and Shimodaira 2006). We identified 11 pairs of sister clusters, neurons in each pair exhibited the same neurotransmitter types (Figure 3-figure supplement 2A). To our surprise, neurons of each sister clusters could be from either the same (n = 6) or different (n = 5) brain regions (Figure 3-figure supplement 2A), which did not reflect the strong brain region preference.

To further examine the TF profiles of these effector gene-based sister clusters, we classified glutamatergic (IIa) and inhibitory (IIb) neurotransmitter-type neurons using TF profiles. Notably, TF-based and effector gene-based trees were distinct in terms of matching node (only one matching node: Clusters 9/61) and tree distances (tree...
distance = 0.71, Figure 3-figure supplement 1B), suggesting that the effector gene-based sister clusters (Figure 3-figure supplement 2A) might express different TF profiles.

We found out of 11 effector genes-based sister clusters, only 1 pairs could be found in TF-based sister clusters (Clusters 9/61, “matched pattern”, Figure 3-figure supplement 2B). And other 10 effector gene-based sister clusters were separated in TF-based classification, suggesting that neurons with similar terminal features mostly expressed different TF profiles (“convergent pattern”; Figure 3-figure supplement 2C). Also, neurons in each of these 10 sister clusters could come from either the same (n = 5) or different (n = 5) brain regions, exhibiting no brain region preference (Figure 3-figure supplement 2C).

Alternatively, we performed the population-level statistical analysis to compare the landscape of TF and effector gene expression accounting for the full spectrum of cell types rather than just the most similar sister clusters. For all glutamatergic/GABAergic neuronal clusters (n=39), we calculated the distances between every two clusters \( C_{39}^2 \) based on either effector gene profiles or TF profiles, and then defined the pairs, which had the lowest 10% distances after ranking, as similar pair clusters (Figure 3A). Then, we defined paired clusters that were similar in both TF and effector gene profiles as “matched pattern”, those paired clusters that were similar in effector gene profiles but not in TF profiles as “convergent pattern”
The population-level analysis identified 19 pairs of effector gene-based similar pair clusters, 5 with matched pattern and 14 with convergent pattern (Figure 3B). Overall, similar pair clusters with either matched or convergent pattern identified by the population-level analysis showed an overlapping but distinct pattern with those identified by the hierarchical sister cluster analysis (Figure 3B). This discrepancy was likely due to the following facts: 1. In the population-level statistical analysis, we arbitrarily set the lowest threshold as a criterion to identify similar pair clusters. Thus, the levels of this threshold could influence the production of similar pair clusters. 2. In population-level statistical analysis, each cluster could use for multiple times, whereas in hierarchical sister cluster analysis, once a cluster was selected as a pair with another cluster, it could not be re-used again. To overcome this discrepancy, we intersected the results from hierarchical sister cluster analysis and population-level statistical analysis, and identified 8 pairs of effector gene-based glutamatergic/GABAergic similar pair clusters, 1 with matched pattern and 7 with convergent pattern (Figure 3B).

We further validated these patterns of paired neuron clusters by subsampling of genes (80% of total either TFs or effector genes) and average statistics over 20 times to re-identify similar paired clusters based on either TFs or effector genes (Materials and methods). Notably, re-identified paired clusters of different patterns completely recapitulated those pairs identified using the population-level statistical analysis above (Figure 3-supplemental 4E), indicating the robustness of these patterns.
By combining sister cluster analysis and the population-level statistical analysis, we identified 1 paired clusters of “matched pattern”, 7 paired clusters of “convergent pattern” (Figure 3B). Here were some representative cases with the convergent pattern. The first case was a glutamatergic pair cluster from different brain regions: tectal glutamatergic cluster 1 and hindbrain glutamatergic cluster 31 shared effector gene profiles including camk2n1a/stmn4/cbln2b/olfm3a/cd63, but differentially expressed TF profiles, atf5b/bhlhe41/lhx1a and ddit3/cebpbl/lef1, respectively (Figure 3C). The second case was GABAergic pair cluster from different brain regions: GABAergic clusters in the forebrain (cluster 36) and the sub-OT (cluster 16) shared effector gene profiles mcl1a/cbln1/tspan18a/dusp5/ncaldb, but differentially expressed TF profiles, tbr1b/foxg1albarhl2 and otpa/otpb/six6a, respectively (Figure 3D). Consistently, each of the above TF profiles has previously been characterized to participate in the specification of either glutamatergic or GABAergic neurons (Li, Yun et al. 2007, Kala, Haugas et al. 2009, Talbot, Johnson et al. 2010, Waite, Skidmore et al. 2011, Achim, Peltopuro et al. 2013).

**The TF Regulatory Landscape of Neuromodulatory Neuron Clusters**

We further examined neuromodulator-type sister clusters based on either TF or effector gene profiles (Figure 3-figure supplement 3A). Using hierarchical sister cluster analysis, 5/8 of neuromodulator-type sister clusters matched with those defined by TF profiles (“matched pattern”), 3/8 sister cluster were found to be
separated in TF-based classification (“convergent pattern”, Figure 3-figure supplement 3C-D).

Then, we also performed the population-level statistical analysis using all genes or 80% genes (subsampling) to compare the landscape of TF and effector gene expression. We found 8 pairs with matched pattern and 1 pair with convergent pattern (Figure 3-figure supplementary 3E-F). Thus, we intersected hierarchical sister cluster analysis and population-level statistical analysis, and identified 6 pairs of similar neuromodulator pair clusters, 5 with matched pattern and 1 with convergent pattern (Figure 3-figure supplementary 3G-I).

Together with the above results of neuromodulator-type and neurotransmitter types, our analysis showed that neuromodulator pairs with similar effector gene profiles predominantly were “matched” pattern, whereas neurotransmitter pairs with similar effector gene profiles mainly were “convergent” pattern (Figure 3E). Moreover, we performed global tree measurement using the R package “TreeDist”, and found that the distance between TF-based and effector-based hierarchical tree of neuromodulator neurons (Tree distance = 0.38, Figure 3-figure supplementary 3A-B) was lower than that of glutamatergic/GABAergic neurons (Tree distance = 0.71, Figure 3-figure supplementary 1B), suggesting distinct TF regulatory logic of effector-based phenotypes between neurotransmitter-type and neuromodulator-type at the global level. The smaller distance for neuromodulator neurons was consistent with our
conclusion that neuromodulator type clusters predominantly expressed the same TF profiles ("matched"). The “matched” pattern suggested the generation of similar neuromodulator pairs shared specific TF programs (“stereotyped programming”), and the “convergent” pattern suggested the generation of similar neurotransmitter pairs could be generated by different TF programs (“flexible programming”). These different strategies may account for the fact that across species, neuromodulator types are more conserved, whereas neurotransmitter types are much diverse and variable (La Manno, Gyllborg et al. 2016, Saunders, Macosko et al. 2018, Tiklova, Bjorklund et al. 2019, Poulin, Gaertner et al. 2020).

**Different Neuron Clusters Exhibit Similar TF Profiles**

On the other hand, in glutamatergic/GABAergic neuronal classification, we surprisingly found that 13 pairs of sister clusters with similar TF profiles were separated in the effector gene-based classification. In other words, different from matched and convergent pairs mentioned above (Figure 3B-D), these 13 glutamatergic/GABAergic clusters exhibited different effector gene profiles but similar TFs, here terms as “divergent” pattern (Figure 3-figure supplement 4A). Also, the population-level statistical analysis identified divergent pairs (n = 15, Figure 3-figure supplement 4B) that were largely overlapped with those identified by hierarchical sister analysis (n = 10, Figure 3B).

Neurons in each of these 10 divergent paired clusters could be from the same (n=6) or
different (n=4) brain regions (Figure 3-figure supplement 4A). More strikingly, 2 pairs of neuronal clusters with divergent pattern expressed different neurotransmitters. For examples, sub-tectal glutamatergic neurons (cluster 24) and hindbrain GABAergic neurons (cluster 17) shared similar TF profiles \textit{id2a/pax2a/zfhx4/meis1b}, but expressed different effector gene profiles, \textit{cpne4a/c1ql4b/slc5a7a} and \textit{gad1b/slc6a1b/slc6a5}, respectively (Figure 3F); forebrain glutamatergic neurons (cluster 30) and tectal GABAergic neurons (cluster 11) shared TF profiles, \textit{bhlhe41/bcl11ba/mef2cb}, but had differentially expressed effector gene profiles, \textit{adcyap1b/syt5b/rpml/gadd45bb}, and \textit{kcnip1b/cxcl4/plxdc1}, respectively (Figure 3G). This result illustrated that neurons with different terminal features could also express the same TF profiles. Together, our analysis demonstrated the landscape of TFs in whole brain-wide neuronal phenotypes in the larval zebrafish brain.

\textbf{Combinatorial TFs in Marking Tectal Morphological Subclasses}

Neuronal morphology and effector gene profile are two critical criteria of neuron diversity classification (Sugino, Clark et al. 2019, Peng, Xie et al. 2021). Also, many previous studies have provided the apparent links between effector genes and neuron morphology (Kristin L. Whitford 2002, Marcette, Chen et al. 2014, Delandre, Amikura et al. 2016, Noblett, Wu et al. 2019, Peng, Xie et al. 2021). Thus, after we found three patterns (“matched”, “convergent”, and “divergent”), we then wondered if similar patterns present between morphological subclasses and TFs.
To experimentally verify the relationship between TFs and neuronal phenotypes, we focused on morphological subtypes of tectal glutamatergic neurons. Higher-coverage transcriptomes of tectal glutamatergic neurons were further achieved by scRNAseq of sorted tectal glutamatergic neurons in 8-dpf using Tg (*vglut2a*::loxp-DsRed-loxp-GFP) fishline (Figure 4A). Canonical correlation analysis (CCA) of cells from two independent experiments indicated the reproducibility of the analysis (Figure 4-figure supplement 1A). Analysis of 3,883 sorted tectal glutamatergic neurons yielded 11 clusters (n = 11, Figure 4A), more than those identified from the whole-brain samples (n = 5, Figure 4-figure supplement 1B). The majority of clusters (10/11) were stable using mean/median Jaccard index > 0.6 as cutoff (Figure 4-figure supplement 1D). Each cluster of tectal glutamatergic neurons was identified by TF marker genes (Figure 4-figure supplement 1E-F, Figure 4-Source Data 1).

We developed a labeling strategy using three different plasmids to analyze morphological subclasses of tectal glutamatergic neurons expressing identified TFs. First, we placed Gal4FF cassette at the start-codon of each TF by bacterial artificial chromosome (BAC) recombination technique (Suster, Abe et al. 2011) and constructed the Gal4FF plasmids for 15 marker TF genes (Figure 4B, Materials and methods). Second, we generated *vglut2a*::CRE BAC plasmid using a similar method. Third, plasmid *uas*:loxP-stop-loxP-tdTomatoCaax was used with the first two plasmids to intersectionally label single tectal glutamatergic neurons, each expressing one specific TF (Figure 4C). These labeled neurons were then used for morphological
analysis. We labeled and defined morphological subclasses of tectal neurons by injecting all three plasmids into zebrafish embryos at 4 to 16-cells stage. We observed the enormous variability in the morphology of tectal glutamatergic neurons in their more refined structures. By the limited number of neurons analyzed (n=574), we were unlikely to define morphological subclasses using a global morphological description. Instead, we used the criterion based on major morphological features, including stratification, soma position, and projection patterns, to define the morphological subclasses in the current study. In addition, similar morphological classification have also been used previously for tectal neurons (Linda M Nevin 2010, Robles, Smith et al. 2011, DeMarco, Xu et al. 2020). We selected 6 TF marker genes (en2b, foxb1a, zic1, bhlhe22, zbtb18, and irx1a) for further analysis based on two criteria: 1. These TFs are highly expressed and specific to individual tectal glutamatergic clusters based on single-cell RNAseq analysis; 2. Their BAC plasmids could reliably mark particular morphological subclasses (at least in four animals).

Furthermore, using confocal imaging, we reconstructed a total of 574 tectal neurons (from 263 zebrafish, Figure 4-Source Data 2) that expressed one of these 6 TFs (Figure 4-figure supplement 1G). These neurons were categorized into 7 morphological subclasses: bi-stratified (I, n = 36), mono-stratified (II, n = 71), non-stratified (III, n = 318), necklace-like (IV, n = 76), cross-hemispheric (V, n = 15), ascending (VI, n = 18), and descending (VII, n = 40) (Figure 4D).
Whether a specific TF could serve as a marker for a particular morphological subclass was determined by the criterion that the TF appeared in a given subclass at least 4 times (from at least 4 fish). Remarkably, all 6 TFs marked multiple morphological subclasses in a combinatorial manner, while each of 7 morphological subclasses was marked by numerous TFs (Figure 4E). For example, *zic1* highly marked non-stratified and bi-stratified subclasses, and all 6 TFs marked the non-stratified subclass.

In summary, we found that single TF could mark multiple morphological subtypes in the optic tectum, and multiple TFs could mark a single morphological subtype. This observation could be inferred from “convergent pattern” (same TFs were expressed in different effector-based subtypes) and “divergent pattern” (different TFs were expressed in similar effector-based subtypes), respectively. However, we could not exclude unknown indirect regulations of morphological subtypes by TFs.

**The Landscape of Post-transcriptional Regulators in Neuronal Clusters with Different Terminal Features but Similar TF Profiles**

The above finding that neuronal clusters with different terminal features expressed similar TF profiles were likely resulted from non-TF determinants (Figure 3F-G). We then performed GO analysis of differentially expressed genes between sister clusters of 10 pairs with different terminal features but similar TF profiles (“divergent pattern”). We surprisingly found that many differentially expressed genes encoded RNA-binding proteins (RBPs, Figure 5A). We thus yielded a library of RBP-encoded
genes that could differentially mark neurons with different terminal features (Figure 5-Source Data 1). Here were two examples: two forebrain glutamatergic neuronal clusters (Cluster 7 and Cluster 38) shared the similar expression pattern of TF profiles eomesal/neurod1/tbr1b/foxg1a, but had different effector gene profiles and RBPs, rprmal/stxbp1/blegf6/khdrbs2/msi2b/rbm5 and nrgnalsusp5/fam43b/qkia/larp6a, respectively; GABAergic neurons in hindbrain (cluster 12) and sub-tectal (cluster 16) shared TF profiles, otpal/bhilhe41/id2a, but differentially expressed effector genes and RBPs, kctd4/rgs5b/cpne2/eif5a2/celf3a and slc6a1b/cplx2l/nrdg2/msi2b/rbfox1/msi2a, respectively (Figure 5B). These RBP-encoded genes exhibited wide RNA functions, including capping, splicing, polyadenylation, transport, stability, and translation (Figure 5C, Figure 5-Source Data 3).

Besides, our analysis also yielded a comprehensive list of genes encoding RBPs marking other patterns (Figure 5-figure supplement 1A-D, Figure 5-Source Data 1-2). We found that some RBPs exhibited differential expression between paired clusters specific to one of the above three patterns (Figure 5-figure supplement 1E-F). Also, we found that the number of RBPs that were specific to divergent pattern (n=52 from 10 pairs) was much more than those specific to either matched pattern (n=7 from 1 pair) or convergent pattern (n=10 from 7 pairs, Figure 5-figure supplementary 1E-F). Thus, in terms of the number of pattern-specific RBPs per pair, RBPs showed the preferential expressions in divergent pairs compared to convergent ones (Figure 5D). Note that since there was only one pair of matched pattern, we did not include it in
this analysis. These pattern-specific RBPs were known to be involved in a wide range of RNA functions, including capping, splicing, polyadenylation, transport, stability, and translation (Figure 5-figure supplement 1F, Figure 5-Source Data 3). Thus, in addition to TFs, post-transcriptional regulatory factors also play significant roles in determining transcriptome profiles of neuronal classification.

Furthermore, we examined effector gene that were targeted by pattern-specific RBPs using the oRNAment database (http://rnabiology.ircm.qc.ca/oRNAment/) to explore a potential causality between the divergence of RBPs and the divergence of effector genes (Benoit Bouvrette, Bovaird et al. 2020). GO analysis showed that pattern-specific RBPs targeted various molecular categories including effector genes, TFs, mRNA processing, metabolism (Figure 5-figure supplement 1H-J, Figure 5-Source Data 4). More importantly, divergent pattern-specific RBPs targeted significantly higher proportions of effector genes than matched and convergent pattern-specific RBPs (Figure 5E). These results suggested the potential causality between RBPs and effector gene profiles in divergent pairs.

Interestingly, genes encoding well-known sequence-specific RNA binding protein Rbfox1-encoded gene was differentially expressed in multiple neuronal clusters, like 23/25, 27/54, 11/30, 13/24, 12/16, 2/3, and 7/38 (Figure 5-Source Data 5). Rbfox1 specifically recognizes UGCAUG motifs, which are often found at 5′- and 3′-regions of introns (Yui Jin 2003, Vuong, Wei et al. 2018). Previous studies have elegantly
revealed the importance of activity-dependent splicing regulator Rbfox1 in the
transition from neuroprogenitors to neurons as well as in the interneuron
subtype-specific splicing in the mouse cortex (Zhang, Chen et al. 2016). Moreover,
we calculated the frequency of individual RBPs in marking all sister pair clusters with
each of three patterns and found that single RBPs were frequently present in multiple
pair clusters (42.5%, 51/120, Figure 5F), indicating that clusters mostly used RBPs in
a combinatorial manner.

Thus, our analysis revealed the importance of RBP-encoded genes in specifying
diverse neuronal phenotypes, and the identification of a comprehensive list of
RBP-encoded genes in our study provides a valuable resource for future investigation.

Discussion

This study analyzed ~65,000 single cells from the whole brain, specific brain regions,
neurotransmitter/neuromodulator-defined neuronal populations of the 8-dpf larval
zebrafish (Figure 1A, Figure 2B, and Figure 4A). Notably, our transcriptome analysis
offered a close-to-full coverage of all cells (~100,000) in the zebrafish brain,
providing multidimensional landscapes of TFs and post-transcriptional regulators in
vertebrate whole-brain neuron classification (Figure 6).

A significant focus of our analysis is the analysis of sister clusters at the termini of
effector gene-based hierarchical classification (Figure 3A). These sister pair clusters
represent the finest molecularly defined neuronal subclasses based on effector gene
profiles. Because sister clusters of the terminus pair with highly similar terminal
features, it offers us an opportunity to study the determinants of their terminal feature
similarity and factors that lead to their slight diversification. Our analysis of the
expression profiles of regulatory genes in sister clusters, including TFs and
post-transcriptional regulators (Figure 6) may thus help to elucidate the molecular
logic underlying neuronal diversification.

Specifically, our analysis identified three patterns of TFs in specifying neuronal
phenotypes, “matched pattern”, “convergent pattern”, and “divergent pattern” (Figure
3B and Figure 6). Glutamatergic/GABAergic sister clusters with similar terminal
features mostly expressed distinct TF profiles (“convergent pattern”), whereas
neuromodulator-type sister clusters largely expressed the same TF profiles (“matched
pattern”, Figure 3E). The results indicated that the “matched” pattern supports the
notion that the same set of TFs plays a crucial role in determining a given profile of
neuromodulator-type terminal features, whereas the “convergent” pattern suggests
that even though TFs are important for subclass determination, different TFs could
still converge onto the same transcriptome phenotype. Combined with earlier findings
of extensive phenotypic convergence of distinct neuron types in worms and flies
2019), our findings in zebrafish strongly suggest that phenotypic convergence is
highly conserved from invertebrates to vertebrates.
More interestingly, we identified 10 terminus clusters exhibiting different transcriptome patterns but similar TF profiles (“divergent pattern”). Thus, these clusters are likely determined by factors other than TFs (Figure 3F-G). Our further analyses revealed the potential importance of the expression pattern of post-transcriptional regulators, particularly RNA-binding proteins, in marking these neuron clusters (Figure 5A). For instance, upf3a (the marker distinguishing subclasses of the pairs 23/25, 12/16, and 60/62, Figure 5-Source Data 5) is known to regulate mRNA stability via nonsense-mediated RNA decay (Shum, Jones et al. 2016), rbfox1/rbfox3a (the marker distinguishing subclasses of the pair 27/54, 13/24, 11/30, 23/25, 12/16, 2/3, and 7/38, Figure 5-Source Data 5) could mediate cell-type-specific splicing in cortical interneurons, assembly of axon initial segment, and synaptic transmission (Jacko, Weyn-Vanhentenryck et al. 2018, Vuong, Wei et al. 2018, Wamsley, Jaglin et al. 2018). In addition to post-transcriptional regulators, other factors could also be expected to involve neuron diversification, such as epigenetic regulators, translational regulators, protein stability, which are interesting to explore in future studies. Together, our findings suggest that combinatorial TFs and post-transcriptional regulators could work in concert to determine neuronal types in the larval zebrafish brain (Figure 6).

Regional identity is an essential factor for classifying brain cells. In the larval zebrafish brain, our results showed that neurons, qRG, and neuronal progenitors
exhibited prominent regional characteristics (Fig 1C). In the hierarchical clustering and population-level statistical analysis, similar pair clusters from the same region or different regions occur in nearly equal probabilities (Figure 3-figure supplement 2A).

Considering the fact that neuronal phenotypes from the same region could exhibit common regional identities, it was surprising to observe that such a higher proportion of sister subclasses with similar transcriptomic phenotypes could arise from two different brain regions. Putting this finding into the context of neuron type evolution, it raises the possibility that different brain regions independently give rise to similar neurotransmitter phenotypes through different TF programs. Alternatively, as brain regions are functionally diversified during the evolution, these highly similar neuronal clusters of different brain regions possibly derive from ancient building blocks, which become divergent through the evolutionary acquisition of different combinatorial TF codes.

Single-cell RNA-sequencing resolves the transcriptomes of brain cells at a given time point (Erhard, Baptista et al. 2019). On the other hand, gene expression is highly dynamic and sensitive to spontaneous or stimulus-evoked neuronal activities (West, Griffith et al. 2002, Kim, Hemberg et al. 2010, Yap and Greenberg 2018), circadian rhythm (Satchidananda Panda 2002, Takahashi 2017), as well as other extrinsic and intrinsic factors. Thus, gene profiles of individual neuron subclasses could represent dynamic cellular states at a given time rather than a static subclass (Bin Wu 2016, Ofengeim, Giagtzoglou et al. 2017). This issue could be resolved by comparing
transcriptome profiles of cell samples from different animals or from different developmental stages. In this study, the 68 cell subclasses that we have identified for the larval zebrafish (8 dpf) were confirmed in another fish brain of the same age and were largely recapitulated by cell subclasses identified from the juvenile zebrafish (23-25 dpf, Figure 1E and Figure 1-figure supplement 2F) (Raj, Wagner et al. 2018). Thus, transcriptome-defined subclasses in our study indeed largely represent stable neuronal subclasses. An alternative demonstration of stable neuronal subclasses is to perform genetic labeling of the brain using promoters of marker genes for transcriptome-defined subclasses. Reproducible labeling of the same morphological subclasses, as demonstrated in our morphological studies of tectal neurons expressing specific TFs, also supports that we have identified stable neuronal subclasses (Figure 4D-E). Identification of neuronal subclasses in the present study paves the way for understanding neuronal diversity. Future studies of temporal changes of single-cell transcriptome profiles in the whole brain could provide important insights into the dynamic changes of the brain.

Materials and methods

<table>
<thead>
<tr>
<th>Key Resources Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reagent type (species) or resource</td>
</tr>
<tr>
<td>Strain, Strain Background <em>(Danio rerio)</em></td>
</tr>
<tr>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Strain, Strain Background <em>(Danio rerio)</em></td>
</tr>
<tr>
<td>Strain, Strain Background <em>(Danio rerio)</em></td>
</tr>
<tr>
<td>Strain, Strain Background <em>(Danio rerio)</em></td>
</tr>
<tr>
<td>Strain, Strain Background <em>(Danio rerio)</em></td>
</tr>
<tr>
<td>Strain, Strain Background <em>(Danio rerio)</em></td>
</tr>
<tr>
<td>Recombinant DNA Reagent</td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>vglut2a:cre</td>
</tr>
<tr>
<td>zic1:gal4FF</td>
</tr>
<tr>
<td>bhlhe22:gal4FF</td>
</tr>
<tr>
<td>en2b:gal4FF</td>
</tr>
<tr>
<td>foxb1a:gal4FF</td>
</tr>
<tr>
<td>zbtb18:gal4FF</td>
</tr>
<tr>
<td>irx1a:gal4FF</td>
</tr>
<tr>
<td>Single Cell 3' Library and Gel Bead kit v2</td>
</tr>
<tr>
<td>Commercial assay or kit</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>ClonExpress MultiS One Step Cloning Kit</td>
</tr>
<tr>
<td>chemical compound, drug</td>
</tr>
<tr>
<td>chemical compound, drug</td>
</tr>
<tr>
<td>chemical compound, drug</td>
</tr>
<tr>
<td>chemical compound, drug</td>
</tr>
<tr>
<td>chemical compound, drug</td>
</tr>
<tr>
<td>chemical compound, drug</td>
</tr>
<tr>
<td>Chemical Compound, Drug</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>FBS</td>
</tr>
<tr>
<td>DPBS</td>
</tr>
<tr>
<td>MS222</td>
</tr>
<tr>
<td>Low Melting Agarose</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Software, Algorithm</th>
<th>Data Analysis</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>R 3.5.1</td>
<td>R-project</td>
<td></td>
</tr>
<tr>
<td>Cell Ranger Single Cell Software Suite (v2.1.0)</td>
<td>10xgenomics</td>
<td>scRNA seq data analysis</td>
</tr>
<tr>
<td>Seurat</td>
<td><a href="http://satijalab.org/seurat/">http://satijalab.org/seurat/</a></td>
<td>scRNA seq data analysis</td>
</tr>
<tr>
<td>bioDist</td>
<td><a href="https://ww.bioconductor.org">https://ww.bioconductor.org</a></td>
<td>scRNA seq data analysis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bioconductor</th>
<th>R 3.5.1</th>
<th><a href="https://ww.r-project.org/">https://ww.r-project.org/</a></th>
<th>Data analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>10xgenomics</td>
<td></td>
<td><a href="https://support.10xgenomics.com/">https://support.10xgenomics.com/</a></td>
<td>scRNA seq data analysis</td>
</tr>
<tr>
<td>Seurat</td>
<td><a href="http://satijalab.org/seurat/">http://satijalab.org/seurat/</a></td>
<td>scRNA seq data analysis</td>
<td></td>
</tr>
<tr>
<td>bioDist</td>
<td><a href="https://ww.bioconductor.org/packages/release/bioc/html/bioDist.html">https://ww.bioconductor.org/packages/release/bioc/html/bioDist.html</a></td>
<td>scRNA seq data analysis</td>
<td></td>
</tr>
<tr>
<td>software, algorithm</td>
<td>scclusteval</td>
<td>R-package</td>
<td><a href="https://github.com/crazyhottomy/scclusteval">https://github.com/crazyhottomy/scclusteval</a></td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>software, algorithm</td>
<td>treeDist</td>
<td>R-package</td>
<td><a href="https://github.com/Treedist/">https://github.com/Treedist/</a></td>
</tr>
<tr>
<td>software, algorithm</td>
<td>FIJI</td>
<td>PMID:2274</td>
<td><a href="http://fiji.sc/">http://fiji.sc/</a></td>
</tr>
<tr>
<td>software, algorithm</td>
<td>GraphPad</td>
<td>GraphPad</td>
<td><a href="http://www.graphpad.com">www.graphpad.com</a></td>
</tr>
<tr>
<td>software, algorithm</td>
<td>FV10-ASW 4.0 Viewer</td>
<td>Olympus</td>
<td><a href="http://www.olympus-global.com">www.olympus-global.com</a></td>
</tr>
<tr>
<td>software, algorithm</td>
<td>NNLS</td>
<td>R-package</td>
<td><a href="https://github.com/rdevits/NNLS.jl">https://github.com/rdevits/NNLS.jl</a></td>
</tr>
</tbody>
</table>

**Animal maintenance and transgenic lines**

For all experiments in this study, zebrafish were maintained, mated, and raised at 28°C according to standard protocols. Animals were staged according to days post fertilization (dpf). Animal procedures performed in this study were approved by the
Transgenic lines used in this study include: Tg (glyT2:GFP) (McLean, Fan et al. 2007), Tg (vglut2a: loxp-DsRed-loxp-GFP) (Satou, Kimura et al. 2012), Tg (Etvmat2:GFP) (Wen, Wei et al. 2008), Tg (gad1b:EGFP) (Wang, Du et al. 2020).

**Single-cell sample preparation**

**Whole-brain and brain region identified cell populations**

Wild-type fish were processed for 10x Genomics single-cell transcriptome sequencing. Whole brains or different brain regions (forebrain, optic tectum (OT), hindbrain, and brain regions underneath OT (sub-OT)) were dissected as anatomical structures (Figure 1B, Figure 1-figure supplement 1A). Tissue were dissociated with 300 μL papain (28 units/mL, Worthington) in papain solution (1% DNase, 12 mg/mL L-cysteine in DMEM/F12), incubated at 37 °C for 15 mins with proper mix methods (Yu and He 2019). Dissociated cells were washed twice with washing buffer (100 ml: 650 μL 45% glucose, 500 μL 1M HEPES and 5 mL FBS into 93.85 mL 1× DPBS, sterilized with a 0.22 μm pore size filter), and sterilized with 40 μm cell strainers (BD Falcon) into 300–400 μL washing buffer.
Whole-brain neuromodulator or tectal glutamatergic neurons

For the transgenic fishlines, brains were dissected and dissociated as above; the cell suspension were used to sort cells with high fluorescence intensity with flow cytometry (Moflo XDP, Beckman Coulter). The interesting cells were transferred to washing buffer. In this paper, we used Tg (ETvmat2:GFP) transgenic fishlines to acquire neuromodulator-defined cell populations; Tg (vglut2a: loxp-DsRed-loxp-GFP) was used to collect glutamatergic neurons. Tg (ETvmat2:GFP) could label most monoaminergic neurons, including dopaminergic (DA) neurons, serotonergic (5-hydroxytryptamine, 5-HT) neurons, and norepinephrinergic (NE) neurons (Figure 2A, Figure 2-figure supplement 1B) (Wen, Wei et al. 2008).

Single-cell RNA sequencing on 10x Genomics Chromium platform

All sampling was carried out with 10x Genomics Chromium Single Cell Kit (Version 2): suspensions prepared as described above were diluted to concentrations 300-1000 cells/mL with washing buffer, then added 10x Chromium RT mix to achieve loading target cell numbers 13,000. Downstream cDNA synthesis (14 PCR cycles), library preparation, and sequencing, were carried out according to manufacturer’s instructions (https://www.10xgenomics.com/solutions/single-cell/).

Data analysis of single-cell RNA-seq
Bioinformatics processing of raw reads

Sequencing data (FASTQ files) were converted to matrices of expression counts using the Cell Ranger software (Version 2.0.1) provided by 10x Genomics. Transcriptome libraries were mapped to a zebrafish reference built from a custom GTF file and the zebrafish GRCz11 (Ensemble release-96) genome assembly. For each sample, cells with less than 600 detected molecules (UMIs), or less than 1.2-fold molecule to gene ratio, and genes detected in fewer than 20 cells or more than 60% of all cells were excluded for the following analysis.

Cell clustering analysis of cells from the whole brain

After filtering, the combined gene expression matrix of whole brain and four major brain region samples were combined (using Cellranger aggr pipeline) and loaded into Seurat package (Version 2.3.4) in R (Version 3.4.3) for the following analysis as described in tutorials (http://satijalab.org/seurat/). In brief, digital gene expression matrices were column-normalized and log-transformed. Cells with fewer than 500 expressed genes, greater than 5% mitochondrial content, very high numbers of UMIs and gene counts that were outliers of a normal distribution (likely doublets/multiplets) were removed from further analysis. Combined all samples together, we got 45,746 cells, which included 4 whole brain samples, four specific brain regions (n = 2 each) including the forebrain (Fore), optic tectum (OT), hindbrain (Hind), and the region underneath the optic tectum (sub-OT). 1,402 variable genes were selected for
principal component analysis (PCA) by binning the average expression of all genes into 300 evenly sized groups, and calculating the median dispersion in each bin (parameters for MeanVarPlot function: $x\.low\.cutoff = 0\.0125$, $x\.high\.cutoff = 8$, $y\.cutoff = 0\.5$). The top 100 principal components (PCs) were used for the first round of clustering with the Louvain modularity algorithm (FindClusters function, resolution = 2.5), which generated 68 distinct clusters (Figure 1A). Marker genes for each cluster were calculated using FindAllMarkers function in Seurat (parameters: min.pct = 0.1, min.diff.pct = 0.25).

**Clustering analysis of neuromodulator-defined cell populations**

After filtering as above and selected all cells with neuromodulators marker gene expression ($vmat2^+$ monoaminergic neurons: $th^+/thph2^+/th2^+/dbh^+$, Figure 2-figure supplement 1C, and Figure 2- Source Data 1), variable genes ($n = 2,120$) were selected for principal component analysis (PCA). The top 46 principal components were used for the first round of clustering with the Louvain modularity algorithm (FindClusters function, resolution = 0.6). In total, 5,398 cells obtained from Tg (ET$vmat2$:GFP) fishline expressed monoaminergic neuromodulators, showing 22 clusters (Figure 2B). Marker genes for each cluster were calculated using FindAllMarkers function in Seurat (parameters: min.pct = 0.1, min.diff.pct = 0.25)
Clustering analysis of glutamatergic neurotransmitter-type cell populations

After filtering as above and selected all cells with tectal glutamatergic marker gene expression (vglut2a, mab21l2, tubb5, and tfap2a, Figure 4-figure supplement 1C), we obtained 3,883 cells from optic tectum with Tg (vglut2a:loxp-DsRed-loxp-GFP) fishline expression tectal glutamatergic neurons, showing 11 clusters using 1,276 variable genes and 37 PCs (Figure 4A). Marker genes for each cluster were calculated using FindAllMarkers function in Seurat (parameters: min.pct = 0.1, min.diff.pct = 0.25)

Jaccard similarity index

To evaluate the stability of each cluster, we performed R package “scclusteval”, in which we re-sampled a subset of the cells (80%) from the population and repeated clustering, and then used the Jaccard index to evaluate cluster similarity before and after re-clustering. If a cluster is robust and stable, random subsampling and re-clustering will keep the cell identities within the same cluster. After repeating the re-clustering 20 times, we used the mean/median of the Jaccard index as the metric to evaluate the stability of the clusters (Tang, Kaymaz et al. 2020). The distribution of the Jaccard index across subsamples measures the robustness of the cluster. Clusters with a mean/median stability score less than 0.6 should be considered unstable (Zumel N. 2014). Here, we used JaccardRainCloudPlot to visualize the Jaccard index
and found most clusters of whole brain (Figure 1-figure supplement 1E), \textit{vmat2}⁺ neuromodulator (Figure 2-figure supplement 1A), and tectal glutamatergic (Figure 4-figure supplement 1D) were stable.

**Identification of brain region markers**

To identify potential brain region-specific markers that exist in all cell types, clusters were assigned when most of the cells in a cluster intermingle with cells from one particular brain region (Figure 1-figure supplement 2B). Then, we defined differentially expressed genes shared by all cell types (\textit{vglut}⁺, glutamatergic neurons; \textit{gad}⁺, GABAergic neurons; \textit{pcna}⁺, neuroprogenitors; \textit{cx43}⁺, radial astrocytes) to represent the targeted brain region-specific markers. We separated each cell type in four different brain regions as individual samples, FindAllMarkers function in Seurat was used to identify the differentially expressed markers of a given cell type. We listed these brain-region specific markers in Figure 1C, and these genes exhibited specific expression pattern in all cell types of particular brain region (Figure 1D). Furthermore, these brain region specific genes were used to identify the regional origins of each cell clusters. Only clusters with over 5% of cells expressing the marker genes or with averaged UMI of genes-expressing cells over 2 were assigned regional identities (Figure 1-Source Data 3).
The regional distribution of whole-brain neurotransmitter-type neurons

According to the marker genes that are specific to primary neurotransmitter phenotypes, including slc17a6b (glutamatergic neurons), gad1b (GABAergic neurons), and slc6a5 (glycinergic neurons), we identified subsets of neurons with different neurotransmitter-type as clusters with over 5% of cells expressing the marker genes or with averaged UMI of genes-expressing cells over 2 (Figure 1-Source Data 3). We found that the forebrain was predominantly populated by glutamatergic neurons, whereas glycinergic neurons mainly resided in the hindbrain (Figure 1-figure supplement 2C-D). Live image with fishline Tg (glyT2:GFP), Tg (vglut2a: loxp-DsRed-loxp-GFP), and Tg (gad1b:EGFP) support the results above (Figure 1-figure supplement 2E).

Annotation of each cluster in neuromodulator-type neurons

To assign brain region identity for each neuromodulator-type neurons, we used markers dlx5a, otpa, dlx2a, barhl2 for sub-OT, en2a for OT, mab21l2 for midbrain and hindbrain, and phox2bb, hoxb3a, mafb, and efnb2a for hindbrain (Figure 2-figure supplement 1D). As for the subtype of neuromodulator-type neurons, we used markers th for dopaminergic (DA) neurons, tph2 for serotonergic neurons, dbh for norepinephrinergic neurons, and slc5a7a, chata, slc18a3a as markers of cholinergic neurons (Figure 2-figure supplement 1C). And defined clusters with over
5% of cells expressing the marker genes or with averaged UMI of genes-expressing cells over 2 as a threshold for neurons with their identities (Figure 2-Source Data 1).

In addition, we used markers *gad1b, gad2* for GABAergic and *vglut1, vglut2a, vglut2b, vglut3* for glutamatergic (Figure 2C) as a method to assign co-expression pattern of neurotransmitter for each neuromodulator-type neurons.

Comparison of the single-cell transcriptome between larval and juvenile zebrafish

The Lawson-Hanson algorithm for non-negative least squares (NNLS) was applied to compare our data (8 dpf) with juvenile zebrafish (23-25 dpf, GSE105010). Top 20 (Listed in Figure 1-Source Data 4) marker genes defined in our data that are also highly variable genes in juvenile data were used for decomposition. Decomposition was performed on mean expression values by averaging our dataset or by averaging other single-cell datasets using cluster assignments provided by the authors. NNLS was implemented using the ‘nnls’ package in R (Version 3.4.3). The 68 clusters we identified exhibited a high overlap with their counterparts in the juvenile zebrafish brain. Meanwhile, we also found clusters with specific regional origins and cell types exhibited high correlation with their juvenile counterparts (Figure 1E and Figure 1-figure supplement 2F).
Hierarchical clustering analysis based on effector gene and TF profiles

The effector genes (n = 1,099) in highly variable genes (n = 1,402) were selected for the clustering analysis of whole-brain clusters, and mean expression value of each gene was calculated for each cluster. The distance matrix was defined as (1-Pearson correlation coefficient between clusters)/2. Hclust function (clustering method: ward.D) implemented in R (Version 3.4.3) was applied for the hierarchically clustering. Six major brain cell types were identified using height = 0.42 as cutoff (shown with different color in Figure 1A, 1F). For each cell type, pan-enriched genes were identified using FindAllMarkers function in Seurat (Parameter: min.pct = 0.1, min.diff.pct = 0.25). We identified these six major cell type as: branch I (cerebellum cells and habenula cells), branch IIa (glutamatergic neurons), branch IIb (inhibitory neurons), branch III (neuroprogenitors), branch IV (radial astrocytes), and branch V (others: oligodendrocytes, microglia, and endothelia cells) according to marker genes of each branch (Figure 1-Source Data 5, Figure 1F).

Besides, we also selected 283 TFs in variable genes (n = 1,402) to do hierarchical clustering analysis as above (Figure 3-figure supplement 1B, 2A). Meanwhile, we performed hierarchical analysis for vmat2 neuromodulator neurons based on effectors (n = 1,783) and TFs (n = 319) in variable genes (Figure 3-figure supplement 3A-B).
Moreover, R package “pvclust” were used to assess the uncertainty in hierarchical clustering by performing bootstrap analysis of clustering (Figure 3-figure supplement 1A). Groups strongly supported with AU (approximately unbiased) p-value > 0.9 were highlight.

**Identification the similarity of pair clusters using two strategies**

**Hierarchical sister clusters analysis**

Before analysis, we compare the similarity of tree based on effector gene and TF profiles with R package “TreeDist”. In glutamatergic/GABAergic neurons, these two tree shown different organization, with only one matching node (node 13, cluster 9 and 61, tree distance = 0.7127991, Figure 3-figure supplementary 1B); while in neuromodulator neurons, these two tree shown much similar organization, with 8 matching nodes (tree distance = 0.3766735, Figure 3-figure supplementary 3A-B).

Sister clusters in the terminus of tree indicate pair cluster with most similar gene expression, we focus on 39 glutamatergic (IIa) and inhibitory (IIb) neurons and identified 11 pairs of sister clusters that exhibited high similarity in effector gene profiles and 14 sister clusters with similar TF expression (Figure 3-figure supplement
2A). And for \textit{vmat2}+ neuromodulator neurons, we identified 8 sister clusters in the terminus branches of both effector-based hierarchy tree and TF-based tree (Figure 3-figure supplement 3C).

Population-level statistical analysis

To compare the landscape of TF and effector gene expression accounting for the full spectrum of cell types rather than just the most similar sister clusters, we also performed population-level statistical analysis. First, we calculate any distance of two clusters that were randomly picked from 39 glutamatergic/GABAergic neuronal clusters ($C_{39}^2$). If the distance is within the lowest 10% population of all distances containing either of two randomly-selected clusters, we defined it a similar pair cluster; while the distance is within the top 80% population of all distances containing either of two randomly-selected clusters, we defined it as a dissimilar pair (Figure 3A). Then we performed these analysis on both effector gene profiles and TFs, and defined those pair clusters similar in both effector-based and TF-based tree as ‘matched pattern’; those pair clusters similar in effector-gene profiles but not TFs as ‘convergent pattern’, and those pair clusters similar in TF profiles but not effector-gene profiles as ‘divergent pattern’ (Figure 3B). In matched pattern, the distance of pair cluster were the lowest distance in both effector gene-based (white) and TFs-based (gray) distances (Figure 3-figure supplement 2D). In convergent pattern, the distance of pair cluster were the lowest distance in effector gene-based
(white) distances, but not the lowest distance in TFs-based (gray) distances (Figure 3-figure supplement 2E). In divergent pattern, the distance of pair cluster were the lowest distances in TFs-based (gray) distance, but not the lowest distance in effector gene-based (white) distances (Figure 3-figure supplement 4B).

Moreover, we performed population-level statistical analysis with vmat2⁺ neuromodulator neurons, and found 9 pairs with matched pattern which showing the lowest distance in both TF and effector gene-based distances; 2 pairs with convergent pattern which showing the lowest distance in effector gene-based distance, but not in TF-based distance (Figure 3-figure supplementary 3E).

However, this analysis is subject to the intrinsic statistics of TF and effector gene expression, as well as the sensitivity of scRNAseq method. Thus, we intersected the results from hierarchical sister cluster and population-level statistical analysis to identify pairs with each patterns (Figure 3B and Figure 3-figure supplementary 3G).

**Subsampling of genes in population-level statistical analysis**

To test the robustness of three patterns based on distance measures, we subsampling 80% gene inputs in population-level statistical analysis. First, we random subsampled 80% effector genes/TFs, calculated their distance defined as (1-Pearson correlation coefficient between clusters)/2. Then average statistics over 20 times sub-sampling...
calculations and performed population-level statistical analysis as above. These re-identified pairs completely recapitulated those identified using the population-level statistical analysis based on total genes, suggesting the robustness of three patterns in glutamatergic/GABAergic neurons (Figure 3-supplemental 2F). Besides, 8 out of 12 re-identified pairs with matched pattern, 1 out of 3 re-identified pairs with convergent pattern were consistent with those identified using the population-level statistical analysis based on total genes (Figure 3-figure supplementary 3F).

Moreover, we quantified the distance of two similar trees that was exemplified by TF-based or effector gene-based clusters with or without subsampling of 80% genes, and the result showed that the tree distances of TF-based or effector gene-based clusters before and after subsampling were only 0.20 (TF-based) and 0.14 (effector gene-based), respectively (Figure 3-figure supplementary 2F-G). These distances are smaller than the distance between TF and effector-gene profiles (0.7127991, Figure 3-figure supplementary 1B). All above results indicated the overall distinction of the landscape of TF-based and effector gene-based clusters.

Comparison of two rounds of tectal glutamatergic transcriptome data

Canonical correlation analysis (CCA) in Seurat was used to test the reproducibility of two rounds of tectal glutamatergic transcriptome data. Two rounds of transcriptome
data were intermingled with each other in tSNE plot showing a better performance of two parallel transcriptome data (Figure 4-figure supplement 1A).

Sorted tectal glutamatergic transcriptome data yielded more clusters

The Lawson-Hanson algorithm for non-negative least squares (NNLS) was applied to compare the transcriptome data from sorted tectal glutamatergic neurons and the subset of these neurons from whole brain. The top 20 marker genes defined in sorted sample which shared in whole-brain sample were used for decomposition. Decomposition was performed on mean expression values by averaging our dataset. NNLS was implemented using the ‘nnls’ package in R (Version 3.4.3). Sorted samples yield more clusters than whole brain samples. Besides, sorted sample clusters exhibited diversity with their counterparts in the whole brain sample (Figure 4-figure supplement 1B).

Differentiation expression of genes encoding RNA binding protein between paired clusters

For each paired clusters, differential expression genes were identified using FindMarkers function in Seurat (Parameter: min.pct = 0.1, min.diff.pct = 0.25). Then we mapped the expression of genes encoding RNA binding protein within each paired
clusters in different patterns (matched, convergent, and divergent patterns, Figure 5-Source Data 1, Figure 5-figure supplement 1B-D). Comparing the RBPs used in each pattern, we identified pattern specific RBPs (Figure 1-figure supplementary 1E-F).

**Searching the binding sites of pattern specific RBPs**

We take advantage of the oRNAment database (Benoit Bouvrette, Bovaird et al. 2020) to search the putative target sites of RBPs. We only found 3 matched pattern specific RBPs (*celf1, noval, pum1*), 3 convergent-pattern specific RBPs (*msi2a, taf15, tial*), 18 divergent pattern specific RBPs (*eif4a2, elavl1, hnrnpa1b, igf2bp3, hnrnp12, mbnl2, mbnl3, pchp4, raly, rbm25b, rbm28, sart3, sf3b4, sfpq, snrnp70, snrpa1, taf13, u2af2b*) in the database. Then, we searched the putative binding sites of these RBPs use Score > 0.5 as a cutoff (matched: 2987 binding sites; convergent: 2755 binding sites; divergent: 2856 binding sites, Figure 5-Source Data 4).

**Gene Ontology**

R package “clusterProfiler” was used to analyze and visualize function profile of gene and gene clusters (Yu, Wang et al. 2012). Bioconductor annotation packages org.Dr.eg.db were imported for genome-wide annotation of mapping Entrez gene
identifiers or ORF identifiers for zebrafish. We performed GO analysis on 1,402 variable genes of whole-brain and 2,120 variable genes of \textit{vmat2}$^+$ neutromodulator-type neurons, and found most genes belong to effector genes, including receptors, neuropeptide, ion channel, and so on (Figure 1-figure supplement 2G).

We performed GO analysis on differential expression genes between sister pair clusters with different patterns (convergent, and divergent), and found these genes belong to TFs, effectors and genes encoding RNA binding protein (Figure 5A, Figure 5-figure supplement 1A).

Besides, we performed GO analysis of binding target sites of pattern specific RBPs, showing the TFs, effectors, mRNA processing are the major targets in these RBPs (Figure 5- figure supplement 1H-J). And found the percentage of effectors in all binding sites of divergent pattern were higher than matched and convergent pattern (matched: 5%; convergent: 3.8%, divergent 7.6%, Figure 5E).

**Morphologies analysis of tectal glutamatergic neurons**

**Plasmids**

To trace the morphology of tectal glutamatergic neurons, we designed three plasmids
In the first plasmid, the Gal4FF cassette was placed at the start-codon site using a bacterial artificial chromosome (BAC) recombination technique (Suster et al., 2011). Using this method, we created the Gal4FF BAC plasmids for 15 TF marker genes covering all 11 clusters of tectal glutamatergic neurons. Considering the promoter and other regulatory elements of higher expression genes may not be powerful to drive BAC plasmid expression, we selected 6 of them for their better performance in labeling tectal glutamatergic neuron morphologies based on two criterions: 1. These TFs are highly expressed and specific to individual tectal glutamatergic clusters based on single-cell RNAseq analysis; 2. Their BAC plasmids could reliably mark particular morphological subclasses (at least in four animals). The second BAC plasmid vglut2a:CRE was constructed using similar method with BAC (CH211-111D5). The CRE cassette was placed at the start-codon site of vglut2a (CH211-111D5) using BAC recombination technique. The third plasmid uas:loxp-stop-loxp-tdTomatax was constructed with the ClonExpressMultiS One Step Cloning Kit (Vazyme, C112-01/02) according to the standard protocol. We ligated three PCR fragments: loxp-stop-loxp, tdTomatax and 10× uas backbone (Figure 4C). With these tree plasmids, vglut2a promoter driven Cre-mediated excision, uas-Gal4 system activate tdTomatax expression as Figure 4D shown.

**Imaging**

Embryos at 8 dpf were used for imaging. The fish were anesthetized with 0.04% MS222 (Sigma, A5040) and embedded in 1% agarose (Sigma, A0701) on imaging
dish. Imaging were performed on Olympus FV1200 confocal using 30 × oil-immersion objectives (Olympus, NA = 1.2). The image data were analysis with Image J (ImageJ, https://imagej.net/Simple_Neurite_Tracer). In total, we collected 574 single neuronal morphologies from 263 distinct fish (Figure 4-Source Data 1 and Figure 4-figure supplement 1G). All neurons were categorized into 7 morphological subclasses: bi-stratified (I, n=36), mono-stratified (II, n=71), non-stratified (III, n=318), necklace-like (IV , n=76), cross-hemispheric (V, n=15), ascending (VI, n=18), and descending (VII, n=40) (Figure 4D).

The analysis of TFs and tectal morphological subclasses

To calculate the expression of 6 TFs (en2b, irx1a, zbtb18, foxb1a, bhlhe22, and zic1) with respect to 11 tectal glutamatergic clusters, we created a binary expression matrix consisting of the gene expression (ON: with expression; OFF: without expression) and cluster identity (Li, Horns et al. 2017). We marked TFs that have expression in each of 11 clusters with cutoff of 5% cells or have averaged UMI of each TF-expressed cell UMI = 2 as ON, shown with black box. The rests were marked as OFF, shown with white box (Figure 4-Source Data 1 and Figure 4-figure supplement 1E-F). All 6 TFs marked multiple transcriptome-based subclasses in a combinatorial manner (Figure 4-figure supplement 1F).
To calculate the expression of 6 TFs with respect to 7 tectal glutamatergic morphological subclasses, we also created a binary expression matrix consisting of the gene expression (ON: with expression; OFF: without expression) and morphological identity. We marked genes labeling one morphological subclass at least 4 times in individual 4 fish as ON, shown with black box. The rests were marked as OFF, shown with white box. All 6 TFs marked multiple morphological subclasses in a combinatorial manner (Figure 4E).

These two binary matrixes were used to test the correlation between transcriptome-based clusters and morphological subclasses of tectal glutamatergic neurons with NNLS. In total, 6 out of 11 transcriptomic clusters exhibited strong correlation with at least one morphological subclass (Figure 4-figure supplement 1H).

**Data availability**

Single cell RNA-seq data has been deposited on BIG (CRA002361)

https://bigd.big.ac.cn/gsub/submit/gsa/subCRA002669.

**Code availability**

All the code for the data analysis is available.

**Acknowledgements**
We thank Dr. Muming Poo for the discussion and manuscript editing. We thank facilities of the Institute of Neuroscience: Dr. Min Zhang (MZ) and Zhenning Zhou (ZZ) from Molecular and Cellular Biology Core Facility for the assistance of single-cell RNAseq; FACS facility Haiyan Wu (HW); the optical imaging facility Qian Hu (QH), Yumei Zhang (YZ) and Yonghong Wang (YW); the bioinformatics core facility and the facility of mapping brain-wide mesoscale connectome. This research was funded by grants from the Shanghai Municipal Science and Technology Major Project (Grant No. 2018SHZDZX05), Strategic Priority Research Program of Chinese Academy of Science (Grant No. XDB32000000), State Key Laboratory of Neuroscience, Shanghai Basic Research Field Project (Grant No. 18JC1410100).

Competing interests
The authors declare no competing interests

References


Blue B. Lake, R. A., 2* Gwendolyn E. Kaeser, 3, 4* Neeraj S. Salathia, 5* Yun C. Yung, 3, A. W. Rui Liu, 2 Derek Gao, 1 Ho-Lim Fung, 1 Song Chen, 1, J. W. Raakhee Vijayaraghavan, 3 Allison


single-cell profiling of the hypothalamic preoptic region." Science 362(6416).


Figure legends

Figure 1
Figure 1. Molecular Classification of Whole-brain Cells in Larval Zebrafish

**Brain.** (A) The t-SNE plot of 45,746 single-cell transcriptomes pooled from whole brains (n = 4) and four different individual brain regions (n = 2 each). The pooled cells were aggregated into 68 clusters, marked by a number. Each color-coded the major cell type as F. (B) The schematic showing different samples separately examined by single-cell RNA-sequencing on 10x Genomics Drop-seq platform: whole brain (n = 4), forebrain (Fore, n = 2), optic tectum (OT, n = 2), hindbrain (Hind, n = 2), and the region underneath the optic tectum (sub-OT, n = 2). OB: olfactory bulb; Tel: telencephalon; OT: optic tectum; Th: thalamus; H: hypothalamus; Pit: pituitary; Ce: cerebellum; MO: medulla oblongata. (C) Venn plots showing the differentially expressed genes in 4 major cell types identified by cell-type marker genes (vglut*, glutamatergic neurons, Glu; gad1b*, GABAergic neurons, Gaba; pcna*, neuroprogenitors, P; cx43*, radial astrocytes, R) in three brain regions (Fore, Hind, and OT). Commonly expressed genes in all cell types for a given brain region were identified as region-specific genes: 6 for forebrain (Fore), 1 for optic tectum (OT),
and I for hindbrain (Hind), with genes listed below. (D) Dot plot showing the expression levels of region-specific marker genes in 4 major cell types (colored circles as C) in three brain regions. The gray level represents the average expression; dot size represents the percentage of cells expressing the marker genes. (E) NNLS analysis showed cell clusters of Fore, OT, and Hind exhibited a high correlation with their counterparts of the juvenile zebrafish. Degree of correlation in marker genes is coded by the gray level and size of circle. (F) The dendrogram for the taxonomy of 68 identified clusters based on effector gene profiles (n = 1,099). Main branches of neuronal and non-neuronal cells were classified into six branches (red dashed line) that include: I, cerebellum and habenula (hb); IIa, Glutamatergic neurons (Glu); IIb, inhibitory neurons (Gaba); III, neuroprogenitors (P); IV, radial astrocytes (R); V, others, including microglia, endothelial cells, and oligodendrocyte et. The colored dots and squares below indicate their regional origins and neurotransmitter-type, respectively.

**Figure 2**
Figure 2. Molecular Classification of Neuromodulator-type Neurons. (A) The schematic showing the procedure of collecting single-cell transcriptomes of neuromodulator neurons with Fluorescence-activated Cell Sorting (FACS). Using Tg (ETvmat2:GFP) fishline, we could isolate dopaminergic (DA), serotoninergic (5-HT), and norepinephrinergic (NE) neurons. (B) The t-SNE clustering analysis of 5,368 cells obtained from Tg (ETvmat2:GFP) fishline expressing monoaminergic neuromodulators, showing 22 clusters, each marked by a number, color-coding brain regions and cell type of each cluster. (C) Dot plot showing the expression of glutamatergic marker (vglut2a/vglut2b/vglu1/vglu3) and GABAergic marker (gad1b/gad2) in each of the 22 vmat2+ clusters in B. The neurotransmitter phenotypes were color-coded. Empty squares depict the ones with undefined neurotransmitter phenotypes. The average expression levels of these genes for all cells in each cluster were coded by the gray level. The percentage of cells expressing each gene within each cluster was coded by dot size.
Figure 3. The TF Regulatory Landscape in Whole-brain Neuronal Clusters. (A)

Schematic showing the strategies to assess the cluster similarity based on effector gene and TF profiles. We focused on clusters of whole-brain glutamatergic/GABAergic neurons and neuromodulator neurons. Similar pair clusters were identified by two strategies: the first strategy was based on hierarchical sister clusters. The second strategy was based on population-level statistical analysis, in which we calculated and ranked the distances of every two clusters from 39 neurotransmitter-type clusters \( C_{39}^2 \) and chose the ones with the lowest 10% distance as similar pair clusters. (B) Left: schematic showing the criteria of three patterns: pair clusters that were similar in both TF and effector gene profiles as “matched pattern”, those pair clusters that were similar in effector gene profiles but not in TF profiles as “convergent pattern”, and those pair clusters that were similar in TF profiles but not in...
effector gene profiles as “divergent pattern”. Right: venn plot showing the number of each pattern using two strategies in A. The red dashed circle showing the number of cluster pairs with given pattern based on hierarchical sister cluster analysis; the black solid circle showing the number of cluster pairs with given pattern based on population-level statistical analysis. (C-D) Violin plots showing the expression of TFs (yellow) or effector genes (black) in glutamatergic and GABAergic similar pair clusters of convergent pattern. (E) The bar plot showing the proportions of different patterns for neuronal clusters with neurotransmitter or neuromodulator-types. The numbers of each pattern were indicated. Fisher’s exact test was used to test the significant association of different patterns, p = 0.02564, *p<0.05. (F-G) Violin plots showing the expression of TFs (yellow) or effector genes (black) in neuronal clusters of divergent pattern.

Figure 4
Figure 4. Combinatorial TFs in Marking Tectal Morphological Subclasses. (A) Left: The schematic showing the procedure of collecting single-cell transcriptomes of tectal glutamatergic neurons with Fluorescence-activated Cell Sorting (FACS) using Tg (vglut2a:loxp-DsRed-loxp-GFP) fishline. Right: the t-SNE plot of 3,883 cells obtained from Tg (vglut2a:loxp-DsRed-loxp-GFP) fishline expressing glutamatergic neurons in the optic tectum showing 11 clusters, each color-coded and marked by a number. (B) Schematic showing the designing of the Gal4FF BAC plasmids for 6 TF marker genes, covering all 11 clusters of tectal glutamatergic neurons shown in A. (C) Schematic showing the method for single TF-based labeling of tectal glutamatergic neurons. CRE expressed in glutamatergic neurons to excise loxp and drive fluorescent tdTomatoaax expression. (D) Representative images of 7 subclasses of tectal glutamatergic neurons with distinct morphological subclasses using the method described in B-C. The number of neurons collected for each subclass was shown. Insets depicted the morphological characteristics. Neurons with ascending and descending projections were those projecting to the forebrain and hindbrain.
respectively. Solid lines marked the boundaries between brain regions. Dashed lines marked the boundary of the tectal neuropile. Scale Bars: 20 μm. (E) The matrix showing the expressions of 6 TFs in each of 7 morphological subclasses. The black squares represented TFs could label particular morphological subclasses, and no expression was indicated by white squares.

**Figure 5**

**Figure 5.** The Post-transcriptional Regulatory Landscape in Neurons with Different Terminal Features but Similar TF Profiles. (A) Dot plot showing GO analysis of differentially expressed genes between each pair clusters with divergent pattern. Red highlighted the category of post-transcriptional regulators. The Number of genes in each category were coded by dot size, p.adjust of each category were color-coded. (B) Violin plots showing the expression of genes encoding TFs (yellow), RBPs (blue), and effector genes (black) in two pairs of neuronal clusters with
(C) 3D pie plot showing the functional annotation of differentially expressed RBP genes identified in A. (D) Bar plot showing the number of pattern specific RBPs per pairs in different patterns. Each pattern was color-coded. (E) Bar plot showing the effector binding target proportion of pattern specific RBPs. Each pattern was color-coded. (F) Bar plot showing the percentage of one RBP used as differentially expressed genes between pair clusters with different patterns.

Figure 6

Figure 6. Organization of Transcriptional and Post-transcriptional Regulators in the Specification of Brain-wide Neuronal Clusters. (A) Graphic summary describes the general organization of TFs and post-transcriptional regulators in the specification of neuronal clusters at the whole-brain level. Effector genes described the terminal features of whole-brain neuronal classification. Neuronal clusters with similar terminal features (named ‘sister clusters’) could share TF profiles (“Matched”, left), or exhibit distinct TF profiles (“Convergent”, middle). Besides, neuronal clusters with distinct effector gene profiles (terminal features) can share TFs profiles (“Divergent”,
right), and in each pairs, two neuronal clusters are differentially marked by the expression of RNA-binding proteins (RBPs) at the post-transcriptional levels. Moreover, sister clusters with ‘convergent’ and ‘divergent’ patterns are glutamatergic/GABAergic neurons, whereas sister clusters with ‘matched’ pattern is neuromodulator-type neuron.

**Supplement files**

**Figure 1-figure supplement 1**

**Figure 1-figure supplement 1. Molecular Classification of Whole-brain Cells in Larval Zebrafish Brain.** (A) The schematic showing each samples of whole brain and different brain regions. (B) t-SNE plot of pooled single-cell transcriptome data from whole-brain samples (n = 4). Four whole-brain replicates were largely overlapping in the plot. Numbers of repeats are color-coded. (C) t-SNE plot of pooled
single-cell transcriptome data from four different brain region (n = 2, each). Two
replicates from forebrain (Fore), optic tectum (OT), hindbrain (Hind), and the region
underneath the optic tectum (sub-OT) were largely overlapping in the plots. Different
replicates were coded by different colors shown on the right. (D) The plot showing
the number of total clusters, stable clusters, and the ratio of stable/total using different
parameters to subsamples and re-clustering using R package “scclusteval”. (E) The
JaccardRainCloudPlot gives an intuitive sense of the stability of clusters (with 0.6
cutoffs) with 68 whole-brain clusters in Figure 1A. Red line indicated the 0.6 cutoff of
Jaccard index to evaluate the cluster stability. Red arrowhead indicated unstable
clusters.

Figure 1-figure supplement 2
Figure 1-figure supplement 2. Molecular Classification of Whole-brain Cells in Larval Zebrafish Brain. (A) t-SNE plots showing the expression (in red) of specific markers (*eomesa*, *foxg1a*, *dlx5a*, *pitx2* as markers for the forebrain; *tal1*, *en2a* as markers for the optic tectum; *phox2a*, and *hoxa3a* as makers for the hindbrain). (B) The 68 clusters obtained in Figure 1A were further marked by their regional origins in colors. (C) The t-SNE plots of glutamatergic marker *vesicular glutamate transporter 2a* (*vglut2a*, shown as *slc17a6b*), GABAergic marker *gad1b*, and glycinerergic maker *glycine transporter 2* (*glyT2*, shown as *slc6a5*). (D) Pie charts showing the compositions of glutamatergic, GABAergic, and glycinerergic neurons in four different brain regions: Fore, OT, sub-OT and Hind (colored dots). (E) Representative images showing whole-brain maps of three major neuronal types using transgenic fish lines.
with specific labeling of neurons expressing different neurotransmitters. Glutamatergic neurons: Tg \(\text{vglut2a:loxP-DsRed-loxP-GFP}\); GABAergic neurons: Tg \(\text{gad1b:EGFP}\); glycinergic neurons: Tg \(\text{glyT2:GFP}\). Solid lines marked the boundaries between brain regions. Tel, telencephalon. Scale bars: 100 μm. Red arrowhead indicated the signal of \(\text{glyT2}^+\) neurons which were enriched in the hindbrain. (F) The NNLS (Lawson-Hanson algorithm for non-negative least squares) analysis using cluster-specific marker genes (top 20) showed that cell clusters of the sub-OT and non-neuronal type exhibit a high correlation with their counterparts of the juvenile zebrafish brain previously reported, with the degree of correlation in marker genes coded by the gray level and size of circle. (G) GO analysis of 1,402 variable genes in Figure 1A. Here we shown the effector genes related to different molecular function. Dots represent term enrichment, degrees of enrichments are color coded, and sizes of dots represent the percentage of each enrichment category.

Figure 2-figure supplement 1
**Figure 2-figure supplement 1. Molecular Classification of Neuromodulator Neurons.** (A) The JaccardRainCloundPlot showing the stability of $vmat2^+$ neuromodulator clusters in Figure 2B. Red line indicated the 0.6 cutoff of Jaccard index to evaluate the cluster stability. Red arrowhead indicated unstable clusters. (B) Representative image of fishline Tg (ET$vmat2$:GFP). The fishline labeled different monoaminergic neuronal types. Scale bars: 100 μm. (C) Dot plot showing the cell type-specific markers of 22 $vmat2^+$ neuromodulator clusters, with the average level of their expressions in all cells of each cluster was color coded, percentage of gene expression was indicated by dot size. (D) tSNE plots showing the expression of different brain region-specific markers ($dlx5a$, $otpa$, $nr4a2b$, $en2a$) in 22 $vmat2^+$ neuromodulator clusters. (E) Representative confocal images of 8-dpf Tg (ET$vmat2$:GFP) and Tg ($gad1b$:gal4::uas:mcherry) zebrafish. Scale bars: 100 μm.
Right: magnified images showing localization of vmat2 and gad1b in hindbrain (Hind), and the regions underneath the optic tectum (sub-OT). Scale bars: 20 μm.

**Figure 3-figure supplement 1**

Figure 3-figure supplement 1. Hierarchical clustering analysis of whole-brain glutamatergic/GABAergic neurotransmitter-type neurons based on effector gene and TF profiles. (A) Hierarchical clustering of 39 glutamatergic/GABAergic neurotransmitter-type neurons based on 1,099 effector gene profiles in highly variable...
genes. Red boxes were highlighted to indicate the terminus sister clusters strongly supported by pvclust (AU > 0.95, p < 0.05). (B) Visualizing the matching nodes between TF-based tree and effector gene-based tree with R package “TreeDist”. Red line indicated the matching node (node 13, Cluster 9 and 61). Each split was labeled with a measure of similarity score. Bottom tree distance (0.7127991) indicated the different organization of TF-based and effector-based tree.

**Figure 3-figure supplement 2**

**Figure 3-figure supplement 2.** The TF Regulatory Landscape in Whole-brain glutamatergic/GABAergic Neuronal Clusters. (A) Schematic showing the sister clusters identification based on hierarchical clustering of 39 whole-brain
glutamatergic/GABAergic neurotransmitter neuronal clusters (IIa: glutamatergic neurons and IIb: GABAergic neurons, in blue and red region of t-SNE plot, respectively). Upper hierarchical clustering was based on 1,099 effector genes, red lines indicated 11 terminus sister pairs. The bottom hierarchical clustering was based on 283 TFs, red lines indicated 14 terminus sister pairs. (B) A schematic for matched pattern (left) and a summary table (right) showing that neuronal clusters at similar termini of TF-based hierarchy were matched as sister cluster pairs within effector gene-based hierarchy (n = 1). (C) A schematic for convergent pattern (left) and a summary table (right) showing that glutamatergic/GABAergic neuromodulator-type neuronal clusters at different termini of TF-based hierarchy were found as sister cluster pairs within effector gene-based hierarchy (n = 10). (D) Scatter plot showing the distance distribution of glutamatergic/GABAergic neurotransmitter-type neuronal clusters with matched pattern (shown below) according to the effector-based (white) and TFs-based (gray) distance. Distance defined as \((1 - \text{Pearson correlation coefficient between clusters})/2\). Red and black dots showed the distance of each pair clusters. In matched pattern, the red and black dots were the lowest distance in both effector gene-based (white) and TF-based (gray) distances. (D') The table showing the brain regions and neurotransmitter type of each pair clusters with matched pattern (n = 5). Regional origins and neurotransmitter type are color-coded. (E) Scatter plot showing the distance distribution of glutamatergic/GABAergic neurotransmitter-type neuronal clusters with convergent pattern (shown below) with other clusters according to the effector-based (white) and TFs-based (gray) distance. Distance defined as \((1-\)
(Pearson correlation coefficient between clusters))/2. Red and black dots showed the distance of each pair clusters. In convergent pattern, the red and black dots were the lowest distance in effector gene-based (white) distances, but not the lowest distance in TF-based (gray) distances. (E') The table showing the brain regions and neurotransmitter type of each pair clusters with convergent pattern (n = 14). Regional origins and neurotransmitter type are color-coded. (F) Heatmap showing the distances distribution of 39 glutamatergic/GABAergic neurons based on effector gene profiles. The hierarchical clustering in the heatmap is similar as Figure 3-figure supplementary 2A (upper). (F') Heatmap showing the distances distribution of 39 glutamatergic/GABAergic neurons based on subsampling 80% of effector genes. The hierarchical clustering in the heatmap is similar as Figure 3-figure supplementary 2A (upper). The tree distance before and after subsampling was 0.1407203 (calculated by “treeDist”). (G) Heatmap showing the distances distribution of 39 glutamatergic/GABAergic neurons based on TF profiles. The hierarchical clustering in the heatmap is similar as Figure 3-figure supplementary 2A (bottom). (G') Heatmap showing the distances distribution of 39 glutamatergic/GABAergic neurons based on subsampling 80% of TF profiles. The hierarchical clustering in the heatmap is similar as Figure 3-figure supplementary 2A (bottom). The tree distance before and after subsampling was 0.2032071 (calculated by “treeDist”). (H) Summary table showing the pairs of neuronal clusters with three patterns before and after subsampling. Re-identified patterns using subsampling completely recapitulated those pairs identified using the population-level statistical analysis based on total genes.
Figure 3-figure supplement 3. Terminus sister clusters of neuromodulator-type neurons. (A) Hierarchical clustering of neuromodulator-type neurons based on 1,783 effector genes or 319 TFs in highly variable genes. The tree plots visualizing the matching nodes between TF-based tree and effector gene-based tree with R package “TreeDist”. Red line indicated the matching node. Bottom tree distance (0.3766735) indicated the similarity organization of TF-based and effector-based tree. (B) Each split node in (A) were labeled with a measure of similarity score. Matching nodes
The table showing the terminus sister pairs of $vmat2^+$ neuromodulator neurons from effector gene-based (C) or TF-based (C') hierarchical sister cluster analysis. The neuromodulator identities, brain region origins, and co-expression neurotransmitters were color-coded. (D-D') A schematic for matched pattern (D), convergent pattern (D') showing 5 and 3 similar neuromodulator pairs with matched and convergent pattern, respectively, based on hierarchical sister cluster analysis. The neuromodulator identities were color-coded. (E) Scatter plots showing the distribution of the distances of clusters of each $vmat2^+$ neuromodulator-type clusters exhibiting matched (light red) and convergent (light orange) patterns to other clusters using the effector-based (white) or TF-based (gray) population-level statistical analysis. Red and black dots represented the distances between two similar clusters. In matched pattern, the red and black dots had the lowest distances in both effector gene-based (white) distance cohort, and TF-based (gray) distance cohort. In convergent pattern, the red and black dots had the lowest distance in effector gene-based (white) cohort, but had higher distances in TF-based (gray) distance cohort. (F) Summary table showing the pairs of similar neuromodulator clusters with two patterns in (E) or after subsampling. (G) Venn plots showing the number of each pattern using two strategies (hierarchical sister cluster and population-level statistical analysis) in Figure 3A. The red dashed circle showing the number of cluster pairs with given pattern based on hierarchical sister clusters; the black solid circle showing the number of cluster pairs with given pattern based on population-level statistical analysis. (H) Dot plot showing the TFs expression in matched terminus pairs from
vmat2+ neuromodulator neurons. Sister clusters in a pair exhibited similar TFs expression patterns (red box). (I) Pheatmap showing the similar effector gene expression between each sister clusters of vmat2+ neuromodulator neurons with matched pattern.

**Figure 3-figure supplement 4**

**Figure 3-figure supplement 4. Divergent pattern of glutamatergic/GABAergic neurotransmitter-type neuronal clusters.** (A) A schematic for divergent pattern (left) and a summary table (right) showing that neuronal clusters as sister cluster pairs within TF-based hierarchy were divergent at different termini within effector gene-based hierarchy (n = 13). Colored dots represent the regional origins of each cluster. (B) Scatter plot showing the distance distribution of glutamatergic/GABAergic neurotransmitter-type neuronal clusters with divergent
pattern (shown with fork) with other clusters according to the effector-based (white) and TF-based (gray) distance measures. Distance defined as \((1 - \text{Pearson correlation coefficient between clusters})/2\). Red and black dots showing the distance of each pair clusters with divergent pattern. In divergent pattern, the red and black dots were the lowest distances in TF-based (gray) distance, but not the lowest distance in effector gene-based (white) distances.

Figure 4-figure supplement 1

Figure 4-figure supplement 1. Combinatorial TFs in Marking Tectal Morphological Subclasses. (A) Canonical correlation analysis (CCA) of cells from two independent experiments of Tg (vglut2a:loxp-DsRed-loxp-gfp), two rounds
overlapping indicated data reproducibility. Two rounds of data were color-coded. (B) NNLS analysis showing the relationship of transcriptome-based clusters of tectal glutamatergic neurons from sorted samples with Tg (vglut2a:loxp-DsRed-loxp-gfp) and whole-brain samples. Sorted clusters could cover all clusters from whole-brain sample. (C) tSNE plots of expression of mab21l2 and vglut2a (shown with slc17a6b) in sorted tectal glutamatergic neurons. Tubb5 and tfap2a are the marker genes of clusters of tectal glutamatergic neurons. (D) The JaccardRainCloundPlot showing the stability of tectal glutamatergic neuronal clusters in Figure 4A. Red line indicated the 0.6 cutoff of Jaccard index to evaluate the cluster stability. Red arrowhead indicated unstable clusters. (E) Left: tSNE plot showing the expression of TF in the clusters of tectal glutamatergic neurons. Right: the percentages of TF-expressed cells in each of 11 clusters (upper) and the averaged UMI of each TF-expressed cell in each cluster (bottom). (F) The matrix showing the expressions of 6 TFs in each of 11 tectal glutamatergic neurons. The labeled were indicated by black squares and no expression was indicated by white squares (G) The matrix showing 574 neurons expressing each of 6 TFs in G-L. All neurons were categorized into 7 morphological subclasses: bi-stratified (n = 36), mono-stratified (n = 71), non-stratified (n = 318), necklace (n = 76), cross-hemisphere (n = 15), ascending projection (n = 18), and descending projection (n = 40). (H) NNLS analysis showing the correlation of morphological subclasses and transcriptome clusters.
**Figure 5-figure supplement 1.** The Post-transcriptional Regulatory Landscape in Neurons with different terminal features but similar TF profiles. (A) Dot graph showing the GO analysis of differential genes between neuronal clusters with convergent pattern. Red highlighted the category of post-transcriptional regulators. The number of genes in each category was coded by dot size. The p.adjust of each category were color-coded. (B) Bar plot showing the ratio of differentially expressed RBPs and all markers between pairs with matched, convergent, and divergent patterns. Different patterns are color-coded. (C) Dot plot showing the p-value of RBPs used as differential genes between pair clusters with matched, convergent, and divergent patterns. Different patterns are color-coded (mean ± SEM). (D) Scatter plot showing the p-value of RBPs which were used to distinguish pair clusters in each pattern. Each
pattern was color-coded (mean ± SEM, matched: 20.73±3.086, convergent: 28.1±3.079, divergent: 13.33±1.493, **p-value < 0.01, ***p-value < 0.001, ns no significant difference; t-test). (E) Venn plot showing the RBPs used as differentially expressed genes between paired clusters with different pattern. Each pattern was color-coded. (F) Bar plot showing the number of pattern-specific RBPs in each patterns. Each pattern was color-coded. (G) The matrix showing the functional category of RBPs which specifically distinguished pair clusters with matched, convergent, and divergent patterns or shared by all three patterns. Different RBPs functions are color-coded. (H-J) Bar plots showing the GO analysis of target sites of pattern specific RBPs (H, matched pattern; I, convergent pattern; and J, divergent pattern) using the oRNAment database.
Figure 1-source data 3. The regional origins and neurotransmitter type annotation of each whole-brain cluster with well-known markers.

Figure 1-Source data 4

Figure 1-source data 4. Top20 marker genes of whole-brain larval zebrafish clusters.

Figure 1-Source data 5

Figure 1-source data 5. Marker genes of major six cell type in whole brain.

Figure 2-source data 1

Figure 2-source data 1. The annotation of neuromodulator-type neuronal types with well-known markers.

Figure 4-source data 1

Figure 4-source data 1. The annotation of TFs expression in each tectal glutamatergic clusters.

Figure 4-source data 2

Figure 4-source data 2. Gene labeled morphological analysis of tectal glutamatergic neurons.
Figure 5-source data 1

Figure 5-source data 1. Differential gene expression between sister pair clusters with matched, convergent, and divergent patterns.

Figure 5-source data 2

Figure 5-source data 2. The p-value of Differential gene expression between sister pair clusters.

Figure 5-source data 3

Figure 5-source data 3. Annotation the function of RBPs.

Figure 5-source data 4

Figure 5-source data 4. The binding sites of pattern specific RBPs.

Figure 5-source data 5

Figure 5-source data 5. The combinatorial usage of RBPs.